

MATHEMATICAL LOGIC  
AND  
HILBERT'S  $\varepsilon$ -SYMBOL

A. C. LEISENRING  
*Reed College, Portland, Oregon*



MACDONALD TECHNICAL & SCIENTIFIC  
LONDON

MACDONALD & CO. (PUBLISHERS) LTD  
49/50 POLAND STREET, LONDON W.1

© 1969 A. C. Leisenring  
SBN 356 02679 5

All Rights Reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of MacDonal & Co. (Publishers) Ltd.

PRINTED AND BOUND IN ENGLAND BY  
HAZELL WATSON AND VINBY LTD  
AYLESBURY, BUCKS

## PREFACE

This study of Hilbert's  $\varepsilon$ -symbol is based on a doctoral dissertation which was submitted to the University of London in June 1967. It contains a number of new results, in particular a strengthening of Hilbert's Second  $\varepsilon$ -Theorem, which should be of some interest to specialists in the field of mathematical logic. However, this book is not written for the expert alone. We have tried to make it as self-contained as possible so that most parts will be intelligible to anyone with a good mathematical background but with a limited knowledge of formal logic. Since the  $\varepsilon$ -symbol can be used to overcome many technical difficulties which arise in formalizing logical reasoning, the present book may be useful as supplementary reading material for an undergraduate course in logic.

In Chapter I we define the formal languages which are used throughout the book and analyze the semantics of such languages. An interesting result concerning finitary closure operations is then established. This result is used in Chapter II to prove the (semantic) completeness of certain formal systems which incorporate the  $\varepsilon$ -symbol. In Chapter II we also establish a number of derived rules of inference for these formal systems. Chapter III, which is more technical than Chapter II, contains proofs of Hilbert's  $\varepsilon$ -Theorems, Skolem's Theorem, and Herbrand's Theorem. If the reader is interested in only these results, he may omit all of Chapter I except for the definitions of formal languages and may skim through most of Chapter II.

Chapter IV deals with formal theories. We see how the  $\varepsilon$ -symbol and  $\varepsilon$ -Theorems may be used to prove the consistency of various mathematical theories. In particular, Hilbert's attempts to prove the consistency of arithmetic are described. We then discuss the role which the  $\varepsilon$ -symbol can play in formalizing set theory, and we give particular attention to the relationship between this symbol and the axiom of choice.

Chapter V may be regarded as an appendix to the book. In this chapter we give alternative proofs of some of the theorems in Chapter III. Our methods illustrate the close connection which exists between Hilbert's  $\varepsilon$ -Theorems and Gentzen's *Hauptsatz*.

The author would like to express his gratitude to G. T. Kneebone (Bedford College, London) under whose helpful supervision his thesis was written. He is also indebted to Marian Cowan for her excellent work in typing the manuscript, and most of all to his wife, Flora, for her help, patience, and encouragement.

A. C. Leisenring  
Reed College  
Portland, Oregon

## CONTENTS

Preface	v
<b>Introduction</b>	
1 Objectives	1
2 Formal languages	2
2.1 The semantic consequence relation $\models$	2
2.2 The deducibility relation $\vdash$	3
2.3 Formal theories and the formalist programme	3
3 The history of the $\varepsilon$ -symbol	4
4 The indeterminacy of the $\varepsilon$ -symbol	5
5 Possible ways of defining formal languages	6
<b>Chapter I Syntax and Semantics</b>	
1 Introduction	9
2 The formal language $\mathcal{L}(\mathcal{V})$	10
2.1 Terms and formulae	12
2.2 A unifying classification of formulae	14
2.3 The cardinality of a language	16
3.1 Truth functions	16
3.2 Tautologies	17
3.3 Models	18
3.4 Satisfiability and the semantic consequence relation	22
4 The Satisfiability Theorem	23
4.1 Logical closures	30
5 Alternative interpretations of the $\varepsilon$ -symbol	32
6 Languages without identity	36
<b>Chapter II Formal Systems</b>	
1.1 Finitary reasoning	37
1.2 General definitions	38
1.3 Axioms and rules of inference	39
2 The $\varepsilon$ -calculus for $\mathcal{V}$	39
3 Derived rules of inference	41
3.1 The $\rightarrow$ -introduction rule (Deduction Theorem)	42
3.2 Some theorems of the $\varepsilon$ -calculus	43
4 Completeness of the $\varepsilon$ -calculus	44

5	The Tautology Theorem	45
6	The consistency of the $\varepsilon$ -calculus	47
7.1	Some derived rules for operating with quantifiers	48
7.2	Substitution instances and universal closures	51
7.3	The equivalence rule	52
7.4	Rule of relabelling bound variables	53
8	Prenex formulae	54
9	The addition of new function symbols	54
9.1	Skolem and Herbrand resolutions of prenex formulae	56
10	The elementary calculus	58
11	The predicate calculus	59
11.1	Derived rules of inference for the predicate calculus	61
11.2	The predicate calculus without identity	62
12	The formal superiority of the $\varepsilon$ -calculus	63
<b>Chapter III The <math>\varepsilon</math>-Theorems</b>		
1	Introduction	64
2.1	The basic problem	65
2.2	Proper and improper formulae	66
2.3	The $\varepsilon^*$ -calculus	66
2.4	The usefulness of the $\varepsilon^*$ -calculus	68
3.1	Subordination and rank	69
3.2	$r, T_r$ -deductions	71
3.3	The Rank Reduction Theorem	72
3.4	The E2 Elimination Theorem	75
3.5	The First and Second $\varepsilon$ -Theorems	77
4.1	Skolem's Theorem	79
4.2	Herbrand's Theorem	80
4.3	The eliminability of the identity symbol	83
<b>Chapter IV Formal Theories</b>		
1	Introduction	85
2	Finitary consistency proofs	85
3	Formal arithmetic	88
3.1	Numerals and numerically true formulae	90
3.2	Formal arithmetic based on the $\varepsilon$ -calculus	92
3.3	Quantifier-free proofs	94
3.4	The consistency of arithmetic	96
4	Theories based on the $\varepsilon$ -calculus	100
4.1	The $\iota$ -symbol	100
4.2	Skolem functions	102
4.3	Definability of indeterminate concepts	103

4.4	The $\varepsilon$ -symbol and the axiom of choice	105
5.1	The predicate calculus with class operator and choice function	107
5.2	The relative consistency of set theory based on the $\varepsilon$ -calculus	112
<b>Chapter V The Cut Elimination Theorem</b>		
1	The sequent calculus	114
2	The axioms and rules of inference of the sequent calculus	115
2.1	Refutations	116
3	The subformula property of normal refutations	117
3.1	The eliminability of the identity symbol	118
4	The contradiction rule	118
4.1	The relationship between the sequent calculus and $\varepsilon$ -calculus	120
5	$k, F_k$ -refutations	121
5.1	The invertibility of the logical rules	121
6	The Cut Elimination Theorem	124
6.1	Applications	126
<b>Bibliography</b>		
<b>Index of Symbols</b>		
<b>Index of Rules of Inference</b>		
<b>General Index</b>		

# INTRODUCTION

## 1 Objectives

The purpose of this book is to examine the nature of Hilbert's  $\varepsilon$ -symbol and to demonstrate the useful role which this logical symbol can play both in proving many of the classical theorems of mathematical logic and also in simplifying the formulation of logical systems and mathematical theories.

The  $\varepsilon$ -symbol is a logical constant which can be used in the formal languages of mathematical logic to form certain expressions known as  $\varepsilon$ -terms. Thus, if  $A$  is a formula of some formal language  $\mathcal{L}$  and  $x$  is a variable of  $\mathcal{L}$ , then the expression  $\varepsilon xA$  is a well-formed term of the language. Intuitively, the  $\varepsilon$ -term  $\varepsilon xA$  says 'an  $x$  such that if anything has the property  $A$ , then  $x$  has that property'. For example, suppose we think of the variables of the language as ranging over the set of human beings and we think of  $A$  as being the statement 'x is an honest politician'. Then  $\varepsilon xA$  designates some politician whose honesty is beyond reproach, assuming of course that such a politician exists. On the other hand, if there are no honest politicians, then  $\varepsilon xA$  must denote someone, but we have no way of knowing who that person is. Similarly, even if there are honest politicians, we have no way of knowing which one of them  $\varepsilon xA$  designates.

Since the  $\varepsilon$ -symbol, or  $\varepsilon$ -operator as it is sometimes called, selects an arbitrary member from a set of objects having some given property, this symbol is often referred to as a 'logical choice function'. It is not surprising then that an investigation of the  $\varepsilon$ -symbol also sheds some light on the nature of the axiom of choice. One of the main theorems of this book, Hilbert's Second  $\varepsilon$ -Theorem, provides a formal justification of the use of the  $\varepsilon$ -symbol in logical systems by showing that this symbol can be eliminated from proofs of formulae which do not themselves contain the symbol. What this theorem says intuitively is that the act of making arbitrary choices is a legitimate logical procedure. However, it has been shown by Cohen [1966] that an application of the axiom of choice in set theory cannot in general be eliminated. It follows then that the real power of the axiom of choice lies not in the fact that it allows one to make arbitrary selections but rather in the fact that it asserts the existence of a set consisting of the selected entities.

Before saying anything more about the nature of the  $\varepsilon$ -symbol and the role it plays, we should first give a brief explanation of the basic objectives and concepts of mathematical logic.

## 2 Formal languages

The objects of study in mathematical logic are formal or symbolic languages, and the primary aim is to provide a formalization of logical reasoning in terms of such languages. A *formal language*  $\mathcal{L}$  consists of a set of symbols together with certain prescribed rules for building well-formed (or grammatical) expressions from these symbols. The well-formed expressions of  $\mathcal{L}$  are usually divided into two categories, *formulae* and *terms*. The formulae correspond to the sentences of ordinary languages and the terms correspond to the nouns, pronouns, and noun clauses. The symbols of the language include certain function and predicate symbols (the *vocabulary* of the language), variables, and logical constants. To the logician the most important symbols are the logical constants. These usually consist of certain *propositional connectives*, such as  $\neg$  ('not'),  $\vee$  ('or'),  $\wedge$  ('and'),  $\rightarrow$  ('implies'), and  $\leftrightarrow$  ('if and only if'), and one or both of the *quantifiers*,  $\forall$  ('for all') and  $\exists$  ('there exists'). The formal languages dealt with in this book also contain the logical constant  $\varepsilon$  (Hilbert's  $\varepsilon$ -symbol).

To avoid any confusion between the formal language being studied and the (informal) language which is used in carrying out this study, one refers to the former as the *object language* and to the latter as the *metalanguage*. Similarly, the mathematical theory which is used in reasoning about the object language is called the *metatheory*. Ideally, the metatheory should be as weak as possible. For example, a proof of the consistency of formal arithmetic would carry little weight if one were to use a strong metatheory which included all of arithmetic and set theory. Therefore, throughout this book whenever possible we shall use a weak 'constructive' metatheory, even though certain theorems can be proved more easily using 'non-constructive' techniques. (A non-constructive argument is one in which the existence of something is proved by deducing a contradiction from the assumption that no such thing exists.)

Having specified a formal language  $\mathcal{L}$ , one can then formalize logical reasoning in either of two ways, (i) in terms of 'models', or (ii) in terms of 'formal systems'.

### 2.1 The semantic consequence relation $\vDash$

Without going into too much detail we can say that a *model* is an abstract mathematical structure which provides an interpretation of the symbols of  $\mathcal{L}$  in such a way that every formula of  $\mathcal{L}$  becomes either a true or a false statement about this structure. The logical constants of  $\mathcal{L}$  are always given their natural interpretation, but the interpretations given to the function and predicate symbols may differ for different models. Thus a formula  $A$  may be true in one model and false in another. A formula  $A$  is said to be a semantic consequence of a set of formulae  $X$ , denoted by  $X \vDash A$ , if  $A$  is true in every model in which all the members of  $X$  are true. For example, suppose that  $\mathcal{L}$

is the language of elementary group theory and  $X$  is the set of formal axioms of this theory. Then to say that  $A$  is a semantic consequence of  $X$  amounts to saying that  $A$  is a statement which holds true for any group  $G$ . Obviously if  $\emptyset \vDash A$ , where  $\emptyset$  is the empty set, then  $A$  is true in all models, and we then say that  $A$  is *valid* (or *logically true*).

### 2.2 The deducibility relation $\vdash$

The semantic consequence relation,  $\vDash$ , provides a useful, though rather abstract, formalization of logical consequence and logical truth. One objection to this approach is that one needs a strong metatheory which permits non-constructive arguments about infinite sets. For many reasons it is often preferable to regard the formulae of a language  $\mathcal{L}$  as concrete objects, i.e., uninterpreted sequences of symbols, and to confine one's metatheory to constructive, combinatorial arguments about these objects.

For this reason one often formalizes logical reasoning in terms of 'deducibility' in a formal system  $\mathcal{F}$ , where a formal system usually consists of certain axioms and rules of inference for the language  $\mathcal{L}$ . The notation  $X \vdash_{\mathcal{F}} A$  is used to denote that  $A$  is deducible from  $X$  in  $\mathcal{F}$ . A formal system  $\mathcal{F}$  is said to be *sound* and (*semantically*) *complete* if the deducibility relation  $\vdash$  for  $\mathcal{F}$  coincides with the semantic consequence relation  $\vDash$ . In other words, by proving that  $\mathcal{F}$  is both sound and complete one justifies the axioms and rules of inference of  $\mathcal{F}$ .

The standard formal system used in mathematical logic is the (first-order) *predicate calculus*, and proofs of the soundness and completeness of this system are included in most textbooks of mathematical logic. The primary formal system which is dealt with in this book is called the  $\varepsilon$ -*calculus*. This system is essentially obtained from the predicate calculus by adjoining the  $\varepsilon$ -symbol as a new logical constant and by introducing some additional axioms for dealing with this symbol. It is our contention that by enlarging the predicate calculus in this way one obtains a much neater and simpler formalization of logical reasoning.

We have now introduced enough of the basic concepts to give a precise statement of Hilbert's Second  $\varepsilon$ -Theorem. Suppose  $A$  is a formula of some language  $\mathcal{L}$ ,  $X$  is a set of formulae of  $\mathcal{L}$ , and the  $\varepsilon$ -symbol does not occur in  $A$  or in any member of  $X$ . The Second  $\varepsilon$ -Theorem states that if  $A$  is deducible from  $X$  in the  $\varepsilon$ -calculus, then  $A$  is deducible from  $X$  in the predicate calculus. A more succinct statement of the theorem would be that the  $\varepsilon$ -calculus is an 'inessential extension' of the predicate calculus.

### 2.3 Formal theories and the formalist programme

Strictly speaking, mathematical logic is a branch of applied mathematics. However, unlike other branches of applied mathematics which are used to

solve problems in the natural and social sciences, mathematical logic is used to prove results about mathematical theories. When mathematical logic is used in this way, the subject is often referred to as *metamathematics*.

A formalization  $\mathcal{F}$  of a particular mathematical theory, e.g. arithmetic or set theory, consists of a formal language  $\mathcal{L}$  which is adequate for expressing the concepts of that theory, a suitable formal system  $\mathcal{F}$ , such as the predicate calculus, and a specified set  $X$  of formulae of  $\mathcal{L}$  which serve as the (non-logical) axioms of the theory. A formula  $A$  of the language  $\mathcal{L}$  is said to be a *theorem* of the formal theory  $\mathcal{F}$  if  $A$  is deducible from  $X$  in  $\mathcal{F}$ . The theory  $\mathcal{F}$  is said to be *inconsistent* if a contradictory formula, i.e., one of the form  $A \wedge \neg A$ , is a theorem of  $\mathcal{F}$ ; otherwise  $\mathcal{F}$  is said to be *consistent*.

By reducing a mathematical theory to a formal theory one can prove various results about that theory in a completely constructive way, since a formal theory is nothing more than a meaningless array of symbols, together with certain prescribed rules for manipulating these symbols.

This study of mathematical theories within a constructive metatheory was first developed by the *formalists* under the leadership of David Hilbert. The discovery of certain paradoxes in set theory around the year 1900 had aroused grave doubts about the legitimacy of the non-constructive techniques which mathematicians often used in dealing with infinite sets. The formalists were convinced that these techniques were justifiable, and they hoped to find such a justification by proving constructively that the basic mathematical theories, such as arithmetic and analysis, were consistent. Unfortunately for them, it was shown by Gödel [1931] that even for such a simple theory as arithmetic no such consistency proof can ever be found.

Despite the fact that their primary goal proved to be unattainable, the formalists made outstanding contributions to the theory of mathematical logic, and paved the way for many important later discoveries.

### 3 The history of the $\varepsilon$ -symbol

The  $\varepsilon$ -symbol was introduced by Hilbert and his collaborators in order to provide explicit definitions of the quantifiers  $\forall$  and  $\exists$ . These definitions are expressed by the formulae

$$(1) \quad \exists xA \leftrightarrow A(\varepsilon xA)$$

and

$$(2) \quad \forall xA \leftrightarrow A(\varepsilon x\neg A),$$

where  $\neg$  is the symbol for 'not'. Hilbert was convinced that by using the  $\varepsilon$ -symbol rather than the quantifiers in formalizing arithmetic and analysis, one could establish the consistency of these two theories. The first published work in which the  $\varepsilon$ -symbol is used is Ackermann's doctoral dissertation [1924], written under Hilbert, in which an attempt is made to prove the

consistency of analysis. However, in the previous year an article by Hilbert [1923] appeared in which a similar symbol, the  $\tau$ -symbol, is used. A comprehensive account of the results which the formalists proved using the  $\varepsilon$ -symbol is given by Hilbert and Bernays [1939].

The  $\varepsilon$ -calculus used by Hilbert and Bernays is essentially the formal system which is obtained from the predicate calculus by adjoining the  $\varepsilon$ -symbol as an additional logical constant and by taking all formulae of the form

$$(3) \quad A(t) \rightarrow A(\varepsilon xA)$$

as additional axioms, where  $t$  is any term. It can then be shown that (1) and (2) follow from (3) by virtue of the axioms and rules of inference of the predicate calculus. The two main results which Hilbert and Bernays prove concerning the  $\varepsilon$ -symbol are known as the First and Second  $\varepsilon$ -Theorems. The first of these is concerned with the eliminability of the quantifiers from the predicate calculus, and as we have already seen, the second is concerned with the eliminability of the  $\varepsilon$ -symbol from the  $\varepsilon$ -calculus.

For the most part, Hilbert and Bernays use their  $\varepsilon$ -calculus only in a subsidiary role to prove that certain deductions in the predicate calculus can be rewritten in a simpler form. However, an  $\varepsilon$ -calculus can be used advantageously as a formal system in its own right. Ackermann [1937-8] and Bourbaki [1954] present interesting formalizations of set theory which are based on an  $\varepsilon$ -calculus. The axioms of their  $\varepsilon$ -calculus include all formulae of the form

$$(4) \quad \forall x(A \leftrightarrow B) \rightarrow \varepsilon xA = \varepsilon xB$$

in addition to formulae of the form (3) above. When this formal system is used in formalizing set theory, there is usually no need to adopt the axiom of choice since this axiom is deducible from an axiom of replacement using axiom (3) above. The conditions under which the axiom of choice is deducible will be discussed in Chapter IV (see page 106).

Naturally, if the  $\varepsilon$ -calculus is used in this way as a formal system in its own right, one would like to know whether the system is sound and complete. The answer to this question depends on the semantic interpretation which is given to the  $\varepsilon$ -symbol. Asser [1957] in his Berlin *Habilitationschrift* interprets the  $\varepsilon$ -symbol as a 'choice function' and proves the soundness and completeness of various forms of the  $\varepsilon$ -calculus. More recently, Hermes [1965] proves the same result for his *Termlogik mit Auswahloperator*. In our proofs of the soundness and completeness of the  $\varepsilon$ -calculus we shall also interpret the  $\varepsilon$ -symbol as a choice function.

### 4 The indeterminacy of the $\varepsilon$ -symbol

One of the most intriguing and useful features of the  $\varepsilon$ -symbol is its indeterminacy. Carnap [1961], pages 162-163, describes this feature as follows:

'The symbol " $\varepsilon$ " was intentionally introduced by Hilbert as an indeterminate constant. Its meaning is specified by the axioms only to the extent that any non-empty set has exactly one representative and this representative is an element of the set. If the set has more than one element, then nothing is said, either officially or unofficially, as to which of the elements is the representative. Thus, for example,  $\varepsilon x(x = 1 \vee x = 2 \vee x = 3)$  must be either 1, or 2, or 3; but there is no way of finding out which it is.'

We might add to Carnap's remarks that the situation is even more mystifying if there is no  $x$  such that  $A$  is true. For example, suppose  $A$  is the formula  $x \neq x$ ; then the term  $\varepsilon x A$  must denote some object, but we have no way of knowing what that object is. A term such as this is called a *null term*.

We shall see that the usefulness of the  $\varepsilon$ -symbol is due essentially to its indeterminate nature. For example, because of its indeterminacy the  $\varepsilon$ -symbol provides useful derived rules of inference for the elimination and introduction of quantifiers (see page 48) and enables one to give explicit definitions of indefinite concepts, such as the concept of cardinal number in set theory (see page 104). However, this indeterminacy also accounts for the suspicion with which the  $\varepsilon$ -symbol is often regarded. We hope that this book will dispel some of that suspicion.

### 5 Possible ways of defining formal languages

We shall conclude these introductory remarks with an explanation of the motivation behind the particular way in which we have defined our formal languages.

In setting up a formal language  $\mathcal{L}$  it is not necessary to take all the logical constants as symbols of the language, since some may be defined in terms of others. For example the symbol  $\leftrightarrow$  may be defined in terms of  $\rightarrow$  and  $\wedge$  by regarding an expression of the form  $A \leftrightarrow B$  as an abbreviation for the expression  $(A \rightarrow B) \wedge (B \rightarrow A)$ . The actual logical symbols of the language are called *primitive* symbols and those symbols which are defined in terms of the primitive symbols are called *defined* symbols.

Our formal languages have the following logical (primitive) symbols:  $f$  (for some false proposition),  $\neg$ ,  $\rightarrow$ ,  $\wedge$ ,  $\vee$ ,  $\exists$ ,  $\forall$ ,  $=$ , and  $\varepsilon$ . The only defined symbol is  $\leftrightarrow$ . It is hoped that by defining our languages with such a rich array of symbols our results are more general than they would be if we were more selective in our choice of logical primitives. In fact, all our main results would still hold if our languages had only a certain subset of these symbols from which the remaining symbols could be defined.

In particular, the symbol  $f$  could be dispensed with by choosing some fixed formula  $A_0$  of  $\mathcal{L}$  and defining  $f$  as  $A_0 \wedge \neg A_0$ . Conversely, the negation symbol  $\neg$  could be defined by letting  $\neg A$  be  $A \rightarrow f$ . For our purposes it is

advisable that at least one of the symbols  $f$  and  $\neg$  be taken as a primitive. With regard to the binary connectives  $\rightarrow$ ,  $\wedge$ , and  $\vee$ , it is possible to define any two of these in terms of the third and the negation symbol. Since we shall be using an elegant unifying notation, due to Smullyan [1965], whereby certain formulae are classified as 'conjunctive' or 'disjunctive' formulae, it will be apparent that all our results would hold if the languages under consideration contained only one or two of these three connectives. Similarly, because of Smullyan's notation, our results would hold if only one of the quantifiers were taken as a primitive and the other were defined in terms of it.

We have seen (page 4) that both the quantifiers can be defined in terms of the  $\varepsilon$ -symbol. However, since we shall be dealing with the predicate calculus, where the  $\varepsilon$ -symbol is unavailable, it is necessary to take at least one of the quantifiers as a primitive. Furthermore, the definition of the quantifiers in terms of the  $\varepsilon$ -symbol leads to certain technical difficulties involving the relabelling of bound variables. Although these difficulties can be overcome in various ways, we shall not deal with them in this book. Methods for defining the quantifiers in terms of the  $\varepsilon$ -symbol appear elsewhere in the literature. See for example, Bourbaki [1954], Hermes [1965], and a paper by the author [1968] in which the main results of Chapter I of this book are proved for languages in which the only logical primitives are  $f$ ,  $\rightarrow$ , and  $\varepsilon$ .

In stating the rules of formation of our languages we do not exclude the possibility of 'vacuous bondage' or 'collisions of bound variables'. Thus an expression of the form  $\exists x A$ ,  $\forall x A$ , or  $\varepsilon x A$  may be well-formed even if the variable  $x$  has no free occurrence in  $A$  (vacuous bondage) or even if  $x$  already has a bound occurrence in  $A$  (collision of bound variables). Nowadays, many logicians allow these two situations because in so doing, various technical difficulties can be avoided, and no new complications arise. For example, we can replace one  $\varepsilon$ -term by another in a formula  $A$  without first relabelling the bound variables in  $A$  to avoid any collisions which might result.

With regard to the actual symbols which we use for individual objects it has proved convenient to have two sorts available: (i) the 'variables'  $x_1, x_2, \dots$  which are denoted syntactically<sup>1</sup> by the letters  $u, v, w, x, y$ , and  $z$ , and (ii) the 'individual symbols'  $a_1, a_2, \dots$  denoted by  $a, b$ , and  $c$ . The variables may be bound by the operators  $\exists, \forall$ , and  $\varepsilon$ , whereas the individual symbols are never bound. The terms and formulae of a language are defined as those well-formed expressions in which no variable has a free (unbound) occurrence. So far, our formulation does not differ from that of Hilbert and Bernays and others who use different symbols for 'free variables' and 'bound variables'. Our approach differs from the Hilbert-type formulation in that we follow the more modern attitude of interpreting the symbols,  $a_1, a_2, \dots$ , as 'arbitrary

<sup>1</sup> In other words, the letters  $u, v, \dots$  are used in the metalanguage as names for unspecified variables.



constants' rather than as 'free variables'. This means that within our formal system there is no general rule of substitution for individual symbols. There are the derived rules that from  $X \vdash A(a)$  one can infer  $X \vdash A(t)$  and  $X \vdash \forall xA(x)$  provided that  $a$  does not appear in any member of  $X$ . However, if  $a$  does appear in the assumption set  $X$ , then the interpretation of  $a$  is temporarily 'fixed' and it cannot be treated as a free variable is treated in a Hilbert-type system. One advantage of this approach is that the statement of the deduction theorem does not involve any complicated restrictions concerning the use of the individual symbols, since these restrictions are already built into the system.

It should be pointed out that instead of specifying a list of individual symbols in our definition of a language, we could require that the vocabulary must contain infinitely many 0-place function symbols. This method is used by Robinson [1963]. These function symbols, or constants, could then be used as the individual symbols. We have not taken this approach since it seems more natural to confine the 0-place function symbols of the vocabulary to those particular constants, such as the symbol for zero in arithmetic, which have a fixed intended interpretation. Within a deduction, the individual symbols do not behave as constants since they are not the names of definite objects. Consequently, these symbols should be regarded more as logical symbols than as vocabulary symbols. In fact, we shall see that the individual symbols in a deduction can be thought of as abbreviations for  $\varepsilon$ -terms. The close connection between these symbols and  $\varepsilon$ -terms is brought out by the fact that in formal systems which incorporate the  $\varepsilon$ -symbol, the individual symbols can be dispensed with, since their role is assumed by the  $\varepsilon$ -terms of the language.

## CHAPTER I

### SYNTAX AND SEMANTICS

#### 1 Introduction

In the study of ordinary languages, such as English, the word 'syntax' is used to refer to the grammatical construction of sentences. Similarly, we define the *syntax* of a formal language  $\mathcal{L}$  to be the basic structure of  $\mathcal{L}$  as laid down by its rules of formation. Thus a syntactical study of  $\mathcal{L}$  officially ignores any intended interpretations of the symbols of  $\mathcal{L}$ . For example, the notion of formula is a syntactic concept, whereas the notions of validity and theorem are not. On the other hand, a *semantic* study of a formal language  $\mathcal{L}$  deals with the possible interpretations which may be given to the symbols, terms, and formulae of  $\mathcal{L}$ . Thus the notions of model, validity, and satisfiability are semantic concepts.

This chapter deals with the syntactic and semantic properties of formal languages which incorporate Hilbert's  $\varepsilon$ -symbol. The main theorem of the chapter, Theorem I.11, establishes an abstract semantic property of such formal languages which is of sufficient generality for both the Compactness Theorem and the Completeness Theorem for the  $\varepsilon$ -calculus to follow from it as corollaries. The Compactness Theorem, which has many applications in other branches of mathematics, can be stated as follows:

For any set  $X$  of formulae of  $\mathcal{L}$ , if every finite subset of  $X$  has a model, then  $X$  has a model whose cardinality is less than or equal to the cardinality of the set of symbols of  $\mathcal{L}$ .

It is well known that Henkin's proof of the completeness of the predicate calculus (Henkin [1949]) can be used to prove the Compactness Theorem. However, since this theorem deals only with models of sets of formulae, it is natural to look for a proof which is purely model-theoretic—that is, one which does not depend on the Completeness Theorem or on the particular set of axioms and rules of inference which have been chosen to give the language a formal deductive structure. One advantage of our abstract approach is that the resulting proof of the Compactness Theorem is model-theoretic in this sense.

In order to carry out our semantic investigations it is necessary to use ordinary mathematical reasoning about sets. Consequently, throughout this chapter our metatheory is set theory with the axiom of choice. The set-

theoretic symbols  $\in$ ,  $\emptyset$ ,  $\subseteq$ ,  $\cup$ ,  $\cap$ ,  $\setminus$ , and  $\{x: \dots x \dots\}$  are used in the usual way.

## 2 The formal language $\mathcal{L}(\mathcal{V})$

A vocabulary  $\mathcal{V}$  is an ordered triple  $\langle Fn, Pr, \rho \rangle$  where  $Fn$  and  $Pr$  are any two disjoint sets and  $\rho$  is a function from  $Fn \cup Pr$  into the set of non-negative integers. The elements of  $Fn$  are called the *function symbols* of  $\mathcal{V}$  and the elements of  $Pr$  are called the *predicate symbols* of  $\mathcal{V}$ . For any  $g$  in  $Fn$  or  $P$  in  $Pr$ ,  $\rho(g)$  is called the *order* of  $g$  and  $\rho(P)$  is called the *order* of  $P$ . Any function symbol (predicate symbol) of order  $n$  is called an *n-place function symbol* (*n-place predicate symbol*). A 0-place function symbol is often called a *constant* and a 0-place predicate symbol is often called a *proposition*. The letters  $g$  and  $h$ , with or without subscripts, will be used as metalinguistic variables to denote arbitrary function symbols, and in particular,  $g^n$  will be used to denote an arbitrary  $n$ -place function symbol. The letter  $P$ , with or without subscripts, will be used to denote an arbitrary predicate symbol, and  $P^n$  to denote an arbitrary  $n$ -place predicate symbol. We place no restriction on the cardinality of the sets  $Fn$  and  $Pr$ . In particular, both may be empty.

Let  $\mathcal{V}$  be the vocabulary  $\langle Fn, Pr, \rho \rangle$  and let  $\mathcal{V}'$  be the vocabulary  $\langle Fn', Pr', \rho' \rangle$ . We say that  $\mathcal{V}'$  is an *extension* of  $\mathcal{V}$  if  $Fn \subseteq Fn'$ ,  $Pr \subseteq Pr'$ , and  $\rho$  and  $\rho'$  agree on  $Fn \cup Pr$ . Given a vocabulary  $\mathcal{V}$ , we often form an extension of that vocabulary by *adjoining* new function symbols or predicate symbols. For example, to formalize arithmetic we might use a vocabulary  $\mathcal{V}$  consisting of the constants 0 and 1, and the 2-place function symbols  $+$  and  $\cdot$ . The vocabulary  $\mathcal{V}'$  obtained from  $\mathcal{V}$  by adjoining the 2-place predicate symbol  $<$  is then an extension of  $\mathcal{V}$ .

We shall now define the unique formal language which is determined by a given vocabulary  $\mathcal{V}$ . We normally denote this language by  $\mathcal{L}(\mathcal{V})$ . However, if it is irrelevant to our discussion what the vocabulary  $\mathcal{V}$  is, we write  $\mathcal{L}$  instead of  $\mathcal{L}(\mathcal{V})$ .

The *symbols* of  $\mathcal{L}(\mathcal{V})$  are:

1. the function symbols and predicate symbols of  $\mathcal{V}$ ;
2. the variables  $x_1, x_2, x_3, \dots$ ;
3. the individual symbols  $a_1, a_2, a_3, \dots$ ;
4. the logical constants  $f, \neg, \rightarrow, \wedge, \vee, \exists, \forall, \varepsilon$ , and  $=$ ;
5. the separation symbols  $(, ),$  and  $,$ .

We require that no symbol comprehended under any one of the clauses 1, 2, 3, 4, and 5 is comprehended under any other. The letters  $u, v, w, x, y, z$ , with or without subscripts, will be used to denote arbitrary variables, and the

letters  $a, b, c$ , with or without subscripts, to denote arbitrary individual symbols. The set of variables will be denoted by  $Vr$  and the set of individual symbols by  $Ind$ .

An *expression of length  $n$*  is any string  $s_1 \dots s_n$  of symbols of  $\mathcal{L}(\mathcal{V})$ . We include the possibility that  $n = 0$ . In this case the expression is called the *empty expression*, and we denote this expression by  $e$ . Given any two expressions  $A$  and  $B$ , we can form a new expression  $AB$  by juxtaposing  $A$  and  $B$ . Thus if  $A$  is an expression of length  $m$  and  $B$  is an expression of length  $n$ , then the length of  $AB$  is  $m + n$ . We say that an expression  $A$  *occurs in* (or *appears in* or *is contained in*) an expression  $B$  if  $B$  is of the form  $B_1AB_2$  where  $B_1$  and  $B_2$  are any two expressions.

The *well-formed expressions* of  $\mathcal{L}(\mathcal{V})$  fall into two disjoint categories, the *quasi-terms* and the *quasi-formulae*. These are defined by the following recursive *rules of formation*.

- G1. Any variable or individual symbol is a quasi-term.
- G2. If  $g^n$  is an  $n$ -place function symbol of  $\mathcal{V}$  ( $n \geq 0$ ), and  $t_1, \dots, t_n$  are quasi-terms, then  $g^n t_1 \dots t_n$  is a quasi-term.
- G3. If  $P^n$  is an  $n$ -place predicate symbol of  $\mathcal{V}$  ( $n \geq 0$ ), and  $t_1, \dots, t_n$  are quasi-terms, then  $P^n t_1 \dots t_n$  is a quasi-formula.
- G4. If  $s$  and  $t$  are quasi-terms, then  $s = t$  is a quasi-formula.
- G5. The symbol  $f$  is a quasi-formula.
- G6. If  $A$  and  $B$  are quasi-formulae, then  $\neg A$ ,  $(A \rightarrow B)$ ,  $(A \wedge B)$ , and  $(A \vee B)$  are quasi-formulae.
- G7. If  $A$  is a quasi-formula, then for any variable  $x$ ,  $\exists xA$  and  $\forall xA$  are quasi-formulae.
- G8. If  $A$  is a quasi-formula, then for any variable  $x$ ,  $\varepsilon xA$  is a quasi-term.
- G9. Only those expressions which are generated by G1–G8 are well-formed.

Any expression which is well-formed by virtue of G8 is called a *quasi  $\varepsilon$ -term* and any which is well-formed by virtue of G3, G4, or G5 is called an *atom*. A well-formed expression is said to be  *$\varepsilon$ -free* if the symbol  $\varepsilon$  does not occur in it, *identity-free* if the symbol  $=$  does not occur in it, and *elementary* if the symbols  $\forall, \exists$ , and  $\varepsilon$  do not occur in it.

Unless otherwise stated, the letters  $A, B$ , and  $C$ , with or without subscripts, will be used to denote arbitrary quasi-formulae, and the letters  $s$  and  $t$ , with or without subscripts, will be used to denote arbitrary quasi-terms. The following syntactic abbreviation is used throughout:

$$(A \leftrightarrow B) \text{ for } ((A \rightarrow B) \wedge (B \rightarrow A)).$$

We shall adopt the following conventions for the omission of parentheses. First of all, we normally omit the outermost pair of parentheses. For example, we write

$$(A \wedge B) \rightarrow C$$

instead of

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

Secondly, we adopt the rule of association to the right. This means that  $A \vee B \vee C$ ,  $A \wedge B \wedge C$ , and  $A \rightarrow B \rightarrow C$  should be read as  $(A \vee (B \vee C))$ ,  $(A \wedge (B \wedge C))$ , and  $(A \rightarrow (B \rightarrow C))$ , respectively. Thus if we restore the parentheses to the abbreviated expression

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

we obtain the expression

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

This rule of association to the right is applied to abbreviated expressions of the form  $A_1 \vee \dots \vee A_n$ ,  $A_1 \wedge \dots \wedge A_n$ , and  $A_1 \rightarrow \dots \rightarrow A_n$ , where the  $A_i$  are any quasi-formulae. (If  $n = 1$ , these three expressions all stand for  $A_1$ .)

Although it is not necessary to enclose quasi-formulae of the form  $s = t$  by parentheses, we often do so for sake of readability. Thus, for example,  $\forall x(x = x)$  and  $\neg(x = y)$  stand for the well-formed expressions  $\forall xx = x$  and  $\neg x = y$ .

## 2.1 Terms and formulae

An occurrence of a variable  $x$  in any well-formed expression  $A$  is said to be a *bound occurrence* if it is an occurrence in a well-formed part of  $A$  of the form  $\exists xB$ ,  $\forall xB$ , or  $\varepsilon xB$ ; otherwise it is a *free occurrence*. The *free variables* of  $A$  are those variables which have free occurrences in  $A$ . We denote the set of free variables in  $A$  by  $V(A)$ . If  $A$  is any well-formed expression and  $x$  any variable, we say  $x$  occurs free in  $A$  within the scope of an  $\varepsilon$ -symbol if  $x$  has a free occurrence in  $A$  within a well-formed part of  $A$  of the form  $\varepsilon yB$ .

For any language  $\mathcal{L}$ , a quasi-formula of  $\mathcal{L}$  in which no variable occurs free is called a *formula* of  $\mathcal{L}$ , and a quasi-term of  $\mathcal{L}$  in which no variable occurs free is called a *term* of  $\mathcal{L}$ . Thus the sets  $F_{\mathcal{L}}$  of all formulae of  $\mathcal{L}$  and  $T_{\mathcal{L}}$  of all terms of  $\mathcal{L}$  are defined by

$$F_{\mathcal{L}} = \{A: A \text{ is a quasi-formula of } \mathcal{L} \text{ and } V(A) = \emptyset\},$$

$$T_{\mathcal{L}} = \{t: t \text{ is a quasi-term of } \mathcal{L} \text{ and } V(t) = \emptyset\}.$$

In particular, a quasi  $\varepsilon$ -term in which no variable occurs free is called an  $\varepsilon$ -term.

We say a quasi-term  $t$  is *free for*  $x$  in a well-formed expression  $A$  if no free occurrence of  $x$  in  $A$  is an occurrence in a well-formed part of  $A$  of the form  $\varepsilon yB$ ,  $\exists yB$ , or  $\forall yB$ , where  $y$  is free in  $t$ . It follows from this definition that if  $t$  is free for  $x$  in  $A$ , then on replacing all free occurrences of  $x$  in  $A$  by  $t$ , no free occurrence of a variable in  $t$  becomes bound. If  $t$  is free for  $x$  in  $A$ , we shall use the notation

$$[A]_t^x$$

to denote that (well-formed) expression which is obtained from  $A$  by replacing all free occurrences of  $x$  in  $A$  by  $t$ . We shall adopt the convention that this notation is used only when  $t$  is free for  $x$  in  $A$ . Thus when we write the formula

$$(1) \quad \forall z([A]_z^x \leftrightarrow [B]_z^y) \rightarrow (\varepsilon xA = \varepsilon yB),$$

it is understood that  $z$  is free for  $x$  in  $A$  and for  $y$  in  $B$ . (As will be seen later, if all formulae of this form are taken as logical axioms and the usual axioms for identity are available, it is unnecessary to adopt a rule for the relabelling of bound variables.)

Since a *term* contains no free variables, then any term is free for  $x$  in  $A$ , for any  $x$  and any  $A$ . In practice, the notation  $[A]_t^x$  is seldom used when  $t$  is not a term—formula (1) will be one of the few cases in which it is so used. Consequently, our convention that  $t$  is free for  $x$  in  $A$  is for the most part superfluous.

When no confusion can arise, the notation  $[A]_t^x$  will often be simplified to  $A_t^x$  or just  $A(t)$ . However, at times the full notation is necessary. For example,  $\varepsilon y[A]_t^x$  and  $[\varepsilon yA]_t^x$  do not necessarily denote the same expression, and therefore the notation  $\varepsilon yA_t^x$  would be ambiguous. In view of our abbreviated notation, the expressions

$$\forall xA \rightarrow A(t) \quad \text{and} \quad \neg \exists xA \rightarrow \neg A(t)$$

are to be understood as

$$\forall xA \rightarrow [A]_t^x \quad \text{and} \quad \neg \exists xA \rightarrow \neg [A]_t^x.$$

The following properties are easy to verify.

### THEOREM I.1.

- (i) If  $s$  and  $t$  are terms and  $x$  and  $y$  are distinct variables, then  $[[A]_s^x]_t^y$  and  $[[A]_t^y]_s^x$  denote the same expression;
- (ii) If  $x$  and  $y$  are distinct variables, then  $[\exists yA]_t^x$  and  $\exists y[A]_t^x$  denote the same expression,  $[\forall yA]_t^x$  and  $\forall y[A]_t^x$  denote the same expression, and  $[\varepsilon yA]_t^x$  and  $\varepsilon y[A]_t^x$  denote the same expression.

In the case of multiple replacements of free variables, in order to economize in the use of brackets, we shall write

$$(1) \quad [A]_{t_1^{x_1} \dots t_n^{x_n}}$$

instead of

$$(2) \quad [\dots [[A]_{t_1^{x_1}}] \dots ]_{t_n^{x_n}}.$$

Thus (1) denotes the expression which is obtained from  $A$  by first replacing each free occurrence of  $x_1$  by  $t_1$ , then each free occurrence of  $x_2$  by  $t_2$ , etc.

By Theorem I.1(i) it follows that if each  $t_i$  is a term and if the variables  $x_1, \dots, x_n$  are all distinct, then the order in which these replacements are performed is immaterial.

A *replacement operator* for a language  $\mathcal{L}$  is any function, usually denoted by  $\Sigma$ , from a finite set of variables, called the *domain* of  $\Sigma$ , into  $T_{\mathcal{L}}$ . We shall denote the domain of  $\Sigma$  by  $\text{dom } \Sigma$ . For any well-formed expression  $A$  of  $\mathcal{L}$  we define  $[A]\Sigma$  as

$$[A]_{\Sigma(x_1) \dots \Sigma(x_n)}^{x_1 \dots x_n}$$

where the  $x_i$  are the distinct members of  $\text{dom } \Sigma$ . Since each  $\Sigma(x_i)$  is a term, this definition is independent of the order in which the replacements are performed. If  $\text{dom } \Sigma = \emptyset$ , then  $[A]\Sigma$  is  $A$ . When there is no possibility of ambiguity, we write  $A\Sigma$  instead of  $[A]\Sigma$ .

If  $\Sigma$  is any replacement operator,  $x$  any variable, and  $t$  any term, then  $\Sigma_t^x$  is defined to be that function whose domain is  $\text{dom } \Sigma \cup \{x\}$ , such that for all  $y \in \text{dom } \Sigma \cup \{x\}$

$$\Sigma_t^x(y) = \begin{cases} \Sigma(y) & \text{if } y \text{ is distinct from } x, \\ t & \text{if } y \text{ is } x. \end{cases}$$

For any  $\Sigma$  and any variable  $x$ , the *x-suppression* of  $\Sigma$  is that function which is obtained from  $\Sigma$  by restricting its domain to  $\text{dom } \Sigma \setminus \{x\}$ . Thus if  $\Sigma'$  is the *x-suppression* of  $\Sigma$ , then for all  $y \in \text{dom } \Sigma \setminus \{x\}$ ,

$$\Sigma'(y) = \Sigma(y).$$

The following properties of replacement operators follow immediately from the definitions and Theorem I.1.

**THEOREM I.2.** *For any replacement operator  $\Sigma$ :*

- (i) *If  $A$  is any quasi-formula and  $V(A) \subseteq \text{dom } \Sigma$ , then  $A\Sigma$  is a formula. Similarly, for any quasi-term  $t$ .*
- (ii) *For any well-formed expression  $A$ , and any variable  $x$ , if  $x \notin \text{dom } \Sigma$ , then  $[A]\Sigma_t^x$  and  $[[A]\Sigma]^x_t$  denote the same expression.*
- (iii) *If  $\Sigma'$  is the  $x$ -suppression of  $\Sigma$ , then  $[\exists x A]\Sigma$  and  $\exists x[A]\Sigma'$  denote the same expression,  $[\forall x A]\Sigma$  and  $\forall x[A]\Sigma'$  denote the same expression, and  $[exA]\Sigma$  and  $ex[A]\Sigma'$  denote the same expression.*

## 2.2 A unifying classification of formulae

The formulae of any language  $\mathcal{L}$  can be classified according to their syntactic structure as follows:

1. the atoms and the negations of atoms;
2. the formulae of the form  $\neg \neg A$ ;
3. the formulae of the form  $A \wedge B$ ,  $\neg(A \vee B)$ , and  $\neg(A \rightarrow B)$ ;

4. the formulae of the form  $A \vee B$ ,  $\neg(A \wedge B)$ , and  $A \rightarrow B$ ;
5. the formulae of the form  $\forall x A$  and  $\neg \exists x A$ ;
6. the formulae of the form  $\exists x A$  and  $\neg \forall x A$ .

Clearly, any formula of  $\mathcal{L}$  is included in one and only one of the above six categories. We shall use the metalinguistic variables  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\delta$  to denote arbitrary formulae of types 3, 4, 5, and 6, respectively.

In particular, a formula of type 3 is called a *conjunctive* formula. The *conjunctive components*,  $\alpha_1$  and  $\alpha_2$ , of a conjunctive formula  $\alpha$  are defined as follows: if  $\alpha$  has the form  $A \wedge B$ ,  $\neg(A \vee B)$ , or  $\neg(A \rightarrow B)$ , then  $\alpha_1$  and  $\alpha_2$  are  $A$  and  $B$ ,  $\neg A$  and  $\neg B$ , or  $A$  and  $\neg B$ , respectively.

A formula of type 4 is called a *disjunctive* formula. The *disjunctive components*,  $\beta_1$  and  $\beta_2$ , of a disjunctive formula  $\beta$  are defined as follows: if  $\beta$  has the form  $A \vee B$ ,  $\neg(A \wedge B)$ , or  $A \rightarrow B$ , then  $\beta_1$  and  $\beta_2$  are  $A$  and  $B$ ,  $\neg A$  and  $\neg B$ , or  $\neg A$  and  $B$ , respectively.

A formula of type 5 is called a *universal* formula. If  $\gamma$  is a universal formula of the form  $\forall x A$  or  $\neg \exists x A$ , then for any term  $t$  we use the notation  $\gamma(t)$  to denote the formula  $A(t)$  or  $\neg A(t)$ , respectively.

A formula of type 6 is called an *existential* formula. If  $\delta$  is an existential formula of the form  $\exists x A$  or  $\neg \forall x A$ , then for any term  $t$ , we use the notation  $\delta(t)$  to denote the formula  $A(t)$  or  $\neg A(t)$ , respectively. Furthermore, we use the notation  $ex\delta$  to denote the term  $exA$ , if  $\delta$  is  $\exists x A$ , and the term  $ex\neg A$ , if  $\delta$  is  $\neg \forall x A$ . Thus  $\delta(ex\delta)$  denotes  $A(exA)$  or  $\neg A(ex\neg A)$ , respectively.

This system of classifying formulae into types 3, 4, 5, and 6, which is due to Smullyan [1965], greatly simplifies the metatheory since in many proofs and definitions we can avoid tiresome considerations of cases. Furthermore, since in the proofs of all our main theorems we shall not be concerned with the particular syntactic structure of a given formula, but only its more general structure as defined in the above six categories, then it is easy to see that our results hold not only for languages as defined on pages 10–12, but also for languages whose logical symbols include only one quantifier, and only one or two of the three binary connectives. In this way the above classifications and unifying notation give our metamathematical investigations an added degree of generality.

For any formula  $A$ , we define the *contrary* of  $A$  as follows: If  $A$  is the negation of some formula  $B$ , then  $B$  is the contrary of  $A$ , and if  $A$  is not a negation, then  $\neg A$  is the contrary of  $A$ . Consequently, the contrary of a disjunctive formula is a conjunctive formula, and vice versa. Similarly, the contrary of an existential formula is a universal formula, and vice versa. Two formulae are said to be *contradictory* if one is the negation of the other. The following very useful properties of contradictoriness will be referred to as the *duality principle*:

1. If  $\alpha$  and  $\beta$  are contradictory, then  $\alpha_1$  and  $\beta_1$  are contradictory, and  $\alpha_2$  and  $\beta_2$  are contradictory.
2. If  $\gamma$  and  $\delta$  are contradictory, then for any term  $t$ ,  $\gamma(t)$  and  $\delta(t)$  are contradictory.

### 2.3 The cardinality of a language

The *cardinality* of a language  $\mathcal{L}$  is defined as the cardinality of its set of symbols, and by an abuse of notation will be denoted by  $\overline{\mathcal{L}}$ . Since every language has enumerably many variables and individual symbols, the cardinality of a language is at least  $\aleph_0$ . Because we allow the vocabulary of a language to be of arbitrary cardinality, it is possible for languages to be non-enumerable. In fact, all mathematical theories can be formulated within an enumerable language; however, in recent years it has been found that non-enumerable languages play an important role in various metamathematical investigations, particularly in the realm of algebra. One metamathematical result involving non-enumerable languages which has fruitful applications is the Compactness Theorem.

In connection with our proof of the Compactness Theorem we shall make use of the following fact. If  $\mathcal{L}$  is any language, and  $T_{\mathcal{L}}$  is the set of its terms, then  $\overline{T_{\mathcal{L}}} \leq \overline{\mathcal{L}}$ . This can be proved by letting  $E_{\mathcal{L}}$  be the set of expressions of  $\mathcal{L}$ . Since the set of symbols of  $\mathcal{L}$  is infinite, and since  $E_{\mathcal{L}}$  consists of all finite sequences of these symbols, then by a familiar set-theoretical argument,  $\overline{E_{\mathcal{L}}} = \overline{\mathcal{L}}$ . The result then follows from the fact that  $T_{\mathcal{L}}$  is a subset of  $E_{\mathcal{L}}$ .

### 3.1 Truth functions

We begin our study of semantics by formalizing our intuitive interpretations of the symbols  $\neg$  ('not'),  $\vee$  ('or'),  $\wedge$  ('and'), and  $\rightarrow$  ('implies').

Regardless of what is meant by the words 'true' and 'false' we would like to arrange matters in such a way that the formula  $\neg A$  is 'true' iff<sup>1</sup>  $A$  is 'false',  $A \wedge B$  is 'true' iff both  $A$  and  $B$  are 'true',  $A \vee B$  is 'true' iff at least one of  $A$  and  $B$  is 'true', and  $A \rightarrow B$  is 'false' iff  $A$  is 'true' and  $B$  is 'false'.

We proceed as follows. Let  $\mathcal{B}$  denote the set  $\{0,1\}$ , where the numbers 0 and 1 are called *truth values*. (Intuitively, 0 and 1 denote falsehood and truth, respectively.) An  $n$ -ary *truth function* is a function from  $\mathcal{B}^n$  to  $\mathcal{B}$ . To the logical symbols  $\neg$ ,  $\wedge$ ,  $\vee$ , and  $\rightarrow$  we assign the truth functions  $H_{\neg}$ ,  $H_{\wedge}$ ,  $H_{\vee}$ , and  $H_{\rightarrow}$ , respectively, where  $H_{\neg}$  is unary and the others are binary. These functions are defined as follows. For any  $m, n \in \mathcal{B}$ :

$$\begin{aligned} H_{\neg}(m) &= 1 \quad \text{iff} \quad m = 0, \\ H_{\wedge}(m,n) &= 1 \quad \text{iff} \quad m = 1 \text{ and } n = 1, \\ H_{\vee}(m,n) &= 1 \quad \text{iff} \quad m = 1 \text{ or } n = 1, \\ H_{\rightarrow}(m,n) &= 0 \quad \text{iff} \quad m = 1 \text{ and } n = 0. \end{aligned}$$

<sup>1</sup> We use *iff* as an abbreviation for *if and only if*.

Consequently, if by some means we have assigned the truth values  $m$  and  $n$  to  $A$  and  $B$  respectively, then  $H_{\neg}(m)$  is the correct truth value of  $\neg A$ ,  $H_{\wedge}(m, n)$  the correct truth value of  $A \wedge B$ , etc. We shall now consider one possible way in which truth values can be assigned to formulae.

### 3.2 Tautologies

A formula is a *molecule* if it is either an atom or a formula of the form  $\forall xB$  or  $\exists xB$ . Thus every formula  $A$  either is a molecule or is 'built up' from molecules by means of the symbols  $\wedge$ ,  $\vee$ ,  $\neg$ , and  $\rightarrow$ . These molecules in  $A$  are called the *molecular constituents* of  $A$ .

A *truth assignment* is any function  $\psi$  from a finite set  $X$  of molecules into the set  $\mathcal{B}$  which is such that  $\psi(f) = 0$  if  $f$  is a member of  $X$ . For any formula  $A$ , if  $\psi$  is a truth assignment whose domain contains all the molecular constituents of  $A$ , then the truth value  $\overline{\psi}(A)$  which  $\psi$  assigns to  $A$  is defined as follows by induction on the length of  $A$ :

- (i) if  $A$  is a molecule,  $\overline{\psi}(A) = \psi(A)$ ;
- (ii) if  $A$  is of the form  $\neg B$ , then  $\overline{\psi}(A) = H_{\neg}(\overline{\psi}(B))$ ;
- (iii) if  $A$  is of the form  $B * C$ , where  $*$  may be either  $\wedge$ ,  $\vee$ , or  $\rightarrow$ , then  $\overline{\psi}(A) = H_{*}(\overline{\psi}(B), \overline{\psi}(C))$ .

A truth assignment  $\psi$  is said to be a *truth assignment for A* if the domain of  $\psi$  is the set of molecular constituents of  $A$ . Thus if  $A$  has  $n$  molecular constituents, then there are  $2^n$  truth assignments for  $A$ . A formula  $A$  is a *tautology* if  $\overline{\psi}(A) = 1$  for every truth assignment  $\psi$  for  $A$ , and  $A$  is a *tautological consequence* of formulae  $B_1, \dots, B_n$  if the formula  $B_1 \rightarrow \dots \rightarrow B_n \rightarrow A$  is a tautology. Thus a tautology is a formula which is 'true' no matter what truth values are assigned to its molecular constituents, and  $A$  is a tautological consequence of  $B_1, \dots, B_n$  if  $A$  is 'true' for every truth assignment which gives each of the  $B_i$  the value 'true'.

Although a tautology must certainly be regarded as a 'logically true' formula, the notions of tautology and tautological consequence do not give us the complete picture. For, in computing  $\overline{\psi}(A)$  one looks at only the 'molecular structure' of  $A$  and disregards any occurrences of the quantifiers,  $\varepsilon$ -symbol, or identity symbol in  $A$ . For example, according to our intuitive understanding of the symbols  $\exists$ ,  $\forall$ , and  $\varepsilon$ , the formula  $\exists xPx$  should be a 'logical consequence' of  $\forall xPx$ , and the formula  $PexPx$  should be a 'logical consequence' of  $\exists xPx$ . However, these are not tautological consequences since  $\forall xPx \rightarrow \exists xPx$  and  $\exists xPx \rightarrow PexPx$  are not tautologies. Thus for the type of languages which we are considering we need a more inclusive definition of 'logical consequence', i.e., one which analyzes the 'sub-atomic structure' of formulae and reflects our intuitive understanding of the symbols  $\forall$ ,  $\exists$ ,  $\varepsilon$ , and  $=$ , as well as the symbols  $\neg$ ,  $\wedge$ ,  $\vee$ , and  $\rightarrow$ .

Despite these inadequacies, the notion of a tautology is still a very useful one, mainly because of its decidability. For, one can effectively determine whether or not a given formula  $A$  is a tautology by computing  $\bar{\psi}(A)$  for each of the  $2^n$  truth assignments for  $A$ , where  $n$  is the number of molecular constituents of  $A$ . For this reason tautologies play an important role in the study of formal systems, as we shall see in Chapters II and III.

### EXERCISES

1. Prove that any formula of the form  $(A \rightarrow f) \rightarrow \neg A$  is a tautology.
2. Prove that if both  $A$  and  $A \rightarrow B$  are tautologies, then  $B$  is a tautology.
3. Let  $\alpha$  be any conjunctive formula and  $\psi$  any truth assignment for  $\alpha$ . Prove that  $\bar{\psi}(\alpha) = 1$  iff  $\bar{\psi}(\alpha_1) = 1$  and  $\bar{\psi}(\alpha_2) = 1$ .
4. Let  $\beta$  be any disjunctive formula and  $\psi$  any truth assignment for  $\beta$ . Prove that  $\bar{\psi}(\beta) = 1$  iff  $\bar{\psi}(\beta_1) = 1$  or  $\bar{\psi}(\beta_2) = 1$ .

### 3.3 Models

In order to give a precise description of the semantics of our formal languages we define the notion of a model. Because of the  $\varepsilon$ -symbol, it is necessary to modify the conventional definition by equipping each model  $\mathfrak{M}$  with a choice function  $\Phi$ . This choice function provides a semantic interpretation of the  $\varepsilon$ -symbol.

The following familiar set-theoretic notation will be used in our definition. If  $M$  is any set and  $n$  is a positive integer, then  $M^n$  denotes the set of ordered  $n$ -tuples of  $M$ ,  $M^{M^n}$  denotes the set of functions from  $M^n$  into  $M$ , and  $\{0,1\}^{M^n}$  denotes the set of functions defined on  $M^n$  and taking as values 0 or 1. The set  $\{0,1\}^{M^n}$  can be identified with the set of  $n$ -ary relations on  $M$ . If  $n = 0$ , then  $M^{M^n}$  is  $M$ , and  $\{0,1\}^{M^n}$  is  $\{0,1\}$ .

Let  $\mathcal{V}$  be any vocabulary  $\langle Fn, Pr, \rho \rangle$ . A model,  $\mathfrak{M}$ , for  $\mathcal{V}$  is an ordered triple  $\langle M, \Pi, \Phi \rangle$  which satisfies the following conditions:

1.  $M$  is a non-empty set, called the *universe* of  $\mathfrak{M}$ .
2.  $\Pi$  is a function defined on  $F_n \cup Pr \cup Ind$  which assigns values in the following way:
  - (i) for any individual symbol  $a$  in  $Ind$ ,  $\Pi(a) \in M$ ;
  - (ii) for any  $n$ -place function symbol  $g$  in  $F_n$ ,  $\Pi(g) \in M^{M^n}$ ;
  - (iii) for any  $n$ -place predicate symbol  $P$  in  $Pr$ ,  $\Pi(P) \in \{0,1\}^{M^n}$ .
3.  $\Phi$  is a choice function on  $M$ , i.e.  $\Phi(N) \in N$  for any non-empty subset  $N$  of  $M$ , and  $\Phi(\emptyset)$  is an arbitrary, but fixed, member of  $M$ .

The *cardinality* of a model is defined as the cardinality of its universe. Members of the universe will be denoted by the Greek letters  $\mu$  and  $\nu$ .

For any vocabulary  $\mathcal{V}$ , any model  $\mathfrak{M}$  for  $\mathcal{V}$ , and any well-formed expression  $A$  of some language  $\mathcal{L}$ , we say that  $\mathfrak{M}$  is *adequate* for  $A$  if  $A$  is a well-formed

expression of  $\mathcal{L}(\mathcal{V})$ . In other words,  $\mathfrak{M}$  is adequate for  $A$  if the function  $\Pi$  in  $\mathfrak{M}$  assigns values (in the proper way) to each function and predicate symbol occurring in  $A$ .

For any model  $\mathfrak{M}$ , an  $\mathfrak{M}$ -assignment is any function from the set of variables  $Vr$  into the universe of  $\mathfrak{M}$ .  $\mathfrak{M}$ -assignments will be denoted by the Greek letter  $\theta$ , with subscripts if necessary. If  $\theta$  is an  $\mathfrak{M}$ -assignment, then  $\theta_\mu^x$  is that  $\mathfrak{M}$ -assignment whose value for  $x$  is  $\mu$  and which otherwise coincides with  $\theta$ .

If  $A$  is any well-formed expression,  $\mathfrak{M}$  is any model which is adequate for  $A$ , and  $\theta$  is an  $\mathfrak{M}$ -assignment, then the *interpretation of  $A$  with respect to  $\mathfrak{M}$  and  $\theta$* , which is denoted by  $\mathfrak{M}\theta(A)$ , is defined as follows by induction on the length of  $A$ :

- G1.  $A$  is a variable  $x$ :  $\mathfrak{M}\theta(x) = \theta(x)$ ;
- $A$  is an individual symbol  $a$ :  $\mathfrak{M}\theta(a) = \Pi(a)$ .
- G2.  $A$  is of the form  $g^n t_1 \dots t_n$ :  
 $\mathfrak{M}\theta(g^n t_1 \dots t_n) = \Pi(g^n)(\mathfrak{M}\theta(t_1), \dots, \mathfrak{M}\theta(t_n))$ .
- G3.  $A$  is of the form  $P^n t_1 \dots t_n$ :  
 $\mathfrak{M}\theta(P^n t_1 \dots t_n) = \Pi(P^n)(\mathfrak{M}\theta(t_1), \dots, \mathfrak{M}\theta(t_n))$ .
- G4.  $A$  is of the form  $(s = t)$ :  
 $\mathfrak{M}\theta(s = t) = 1$  if  $\mathfrak{M}\theta(s) = \mathfrak{M}\theta(t)$ , and 0 otherwise.
- G5.  $A$  is the formula  $f$ :  
 $\mathfrak{M}\theta(f) = 0$ .
- G6.  $A$  is of the form  $\neg B$ ,  $(B \wedge C)$ ,  $(B \vee C)$ , or  $(B \rightarrow C)$ :  
 $\mathfrak{M}\theta(\neg B) = H_{\neg}(\mathfrak{M}\theta(B))$ ,  
 $\mathfrak{M}\theta(B \wedge C) = H_{\wedge}(\mathfrak{M}\theta(B), \mathfrak{M}\theta(C))$ ,  
 $\mathfrak{M}\theta(B \vee C) = H_{\vee}(\mathfrak{M}\theta(B), \mathfrak{M}\theta(C))$ ,  
 $\mathfrak{M}\theta(B \rightarrow C) = H_{\rightarrow}(\mathfrak{M}\theta(B), \mathfrak{M}\theta(C))$ .
- G7.  $A$  is of the form  $\exists y B$ :  
 $\mathfrak{M}\theta(\exists y B) = 1$  if there exists a  $\mu \in M$  such that  $\mathfrak{M}\theta_\mu^y(B) = 1$ ; otherwise  $\mathfrak{M}\theta(\exists y B) = 0$ ;
- $A$  is of the form  $\forall y B$ :  
 $\mathfrak{M}\theta(\forall y B) = 1$  if for all  $\mu \in M$ ,  $\mathfrak{M}\theta_\mu^y(B) = 1$ ; otherwise  $\mathfrak{M}\theta(\forall y B) = 0$ .
- G8.  $A$  is of the form  $\varepsilon y B$ :  
 $\mathfrak{M}\theta(\varepsilon y B) = \Phi\{\mu: \mathfrak{M}\theta_\mu^y(B) = 1\}$ .

Obviously, if  $A$  is a quasi-formula,  $\mathfrak{M}\theta(A) \in \{0,1\}$ , and if  $A$  is a quasi-term,  $\mathfrak{M}\theta(A) \in M$ .

For any subset  $V$  of  $Vr$ , and any  $\mathfrak{M}$ -assignments  $\theta_1$  and  $\theta_2$ , we write  $\theta_1 \sim_V \theta_2$  to denote that  $\theta_1(x) = \theta_2(x)$ , for all  $x \in V$ .

**THEOREM I.3.** For any well-formed expression  $A$ , any model  $\mathfrak{M}$  which is adequate for  $A$ , and any  $\mathfrak{M}$ -assignments  $\theta_1$  and  $\theta_2$ , if  $\theta_1 \sim_{V(A)} \theta_2$ , then  $\mathfrak{M}\theta_1(A) = \mathfrak{M}\theta_2(A)$ .

*Proof.* The proof is by induction on the length of  $A$ . For cases G1–G6, the proof is trivial.

G7: Suppose  $A$  is of the form  $\exists yB$ . Since  $\theta_1 \sim \theta_2$ , then for any  $\mu \in M$ ,  $\theta_1 \stackrel{V(A)}{\sim} \theta_2$ . Consequently, by the induction hypothesis,  $\mathfrak{M}\theta_1 \stackrel{V(B)}{\sim} \mathfrak{M}\theta_2$ . It follows that there exists a  $\mu \in M$  such that  $\mathfrak{M}\theta_1 \stackrel{V(B)}{\sim} \mathfrak{M}\theta_2$ . Hence,  $\mathfrak{M}\theta_1(\exists yB) = \mathfrak{M}\theta_2(\exists yB)$ . Similarly, it follows that  $\mathfrak{M}\theta_1(\forall yB) = \mathfrak{M}\theta_2(\forall yB)$ .

G8: Suppose  $A$  is of the form  $\varepsilon yB$ . As in G7, for all  $\mu \in M$ ,  $\mathfrak{M}\theta_1 \stackrel{V(B)}{\sim} \mathfrak{M}\theta_2$ . Consequently,  $\{\mu: \mathfrak{M}\theta_1 \stackrel{V(B)}{\sim} \mathfrak{M}\theta_2\} = \{\mu: \mathfrak{M}\theta_2 \stackrel{V(B)}{\sim} \mathfrak{M}\theta_1\}$ . Hence,  $\mathfrak{M}\theta_1(\varepsilon yB) = \mathfrak{M}\theta_2(\varepsilon yB)$ .

**COROLLARY.** *If  $x \notin V(A)$ , then for any  $\mu \in M$ ,  $\mathfrak{M}\theta(A) = \mathfrak{M}\theta_\mu^x(A)$ .*

**THEOREM I.4.** *For any well-formed expression  $A$ , any variable  $x$ , any quasi-term  $t$  which is free for  $x$  in  $A$ , and any model  $\mathfrak{M}$  which is adequate for  $A$  and  $t$ ,  $\mathfrak{M}\theta([A]_t^x) = \mathfrak{M}\theta_{\mathfrak{M}\theta(t)}^x(A)$ .*

*Proof.* If  $x \notin V(A)$ , then  $[A]_t^x$  is  $A$ , and by the above corollary,  $\mathfrak{M}\theta_{\mathfrak{M}\theta(t)}^x(A) = \mathfrak{M}\theta(A)$ . Hence, in this case the theorem obviously holds. We now consider the case where  $x \in V(A)$ . The proof is by induction on the length of  $A$ . For cases G1–G6, the proof is trivial.

G7 and G8: Suppose  $A$  is of the form  $\exists yB$ ,  $\forall yB$ , or  $\varepsilon yB$ . Since  $x \in V(A)$ , then  $x$  and  $y$  are distinct variables. Furthermore, since  $t$  is free for  $x$  in  $A$ , then  $y \notin V(t)$ . Take any  $\mu \in M$ . By the above corollary,  $\mathfrak{M}\theta_\mu^x(t) = \mathfrak{M}\theta(t)$ . Hence, by the induction hypothesis

$$\mathfrak{M}\theta_\mu^x(B_t^x) = \mathfrak{M}\theta_\mu^x \stackrel{x}{\mathfrak{M}\theta(t)}(B),$$

and therefore since  $x$  and  $y$  are distinct variables,

$$\mathfrak{M}\theta_\mu^x(B_t^x) = \mathfrak{M}\theta_{\mathfrak{M}\theta(t)}^x \stackrel{y}{\mu}(B).$$

Consequently,

$$\{\mu: \mathfrak{M}\theta_\mu^x(B_t^x) = 1\} = \{\mu: \mathfrak{M}\theta_{\mathfrak{M}\theta(t)}^x \stackrel{y}{\mu}(B) = 1\}.$$

This implies:

$$\mathfrak{M}\theta(\exists y[B]_t^x) = \mathfrak{M}\theta_{\mathfrak{M}\theta(t)}^x(\exists yB),$$

$$\mathfrak{M}\theta(\forall y[B]_t^x) = \mathfrak{M}\theta_{\mathfrak{M}\theta(t)}^x(\forall yB),$$

and

$$\mathfrak{M}\theta(\varepsilon y[B]_t^x) = \mathfrak{M}\theta_{\mathfrak{M}\theta(t)}^x(\varepsilon yB).$$

The theorem now follows by Theorem I.1.

If  $A$  is a formula or term, then by Theorem I.3 the interpretation of  $A$  with respect to  $\mathfrak{M}$  and  $\theta$  is independent of  $\theta$ , since  $V(A) = \emptyset$ . In this case we may refer to the interpretation of  $A$  with respect to  $\mathfrak{M}$  and write  $\mathfrak{M}(A)$ .

**THEOREM I.5.** *For any adequate model  $\mathfrak{M}$ :*

- (i) *If  $\alpha$  is a conjunctive formula with components  $\alpha_1$  and  $\alpha_2$ , then  $\mathfrak{M}(\alpha) = 1$  iff  $\mathfrak{M}(\alpha_1) = 1$  and  $\mathfrak{M}(\alpha_2) = 1$ .*
- (ii) *For any formulae  $A$  and  $B$ ,  $\mathfrak{M}(A \leftrightarrow B) = 1$  iff  $\mathfrak{M}(A) = \mathfrak{M}(B)$ .*
- (iii) *For any universal formula  $\gamma$  and any term  $t$ , if  $\mathfrak{M}(\gamma) = 1$ , then  $\mathfrak{M}(\gamma(t)) = 1$ .*
- (iv) *For any existential formula  $\delta$ , if  $\mathfrak{M}(\delta) = 1$ , then  $\mathfrak{M}(\delta(\varepsilon\delta)) = 1$ .*
- (v) *If  $\mathfrak{M}(\forall x(A \leftrightarrow B)) = 1$ , then  $\mathfrak{M}(\varepsilon xA) = \mathfrak{M}(\varepsilon xB)$ .*
- (vi) *For any  $\varepsilon$ -term  $\varepsilon xA$ ,  $\mathfrak{M}(\varepsilon xA) = \mathfrak{M}(\varepsilon y[A]_y^x)$ .*
- (vii) *For any terms  $s$  and  $t$ , if  $\mathfrak{M}(s) = \mathfrak{M}(t)$  and  $\mathfrak{M}(A_s^x) = 1$ , then  $\mathfrak{M}(A_t^x) = 1$ .*

*Proof.* The proofs of (i) and (ii) follow trivially from our truth functional interpretations of  $\neg$ ,  $\wedge$ ,  $\vee$ , and  $\rightarrow$ , and from the definition of  $\leftrightarrow$ .

(iii): Suppose  $\gamma$  is of the form  $\forall xA$ . Let  $\theta$  be any  $\mathfrak{M}$ -assignment. Since  $\mathfrak{M}(\forall xA) = 1$ , then  $\mathfrak{M}\theta(\forall xA) = 1$ , and for all  $\mu \in M$ ,  $\mathfrak{M}\theta_\mu^x(A) = 1$ . In particular,  $\mathfrak{M}\theta_{\mathfrak{M}\theta(t)}^x(A) = 1$ , and by Theorem I.4  $\mathfrak{M}\theta(\gamma(t)) = 1$ . Hence,  $\mathfrak{M}(\gamma(t)) = 1$ . The proof is similar for the case where  $\gamma$  is of the form  $\neg\exists xA$ .

(iv): Suppose  $\delta$  is of the form  $\neg\forall xA$ . Let  $\theta$  be any  $\mathfrak{M}$ -assignment. Since  $\mathfrak{M}(\delta) = 1$ , then  $\mathfrak{M}\theta(\neg\forall xA) = 1$ , and therefore  $\mathfrak{M}\theta(\forall xA) = 0$ . Hence there exists a  $\mu \in M$  such that  $\mathfrak{M}\theta_\mu^x(A) = 0$ . Consequently,  $\mathfrak{M}\theta_\mu^x(\neg A) = 1$ . Let  $N = \{\mu: \mathfrak{M}\theta_\mu^x(\neg A) = 1\}$ . Since  $N \neq \emptyset$ ,  $\Phi(N) \in N$ . But  $\mathfrak{M}\theta(\varepsilon x\neg A) = \Phi(N)$ . Therefore,  $\mathfrak{M}\theta_{\mathfrak{M}\theta(\varepsilon x\neg A)}^x(\neg A) = 1$ , and by Theorem I.4  $\mathfrak{M}\theta(\neg A(\varepsilon x\neg A)) = 1$ . Consequently,  $\mathfrak{M}(\delta(\varepsilon\delta)) = 1$ . The proof is similar for the case where  $\delta$  is of the form  $\exists xA$ .

(v): Let  $\theta$  be any  $\mathfrak{M}$ -assignment. Since  $\mathfrak{M}(\forall x(A \leftrightarrow B)) = 1$ , then for all  $\mu \in M$ ,  $\mathfrak{M}\theta_\mu^x(A \leftrightarrow B) = 1$ . Hence by (ii), for all  $\mu \in M$ ,  $\mathfrak{M}\theta_\mu^x(A) = \mathfrak{M}\theta_\mu^x(B)$ . Thus  $\{\mu: \mathfrak{M}\theta_\mu^x(A) = 1\} = \{\mu: \mathfrak{M}\theta_\mu^x(B) = 1\}$ , and it follows that  $\mathfrak{M}(\varepsilon xA) = \mathfrak{M}(\varepsilon xB)$ .

(vi): If  $x$  and  $y$  are the same variable, the proof is trivial. Suppose  $x$  and  $y$  are distinct variables. Since  $\varepsilon xA$  is a term, then  $y \notin V(A)$ . Furthermore, by our convention (p. 13),  $y$  is free for  $x$  in  $A$ . Let  $N_1 = \{\mu: \mathfrak{M}\theta_\mu^x(A) = 1\}$  and let  $N_2 = \{\mu: \mathfrak{M}\theta_\mu^y(A_s^x) = 1\}$ . It will be sufficient to prove  $N_1 = N_2$ . Since  $\theta_\mu^y(y) = \mu$ , then by Theorem I.4

$$\begin{aligned} \mathfrak{M}\theta_\mu^y(A_s^x) &= \mathfrak{M}\theta_\mu^y \stackrel{x}{\mu}(A), \\ &= \mathfrak{M}\theta_\mu^x \stackrel{y}{\mu}(A), \\ &= \mathfrak{M}\theta_\mu^x(A) \quad \text{since } y \notin V(A). \end{aligned}$$

Consequently  $N_1 = N_2$ .

(vii): The proof follows immediately from Theorem I.4.

### EXERCISES

1. Prove that for any formula  $A$ , if  $\mathfrak{M}$  is adequate for  $A$  and  $A$  is a tautology, then  $\mathfrak{M}(A) = 1$ .
2. Prove that, if  $\mathfrak{M}(A) = 1$  and  $\mathfrak{M}(A \rightarrow B) = 1$ , then  $\mathfrak{M}(B) = 1$ .

### 3.4 Satisfiability and the semantic consequence relation

Let  $A$  be a formula of some language  $\mathcal{L}$  and  $\mathfrak{M}$  any model. We say  $\mathfrak{M}$  *satisfies*  $A$ , or  $A$  is *true* in  $\mathfrak{M}$ , if  $\mathfrak{M}$  is adequate for  $A$  and  $\mathfrak{M}(A) = 1$ . We write  $\mathfrak{M} \text{ Sat } A$  to denote that  $\mathfrak{M}$  satisfies  $A$ . If  $X$  is any set of formulae, we say  $\mathfrak{M}$  *satisfies*  $X$ , denoted by  $\mathfrak{M} \text{ Sat } X$ , if  $\mathfrak{M}$  satisfies all the members of  $X$ . We say  $X$  is *satisfiable* or *has a model*, denoted by  $\text{Sat } X$ , if there exists a model which satisfies  $X$ , and we say  $X$  is *m-satisfiable*, denoted by  $\text{m-Sat } X$ , if there exists a model of cardinality  $m$  which satisfies  $X$ . For any set of formulae  $X$  and any formula  $A$ , we say  $A$  is a *semantic consequence* of  $X$ , denoted by  $X \vDash A$ , if every model which is adequate for  $A$  and which satisfies  $X$  also satisfies  $A$ . Finally  $A$  is *valid*, denoted by  $\vDash A$ , if  $A$  is a semantic consequence of the null set. It follows that  $A$  is valid if and only if every model which is adequate for  $A$  satisfies  $A$ .

For any language  $\mathcal{L}$ , we define the *finitary semantic closure* for  $\mathcal{L}$  as the function, denoted by  $C_s$ , such that for every  $X \subseteq F_{\mathcal{L}}$ :

$$C_s(X) = \{A : A \in F_{\mathcal{L}} \text{ and there exists a finite subset } Y \text{ of } X \text{ such that } Y \vDash A\}.$$

Clearly, if  $A \in C_s(X)$ , then  $X \vDash A$ . For finite  $X$ , the converse is also true. The fact that the converse is true for infinite  $X$  will follow from the Compactness Theorem (p. 29).

For convenience of notation we shall write  $C_s(X, A)$  instead of  $C_s(X \cup \{A\})$ , where  $X$  is a set of formulae and  $A$  is a formula. The following theorem follows readily from the definition of  $C_s$ .

**THEOREM I.6.** For any  $X \subseteq F_{\mathcal{L}}$ :

- (i)  $X \subseteq C_s(X)$ ;
- (ii)  $C_s(C_s(X)) \subseteq C_s(X)$ ;
- (iii) if  $Y \subseteq X$ , then  $C_s(Y) \subseteq C_s(X)$ ;
- (iv) if  $A \in X$ , then there exists a finite  $Y \subseteq X$  such that  $A \in C_s(Y)$ .

The proof is left as an exercise.

**THEOREM I.7.** If  $f \in C_s(X)$ , then there exists a finite subset  $Y$  of  $X$  which is unsatisfiable.

*Proof.* Since  $f \in C_s(X)$ , then there exists some finite subset  $Y$  of  $X$  such that  $Y \vDash f$ . But for any model  $\mathfrak{M}$ ,  $\mathfrak{M}(f) = 0$ . Hence there is no model which satisfies  $Y$ .

In view of Theorem I.7, to prove the Compactness Theorem it will be sufficient to prove that for any  $X \subseteq F_{\mathcal{L}}$ :

- (1) if  $f \notin C_s(X)$ , then  $\text{m-Sat } X$ , for some  $m \leq \overline{\mathcal{L}}$ .

This assertion follows as a special case from the main theorem to be proved in §4, Theorem I.11. The following theorem will be used in connection with that result.

**THEOREM I.8.** Let  $C_s$  be the finitary semantic closure for some language  $\mathcal{L}$ . Then for any  $X \subseteq F_{\mathcal{L}}$ , any  $A \in F_{\mathcal{L}}$ , any  $\alpha, \gamma, \delta \in F_{\mathcal{L}}$ , and any  $t \in T_{\mathcal{L}}$ :

- (i)  $A \in C_s(X)$  iff  $f \in C_s(X, \neg A)$ ;
- (ii)  $\alpha \in C_s(X)$  iff  $\alpha_1, \alpha_2 \in C_s(X)$ ;
- (iii) if  $\gamma \in X$ , then  $\gamma(t) \in C_s(X)$ ;
- (iv) if  $\delta \in X$ , then  $\delta(\varepsilon\delta) \in C_s(X)$ ;
- (v) if  $(s = t) \in X$  and  $B_s^x \in X$ , then  $B_t^x \in C_s(X)$ ;
- (vi) for any  $\varepsilon$ -terms  $\varepsilon xA$  and  $\varepsilon yB$  of  $\mathcal{L}$ , if  $\forall z(A_z^x \leftrightarrow B_z^y) \in X$ , then  $(\varepsilon xA = \varepsilon yB) \in C_s(X)$ .

*Proof.* (i): Assume  $A \in C_s(X)$ . Then  $Y \vDash A$  for some finite  $Y \subseteq X$ . Let  $Z = Y \cup \{\neg A\}$ . Clearly,  $Z$  is unsatisfiable; hence,  $Z \vDash f$ . Since  $Z \subseteq X \cup \{\neg A\}$ , then  $f \in C_s(X, \neg A)$ . Conversely, assume that  $f \in C_s(X, \neg A)$ . By Theorem I.7, there exists a finite subset  $Y$  of  $X \cup \{\neg A\}$  which is unsatisfiable. Let  $Z = Y \setminus \{\neg A\}$ . It is enough to prove that  $Z \vDash A$ . Assume the contrary, i.e. assume that there exists a model  $\mathfrak{M}$  such that  $\mathfrak{M} \text{ Sat } Z$ , but  $\mathfrak{M}(A) = 0$ . Then  $\mathfrak{M}(\neg A) = 1$  and  $\mathfrak{M} \text{ Sat } Y$ , which is impossible.

The proofs of (ii)–(vi) follow by Theorem I.5.

## 4 The Satisfiability Theorem

In this section we prove our main result, the Satisfiability Theorem, from which the Compactness Theorem follows as a special case. This abstract result is obtained by introducing the general notion of a logical closure  $C$  and then proving that for any set  $X$  of formulae of  $\mathcal{L}$ :

$$\text{if } f \notin C(X), \text{ then } \text{m-Sat } X, \text{ for some } m \leq \overline{\mathcal{L}}.$$

We first define the familiar notion of a *finitary closure operation*  $C$  on any set  $S$  as a function from the power set of  $S$  into itself such that, for any  $X \subseteq S$ , the following conditions hold:

- C1.  $X \subseteq C(X)$ ;
- C2.  $C(C(X)) \subseteq C(X)$ ;
- C3. if  $Y \subseteq X$ , then  $C(Y) \subseteq C(X)$ ;
- C4. if  $A \in C(X)$ , then there exists a finite  $Y \subseteq X$  such that  $A \in C(Y)$ .

The notion of a finitary closure operation, which is fundamental to both algebra and logic, was apparently first applied to logic by Tarski [1930].

If  $\mathcal{L}$  is a language and  $C$  is a finitary closure operation on  $F_{\mathcal{L}}$ , then for any  $X \subseteq F_{\mathcal{L}}$ , we say  $X$  is *maximal under*  $C$ , denoted by  $\text{Max}_C X$ , if:



1.  $f \notin C(X)$ ;
2. for all  $Y \subseteq F_{\mathcal{L}}$ , if  $X \subseteq Y$  and  $f \notin C(Y)$ , then  $X = Y$ .

**THEOREM I.9.** *If  $C$  is a finitary closure operation on  $F_{\mathcal{L}}$  and  $Max_C X$ , then  $C(X) = X$ .*

*Proof.* By C1,  $X \subseteq C(X)$ . Since  $Max_C X$ , then  $f \notin C(X)$ , and hence, by C2,  $f \notin C(C(X))$ . But this implies  $X = C(X)$  by the maximality of  $X$ .

**THEOREM I.10.** *If  $f \notin C(X)$ , then there exists some  $Y \subseteq F_{\mathcal{L}}$ , such that  $X \subseteq Y$  and  $Max_C Y$ .*

*Proof.* Let  $\mathfrak{S} = \{X' : X \subseteq X', X' \subseteq F_{\mathcal{L}}, \text{ and } f \notin C(X')\}$ . We shall use Zorn's lemma to prove that  $\mathfrak{S}$  contains maximal elements. Let  $\mathfrak{I}$  be any subset of  $\mathfrak{S}$  which is totally ordered by inclusion, and let  $Y = \bigcup_{X' \in \mathfrak{I}} X'$ . In order to apply

Zorn's lemma we need only show that  $Y \in \mathfrak{S}$ . Clearly  $Y \subseteq F_{\mathcal{L}}$  and  $X \subseteq Y$ . To prove  $f \notin C(Y)$  we assume the contrary and produce a contradiction. If  $f \in C(Y)$ , then by C4,  $f \in C(Y_0)$  for some finite subset  $Y_0$  of  $Y$ . Since  $Y_0$  is finite, then  $Y_0 \subseteq X'$  for some  $X' \in \mathfrak{I}$ , and by C3,  $C(Y_0) \subseteq C(X')$ . Hence  $f \in C(X')$ , which contradicts the definition of  $\mathfrak{I}$ . This completes the proof.

*Note:* If the vocabulary of  $\mathcal{L}$  is finite or denumerable, then Zorn's lemma (and the axiom of choice) can be avoided in the usual way by enumerating the formulae in  $F_{\mathcal{L}}$ .

Let  $\mathcal{L}$  be any formal language. A finitary closure operation  $C$  on  $F_{\mathcal{L}}$  which possesses the following six properties is called a *logical closure* for  $\mathcal{L}$ : For any  $X \subseteq F_{\mathcal{L}}$ , any  $A, \alpha, \gamma$ , and  $\delta \in F_{\mathcal{L}}$ , and any  $s$  and  $t \in T_{\mathcal{L}}$ :

- L1.  $A \in C(X)$  iff  $f \in C(X, \neg A)$ .
- L2.  $\alpha \in C(X)$  iff  $\alpha_1, \alpha_2 \in C(X)$ .
- L3. If  $\gamma \in X$ , then  $\gamma(t) \in C(X)$ .
- L4. If  $\delta \in X$ , then  $\delta(\varepsilon\delta) \in C(X)$ .
- L5. For any atom  $B$  and any variable  $x$  which does not occur free in  $B$  within the scope of an  $\varepsilon$ -symbol, if  $s = t \in X$  and  $B_s^x \in X$ , then  $B_t^x \in C(X)$ .
- L6. For any  $\varepsilon$ -terms  $\varepsilon xA$  and  $\varepsilon yB$ , if  $\forall z(A_z^x \leftrightarrow B_z^y) \in X$ , then  $(\varepsilon xA = \varepsilon yB) \in C(X)$ .

For example, by Theorems I.6 and I.8, the finitary semantic closure,  $C_s$ , is a logical closure.

**THEOREM I.11 (The Satisfiability Theorem).** *Let  $C$  be a logical closure for a language  $\mathcal{L}$ . For any  $X \subseteq F_{\mathcal{L}}$ , if  $f \notin C(X)$ , then  $m\text{-Sat } X$ , for some  $m \leq \overline{\mathcal{L}}$ .*

By Theorem I.10, if  $f \notin C(X)$ , then  $X$  is contained in some set  $Y$  which is

maximal under  $C$ . Also, if  $X \subseteq Y$ , then  $m\text{-Sat } Y$  implies  $m\text{-Sat } X$ . Consequently, to prove Theorem I.11, it will be sufficient to prove that:

$Max_C X$  implies  $m\text{-Sat } X$ , for some  $m \leq \overline{\mathcal{L}}$ .

Throughout the following lemmas and definitions it is assumed that  $X$  is some fixed set which is maximal under the logical closure  $C$ . Thus, by Theorem I.9,  $X$  is closed, i.e.  $C(X) = X$ . The following lemmas and definitions will enable us to construct the required model. The basic idea of this proof is in the spirit of Henkin's proof of the completeness of the predicate calculus (cf. Henkin [1949]), since the notion of a maximal set under  $C$  is a generalization of his notion of a 'maximal consistent set'. Although the presence of the  $\varepsilon$ -symbol has a complicating effect in some respects, in other respects this symbol simplifies the proof since the availability of  $\varepsilon$ -terms makes it unnecessary to adjoin new constant symbols to the language  $\mathcal{L}$ , as is done by Henkin.

Some of the techniques used here are adaptations of methods used by Hermes [1965] in his proof of the completeness of his *Termlogik mit Auswahloperator*. The proof given here is an improvement on Hermes' result, however, since by its abstract nature our proof does not depend on a particular set of logical axioms and rules of inference. Furthermore, Hermes' result holds only for languages with a denumerable set of symbols, whereas our result holds for languages of arbitrary infinite cardinality.

**LEMMA 1.**

- (i)  $A \in X$  iff  $\neg A \notin X$ ;
- (ii)  $A \rightarrow B \in X$  iff  $A \notin X$  or  $B \in X$ ;
- (iii)  $A \vee B \in X$  iff  $A \in X$  or  $B \in X$ ;
- (iv)  $A \wedge B \in X$  iff  $A \in X$  and  $B \in X$ ;
- (v)  $A \leftrightarrow B \in X$  iff  $(A \in X \text{ iff } B \in X)$ ;
- (vi)  $\exists xA \in X$  iff  $A(t) \in X$ , for some  $t \in T_{\mathcal{L}}$ ;
- (vii)  $\forall xA \in X$  iff  $A(t) \in X$ , for all  $t \in T_{\mathcal{L}}$ ;
- (viii) if  $A(\varepsilon x \neg A) \in X$ , then  $\forall xA \in X$ ;
- (ix)  $(t = t) \in X$ , for all  $t \in T_{\mathcal{L}}$ .

*Proof.*

- (i):  $A \in X$  iff  $f \in C(X, \neg A)$  by L1,  
iff  $\neg A \notin X$  by  $Max_C X$ .
- (ii):  $A \rightarrow B \notin X$  iff  $\neg(A \rightarrow B) \in X$  by (i),  
iff  $A \in X$  and  $\neg B \in X$  by L2,  
iff  $A \in X$  and  $B \notin X$  by (i).

(iii): The proof is similar to that of (ii).

(iv): The proof follows immediately by L2.

- (v): The proof is immediate by the definition of  $\leftrightarrow$ , using (ii) and (iv).  
 (vi): Assume  $\exists xA \in X$ . By L4,  $A(\varepsilon xA) \in X$ . Hence there exists a  $t \in T_{\mathcal{L}}$ , such that  $A(t) \in X$ . Conversely, assume  $A(t) \in X$ , for some  $t \in T_{\mathcal{L}}$ . By (i),  $\neg A(t) \notin X$ . Consequently, by (the contrapositive of) L3,  $\neg \exists xA \notin X$ . Hence, by (i) again,  $\exists xA \in X$ .  
 (vii): The proof is similar to that of (vi).  
 (viii): If  $A(\varepsilon x \neg A) \in X$ , then by (i),  $\neg A(\varepsilon x \neg A) \notin X$ . Consequently, by (the contrapositive of) L4,  $\neg \forall xA \notin X$ . Hence by (i),  $\forall xA \in X$ .  
 (ix): Let  $t_0$  be the term  $\varepsilon x \neg (\neg(x = x) \leftrightarrow \neg(x = x))$ . By (v) we have  $\neg(t_0 = t_0) \leftrightarrow \neg(t_0 = t_0) \in X$ . Thus, by (viii) and the definition of  $t_0$ ,  $\forall x(\neg(x = x) \leftrightarrow \neg(x = x)) \in X$ . So, by L6,  $(\varepsilon x \neg(x = x) = \varepsilon x \neg(x = x)) \in X$ . Again by (viii),  $\forall x(x = x) \in X$ . Consequently, by (vii), for any  $t \in T_{\mathcal{L}}$ ,  $(t = t) \in X$ .

We now define a binary relation  $\sim$  on  $T_{\mathcal{L}}$  as follows: for any  $s, t, \in T_{\mathcal{L}}$ ,  $s \sim t$  iff  $(s = t) \in X$ .

LEMMA 2. The relation  $\sim$  is an equivalence relation on  $T_{\mathcal{L}}$ .

- Proof.* (i): That  $\sim$  is reflexive follows from Lemma 1(ix).  
 (ii): The symmetry of  $\sim$  is proved as follows. Assume  $s \sim t$ . Thus  $(s = t) \in X$ . Let  $B$  be the quasi-formula  $(x = s)$ . Thus  $B_s^x \in X$ , by Lemma 1(ix). Hence, by L5,  $B_t^x \in X$ , i.e.  $(t = s) \in X$ . Therefore,  $t \sim s$ .  
 (iii): To prove that  $\sim$  is transitive, assume  $r \sim s$  and  $s \sim t$ . Thus  $(s = t) \in X$ , and if  $B$  is the quasi-formula  $(r = x)$ , then  $B_s^x \in X$ . Thus, by L5,  $B_t^x \in X$ , i.e.  $(r = t) \in X$ , and so  $r \sim t$ .

LEMMA 3. If  $g^n$  and  $P^n$  are  $n$ -place function and predicate symbols, respectively, of  $\mathcal{L}$ , and  $s_i \sim t_i$  for each  $i = 1, \dots, n$ , then

- (i)  $g^n s_1 \dots s_n \sim g^n t_1 \dots t_n$ ,  
 (ii)  $P^n s_1 \dots s_n \in X$  iff  $P^n t_1 \dots t_n \in X$ .

*Proof.* (i): For each  $i = 1, \dots, n$ , let  $A_i$  be the quasi-formula

$$(g^n s_1 \dots s_n = g^n t_1 \dots t_{i-1} x s_{i+1} \dots s_n).$$

Thus  $A_1 \frac{x}{s_1}$  is the formula

$$(g^n s_1 \dots s_n = g^n s_1 \dots s_n).$$

Consequently, by Lemma 1(ix),  $A_1 \frac{x}{s_1} \in X$ , and by L5,  $A_1 \frac{x}{t_1} \in X$ . However, for each  $i = 1, \dots, n$ ,  $A_i \frac{x}{t_i}$  and  $A_{i+1} \frac{x}{s_{i+1}}$  are the same formula. Therefore, by  $n$  applications of L5, we obtain  $A_n \frac{x}{t_n} \in X$ , i.e.  $(g^n s_1 \dots s_n = g^n t_1 \dots t_n) \in X$ , and so  $g^n s_1 \dots s_n \sim g^n t_1 \dots t_n$ .

(ii): The proof is similar to that of (i) except that we let  $A_i$  be the quasi-formula  $P t_1 \dots t_{i-1} x s_{i+1} \dots s_n$ .

We now let  $|t|$  denote the equivalence class of any term  $t$ , and let  $M$  denote the set of equivalence classes on  $T_{\mathcal{L}}$ . The following definition is due to Hermes [1965].

For any subset  $N$  of  $M$ , we say  $N$  is *representable* if there exists a quasi-formula  $A$  of  $\mathcal{L}$  and a variable  $x$ , such that for all  $t \in T_{\mathcal{L}}$ ,  $|t| \in N$  iff  $A_t^x \in X$ . In this case we say  $\langle A, x \rangle$  is a *representative* of  $N$ . Clearly, if  $\langle A, x \rangle$  represents  $N$ , then no variable other than  $x$  occurs free in  $A$ , and therefore  $\varepsilon x A$  is a term.

LEMMA 4. If  $\langle A, x \rangle$  and  $\langle B, y \rangle$  are both representatives of  $N$ , then  $\varepsilon x A \sim \varepsilon y B$ .

*Proof.* Let  $z$  be any variable not appearing (either free or bound) in  $A$  or  $B$ . Let  $A'$  be  $A_z^x$  and  $B'$  be  $B_z^y$ . Thus for any term  $t$ ,  $A'_t^z$  is  $A_t^x$ , and  $B'_t^z$  is  $B_t^y$ . Consequently  $\langle A', z \rangle$  and  $\langle B', z \rangle$  are representatives of  $N$ . Let  $t_0$  be the term  $\varepsilon z \neg(A' \leftrightarrow B')$ . Since  $\langle A', z \rangle$  and  $\langle B', z \rangle$  represent  $N$ , then  $|t_0| \in N$  iff  $A'_{t_0}^z \in X$  and  $|t_0| \in N$  iff  $B'_{t_0}^z \in X$ . Thus  $A'_{t_0}^z \in X$  iff  $B'_{t_0}^z \in X$ , and by Lemma 1(v),  $A'_{t_0}^z \leftrightarrow B'_{t_0}^z \in X$ . By Lemma 1(viii) and our choice of  $t_0$ , this implies that  $\forall z(A' \leftrightarrow B') \in X$ . Consequently, by L6,  $(\varepsilon x A = \varepsilon y B) \in X$ , and  $\varepsilon x A \sim \varepsilon y B$ .

LEMMA 5. If  $\langle A, x \rangle$  is a representative of a non-empty set  $N$ , then  $|\varepsilon x A| \in N$ .

*Proof.* Since  $N$  is non-empty and  $\langle A, x \rangle$  represents  $N$ , then there exists a term  $t$  such that  $A(t) \in X$ . Thus by Lemma 1(vi),  $\exists x A \in X$ , and by L4,  $A(\varepsilon x A) \in X$ . Hence  $|\varepsilon x A| \in N$ .

The model  $\mathfrak{M}$  for the set  $X$  that we need in order to prove Theorem I.11 is now defined in the following way.

1. The universe of  $\mathfrak{M}$  is the set  $M$  above, i.e. the set of equivalence classes on  $T_{\mathcal{L}}$  under the equivalence relation  $\sim$ .
2. The function  $\Pi$  is defined as follows:
  - (i) For any individual symbol  $a$ ,  $\Pi(a) = |a|$ .
  - (ii) For any  $g^n$  in the vocabulary of  $\mathcal{L}$  and any  $|t_1|, \dots, |t_n|$  in  $M$ ,  $\Pi(g^n)(|t_1|, \dots, |t_n|) = |g^n t_1 \dots t_n|$ . By Lemma 3, this definition is independent of the choice of the  $t_i$ .
  - (iii) For any  $P^n$  in the vocabulary of  $\mathcal{L}$  and any  $|t_1|, \dots, |t_n|$  in  $M$ ,  $\Pi(P^n)(|t_1|, \dots, |t_n|) = 1$  iff  $P^n t_1 \dots t_n \in X$ . Again by Lemma 3, this definition is independent of the choice of the  $t_i$ .
3. The choice function  $\Phi$  is defined as follows for any  $N \subseteq M$ .

*Case 1:* If  $N$  is representable,  $\Phi(N) = |\varepsilon x A|$ , where  $\langle A, x \rangle$  is one of its representatives. By Lemma 4, this definition is independent of the choice of representative.

*Case 2:* If  $N$  is not representable, (and hence  $N \neq \emptyset$ , since the null set is represented by  $\langle \neg(x = x), x \rangle$ ), then  $\Phi(N)$  is an arbitrary member of  $N$ .

By Lemma 5, if  $N \neq \emptyset$ , then  $\Phi(N) \in N$ . Thus under this definition  $\Phi$  is a choice function on  $M$ .

Under these definitions of  $M$ ,  $\Pi$ , and  $\Phi$ , the triple  $\langle M, \Pi, \Phi \rangle$ , which we shall denote by  $\mathfrak{M}$ , is a model for the vocabulary of  $\mathcal{L}$ . Furthermore, since its universe,  $M$ , is  $T_{\mathcal{L}}/\sim$ , then  $\overline{M} \leq \overline{T}_{\mathcal{L}}$ . Therefore the cardinality of the model  $\mathfrak{M}$  is less than or equal to the cardinality of the language  $\mathcal{L}$  (see page 16). All that remains to prove is that  $\mathfrak{M} \text{ Sat } X$ .

Let  $\theta$  be any  $\mathfrak{M}$ -assignment and  $\Sigma$  any replacement operator for the language  $\mathcal{L}$ . Since for any  $x \in \text{dom } \Sigma$ ,  $\Sigma(x) \in T_{\mathcal{L}}$ , then  $|\Sigma(x)|$  is defined for  $x \in \text{dom } \Sigma$ . Let  $V$  be any subset of the set of variables,  $\forall r$ . We write  $\theta \simeq \Sigma$  to denote that  $V \subseteq \text{dom } \Sigma$  and for all  $x \in V$ ,  $\theta(x) = |\Sigma(x)|$ .

LEMMA 6. Let  $A$  be any well-formed expression of  $\mathcal{L}$ . If  $\theta$  is any  $\mathfrak{M}$ -assignment and  $\Sigma$  is any replacement operator such that  $\theta \simeq \Sigma$ , then:

- (i)  $\mathfrak{M}\theta(A) = |A\Sigma|$ , if  $A$  is a quasi-term;
- (ii)  $\mathfrak{M}\theta(A) = 1$  iff  $A\Sigma \in X$ , if  $A$  is a quasi-formula.

*Proof.* The proof is by induction on the length of  $A$ .

G1.  $A$  is a variable  $x$ :  $\mathfrak{M}\theta(x) = \theta(x) = |\Sigma(x)| = |x\Sigma|$ ;  
 $A$  is an individual symbol  $a$ :  $\mathfrak{M}\theta(a) = \Pi(a) = |a| = |a\Sigma|$ .

G2.  $A$  is of the form  $g^n t_1 \dots t_n$ : Since for each  $i = 1, \dots, n$ ,  $V(t_i) \subseteq V(A)$ , then  $\theta \simeq \Sigma$ . Therefore, by the induction hypothesis,  $\mathfrak{M}\theta(t_i) = |t_i\Sigma|$ . Now,

$$\begin{aligned} \mathfrak{M}\theta(g^n t_1 \dots t_n) &= \Pi(g^n)(\mathfrak{M}\theta(t_1), \dots, \mathfrak{M}\theta(t_n)), \\ &= \Pi(g^n)(|t_1\Sigma|, \dots, |t_n\Sigma|) \text{ by induction hypothesis,} \\ &= |g^n[t_1]\Sigma \dots [t_n]\Sigma| \text{ by the definition of } \Pi, \\ &= |[g^n t_1 \dots t_n]\Sigma|. \end{aligned}$$

G3.  $A$  is of the form  $P^n t_1 \dots t_n$ :

$$\begin{aligned} \mathfrak{M}\theta(P^n t_1 \dots t_n) = 1 &\text{ iff } \Pi(P^n)(\mathfrak{M}\theta(t_1), \dots, \mathfrak{M}\theta(t_n)) = 1, \\ &\text{ iff } \Pi(P^n)(|t_1\Sigma|, \dots, |t_n\Sigma|) = 1 \text{ by induction} \\ &\quad \text{hypothesis,} \\ &\text{ iff } P^n[t_1]\Sigma \dots [t_n]\Sigma \in X \text{ by the definition of } \Pi, \\ &\text{ iff } [P^n t_1 \dots t_n]\Sigma \in X. \end{aligned}$$

G4.  $A$  is of the form  $(s = t)$ :

$$\begin{aligned} \mathfrak{M}\theta(s = t) = 1 &\text{ iff } \mathfrak{M}\theta(s) = \mathfrak{M}\theta(t) \\ &\text{ iff } |s\Sigma| = |t\Sigma| \text{ by induction hypothesis,} \\ &\text{ iff } (s\Sigma = t\Sigma) \in X \text{ by the definition of } \sim, \\ &\text{ iff } [(s = t)]\Sigma \in X. \end{aligned}$$

G5.  $A$  is the formula  $f$ :  $\mathfrak{M}\theta(f) = 0$ , and by  $\text{Max}_C X$ ,  $f \notin X$ . Consequently,  $\mathfrak{M}\theta(f) = 1$  iff  $f\Sigma \in X$ , since both sides of this equivalence are false.

G6.  $A$  is of the form  $\neg B$ ,  $(B \rightarrow C)$ ,  $(B \vee C)$ , or  $(B \wedge C)$ : The proof follows immediately by Lemma 1 (i)–(iv).

G7–G8.  $A$  is of the form  $\exists x B$ ,  $\forall x B$ ,  $exB$ : Let  $\Sigma'$  be the  $x$ -suppression of  $\Sigma$  (see p. 14). Let  $N = \{\mu: \mathfrak{M}\theta_{\mu}^x(B) = 1\}$ . Since  $\theta \simeq \Sigma$ , then for any  $t \in T_{\mathcal{L}}$   $\theta_{|t|}^x \simeq \Sigma'_{|t|}$ . We first prove that (1) for all  $t \in T_{\mathcal{L}}$ ,  $|t| \in N$  iff  $[B\Sigma']_{|t|}^x \in X$ .

For any  $t \in T_{\mathcal{L}}$ ,  $|t| \in N$  iff  $\mathfrak{M}\theta_{|t|}^x(B) = 1$  by definition of  $N$ ,  
iff  $[B]\Sigma'_{|t|}^x \in X$  by induction hypothesis,  
iff  $[B\Sigma']_{|t|}^x \in X$  Theorem I.2(ii) (p. 14).

Consequently  $\langle B\Sigma', x \rangle$  is a representative of  $N$ , as required.

Case 1.  $A$  is of the form  $\exists x B$ :

$$\begin{aligned} \mathfrak{M}\theta(\exists x B) = 1 &\text{ iff } N \neq \emptyset \\ &\text{ iff there exists a } t \in T_{\mathcal{L}} \text{ such that } [B\Sigma']_{|t|}^x \in X, \text{ by (1),} \\ &\text{ iff } \exists x [B]\Sigma' \in X \text{ by Lemma 1(vi),} \\ &\text{ iff } [\exists x B]\Sigma \in X \text{ by Theorem I.2(iii) (p. 14).} \end{aligned}$$

Case 2.  $A$  is of the form  $\forall x B$ :

$$\begin{aligned} \mathfrak{M}\theta(\forall x B) = 1 &\text{ iff } N = M, \\ &\text{ iff for all } t \in T_{\mathcal{L}}, [B\Sigma']_{|t|}^x \in X, \text{ by (1),} \\ &\text{ iff } \forall x [B]\Sigma' \in X \text{ by Lemma 1(vii),} \\ &\text{ iff } [\forall x B]\Sigma \in X \text{ by Theorem I.2(iii).} \end{aligned}$$

Case 3.  $A$  is of the form  $exB$ :

$$\begin{aligned} \mathfrak{M}\theta(exB) &= \Phi(N) \\ &= |ex[B]\Sigma'| \text{ by definition of } \Phi, \text{ since } \langle B\Sigma', x \rangle \text{ represents } N, \\ &= |[exB]\Sigma| \text{ by Theorem I.2(iii).} \end{aligned}$$

This completes the proof of Lemma 6.

Using Lemma 6, we now prove that for any  $A \in F_{\mathcal{L}}$ ,  $\mathfrak{M}(A) = 1$  iff  $A \in X$ . Take any  $\mathfrak{M}$ -assignment  $\theta$  and any replacement operator  $\Sigma$ . Since  $A$  is a formula, then  $V(A) = \emptyset$ , and therefore  $\theta \simeq \Sigma$ . Thus by Lemma 6,  $\mathfrak{M}\theta(A) = 1$  iff  $A\Sigma \in X$ . But  $A\Sigma$  is  $A$  itself, and  $\mathfrak{M}\theta(A) = \mathfrak{M}(A)$ . Therefore,  $\mathfrak{M}(A) = 1$  iff  $A \in X$ . Consequently,  $\mathfrak{M} \text{ Sat } X$ . This completes the proof of Theorem I.11.

THEOREM I.12 (The Compactness Theorem). If  $X$  is a set of formulae of some language  $\mathcal{L}$  such that every finite subset of  $X$  is satisfiable, then  $X$  is  $m$ -satisfiable, for some  $m \leq \overline{\mathcal{L}}$ .

*Proof.* By Theorems I.6 and I.8,  $C_s$  is a logical closure for  $\mathcal{L}$ . By Theorem I.7, if every finite subset of  $X$  is satisfiable, then  $f \notin C_s(X)$ . The theorem, therefore, follows from Theorem I.11.

**THEOREM I.13** (Löwenheim-Skolem Theorem). *If  $X$  is any satisfiable set of formulae of some language  $\mathcal{L}$ , then  $X$  is  $m$ -satisfiable, for some  $m \leq \overline{\mathcal{L}}$ .*

*Proof.* If  $X$  is satisfiable, then every finite subset of  $X$  is satisfiable. The result then follows from Theorem I.12.

We can strengthen the statements of Theorems I.12 and I.13 by observing that no condition is imposed on the language  $\mathcal{L}$  except that  $X$  is a set of its formulae. Thus we may consider  $\mathcal{L}$  to be the language whose vocabulary consists of just those function symbols and predicate symbols which appear in the members of  $X$ . In this case,  $\overline{\mathcal{L}} = \max \{\aleph_0, \overline{X}\}$ , and we can state the conclusions of Theorems I.12 and I.13 as ' $X$  is  $m$ -satisfiable, for some  $m \leq \max \{\aleph_0, \overline{X}\}$ '.

Consequently, using the abstract notion of a logical closure, we have managed to prove the Compactness Theorem for a formal language  $\mathcal{L}$  without becoming involved in a particular deductive structure for that language, i.e. without choosing some particular set of logical axioms and rules of inference and then first proving the Completeness Theorem. Another advantage of this abstract approach is that when we do impose a particular deductive structure on the language  $\mathcal{L}$ , by specifying certain axioms and rules of inference, we can then prove the completeness of this formal system merely by showing that its deductive closure is a logical closure.

There are other ways of proving the Compactness Theorem without first proving the Completeness Theorem. One such method involves the use of ultraproducts (cf. Frayne, Morel, and Scott [1962]).

### EXERCISES

1. Prove that if  $X \models A$ , then there exists a finite  $Y \subseteq X$  such that  $Y \models A$ .
2. Use the Compactness Theorem to prove that a partial ordering  $R$  on a set  $S$  can be extended to a total ordering on that set. (*Hint:* Let  $\mathcal{V}$  be the vocabulary consisting of the 2-place predicate symbol  $<$  and constant symbols for each member of the set  $S$ . Let  $X$  be the set of formulae of  $\mathcal{L}(\mathcal{V})$  consisting of the axioms for a total ordering (cf. page 86, S1-S3) and all formulae of the form  $c_\alpha < c_\beta$  where  $c_\alpha$  and  $c_\beta$  are constant symbols which correspond to members of  $S$  which are in the relation  $R$ . Prove that every finite subset of  $X$  is satisfiable.)

#### 4.1 Logical closures

We now use Theorem I.11 to prove a few general results about logical closures.

**THEOREM I.14.** *Let  $C$  be any logical closure on  $\mathcal{L}$ . For any  $A \in F_{\mathcal{L}}$  and any  $X \subseteq F_{\mathcal{L}}$ , if  $X \models A$  then  $A \in C(X)$ .*

*Proof.* Assume  $X \models A$ . Then the set  $X \cup \{\neg A\}$  is unsatisfiable. Consequently by Theorem I.11,  $\neg A \in C(X, \neg A)$ . By L1, this implies  $A \in C(X)$ .

**COROLLARY.** *For any  $X \subseteq F_{\mathcal{L}}$  and any  $A \in F_{\mathcal{L}}$ ,  $A \in C_s(X)$  if and only if  $X \models A$ .*

*Proof.* If  $A \in C_s(X)$ , then  $X \models A$  by the definition of  $C_s$ . If  $X \models A$ , then  $A \in C_s(X)$  by Theorem I.14.

It is reasonable to ask the following question. Are there any logical closures for  $\mathcal{L}$  which do not coincide with  $C_s$ , and if so, what are they?

The first half of this question can be answered easily. Suppose that  $X_0$  is any fixed set of formulae in the language  $\mathcal{L}$ . If we define  $C$  by

$$(1) \quad C(X) = C_s(X \cup X_0), \text{ for any } X \subseteq F_{\mathcal{L}},$$

then it is easy to prove that  $C$  is a logical closure. (The details of this proof are left as an exercise. The only slightly difficult part is verifying that  $C$  satisfies property C2.) If the formulae in  $X_0$  are not all logically valid, then  $C$  does not coincide with  $C_s$ .

The interesting fact is that every logical closure for the language  $\mathcal{L}$  has the form (1), as we shall prove in the following theorem.

**THEOREM I.15.** *If  $C$  is a logical closure for  $\mathcal{L}$ , then for every  $X \subseteq F_{\mathcal{L}}$*

$$C(X) = C_s(X \cup C(\emptyset)).$$

We first prove two lemmas.

**LEMMA 1.** *If  $C$  is a logical closure for  $\mathcal{L}$ , and  $\alpha$  is a conjunctive formula of  $\mathcal{L}$ , then for any  $X \subseteq F_{\mathcal{L}}$ ,*

$$C(X, \alpha) = C(X, \alpha_1, \alpha_2).$$

*Proof.* Since  $\alpha \in C(X, \alpha)$ , then by L2,  $\alpha_1, \alpha_2 \in C(X, \alpha)$ . Also  $X \subseteq C(X, \alpha)$ . Hence  $X \cup \{\alpha_1, \alpha_2\} \subseteq C(X, \alpha)$ . Thus, by C2 and C3  $C(X, \alpha_1, \alpha_2) \subseteq C(X, \alpha)$ . Conversely, since  $\alpha_1, \alpha_2 \in C(X, \alpha_1, \alpha_2)$ , then by L2,  $\alpha \in C(X, \alpha_1, \alpha_2)$ . Thus  $X \cup \{\alpha\} \subseteq C(X, \alpha_1, \alpha_2)$ , and  $C(X, \alpha) \subseteq C(X, \alpha_1, \alpha_2)$ .

**LEMMA 2.** *For any logical closure  $C$  for  $\mathcal{L}$ , any  $X \subseteq F_{\mathcal{L}}$ , and any  $A, B \in F_{\mathcal{L}}$ :*

- (i) *if  $B \in C(X, A)$ , then  $A \rightarrow B \in C(X)$ ;*
- (ii) *if  $A \in X$  and  $A \rightarrow B \in C(X)$ , then  $B \in C(X)$ .*

*Proof.* (i): Since  $B \in C(X, A)$ , then  $\neg B \in C(X, A, \neg B)$ , and by Lemma 1;  $\neg B \in C(X, \neg(A \rightarrow B))$ . Hence  $A \rightarrow B \in C(X)$ .

(ii): Since  $A \rightarrow B \in C(X)$ , then  $\neg(A \rightarrow B) \in C(X)$ , and by Lemma 1,  $\neg(A \rightarrow B) \in C(X, A, \neg B)$ . But since  $A \in X$ , then  $\neg B \in C(X, \neg B)$ , and it follows that  $B \in C(X)$ .

To prove Theorem I.15, we want to show that for any  $X \subseteq F_{\mathcal{L}}$ ,

$$C(X) = C_s(X \cup C(\emptyset)).$$

First take any  $A \in C(X)$ . By C4, there exists a finite subset  $Y$  of  $X$  such that  $A \in C(Y)$ . If  $A_1, A_2, \dots, A_n$  are the members of  $Y$ ,  $A \in C(A_1, A_2, \dots, A_n)$ , and by  $n$  applications of Lemma 2(i),

$$A_1 \rightarrow \dots \rightarrow A_n \rightarrow A \in C(\emptyset).$$

By Lemma 2(ii)

$$A_2 \rightarrow \dots \rightarrow A_n \rightarrow A \in C_s(X \cup C(\emptyset)),$$

and by repeated applications of this lemma, we obtain  $A \in C_s(X \cup C(\emptyset))$ . Conversely, take any  $A \in C_s(X \cup C(\emptyset))$ . Thus  $X \cup C(\emptyset) \vDash A$  and by Theorem I.14  $A \in C(X \cup C(\emptyset))$ . But  $X \subseteq C(X)$  and  $C(\emptyset) \subseteq C(X)$ , so that it follows that  $C(X \cup C(\emptyset)) \subseteq C(X)$ . Therefore,  $A \in C(X)$ . This completes the proof of the theorem.

**COROLLARY.** *Let  $C$  be any function from the power set of  $F_{\mathcal{L}}$  into itself. Then  $C$  is a logical closure for  $\mathcal{L}$  if and only if for every  $X \subseteq F_{\mathcal{L}}$*

$$C(X) = C_s(X \cup C(\emptyset)).$$

### 5 Alternative interpretations of the $\varepsilon$ -symbol

In §3.3 we formulated the semantics of a language  $\mathcal{L}$  by interpreting the  $\varepsilon$ -symbol as a choice function  $\Phi$  for a universe  $M$ , such that  $\Phi$  assigns to the empty set some arbitrary but fixed member of  $M$ . We shall now discuss the general problem of finding a suitable interpretation for the  $\varepsilon$ -symbol and the particular solutions to this problem which appear in the literature.

Since Hilbert introduced the  $\varepsilon$ -symbol merely as a formal syntactic device to facilitate proof-theoretic investigations of the predicate calculus and of mathematical theories, such as arithmetic, which are based on the predicate calculus, the status of the  $\varepsilon$ -symbol is somewhat different from that of the other logical primitives. Although the basic methodology of Hilbert's formalist programme is to treat all symbols as meaningless, there is little doubt as to the *intended* interpretation of the symbols  $\bar{\quad}$ ,  $\rightarrow$ ,  $\&$ ,  $\vee$ ,  $\sim$ ,  $E$ , and  $()$ . On the other hand, it is by no means clear what interpretation is intended for the  $\varepsilon$ -symbol, or whether, in fact, any interpretation is intended. Hilbert's informal remarks about this symbol amount to little more than ' $\varepsilon xA$  is some object of the domain of individuals, such that if anything satisfies the formula  $A$ , then  $\varepsilon xA$  does' ([1939], page 12). Hilbert's main concern is with the formal system which is obtained when the  $\varepsilon$ -symbol and the  $\varepsilon$ -formula,

$$A(a) \rightarrow A(\varepsilon xA),$$

are adjoined to the predicate calculus, and the rule of substitution is extended to allow for the replacement of free variables by  $\varepsilon$ -terms. The significance of his  $\varepsilon$ -Theorems is that a given deduction in the predicate calculus can be converted, using the  $\varepsilon$ -calculus, into another deduction of a certain special form in the predicate calculus. Consequently, the question of an interpretation, or even an intended interpretation, of the  $\varepsilon$ -symbol is unimportant both to his methods and to his results. The intended interpretation of the other logical symbols is unimportant only with respect to his methods.

Although Hilbert provides us with no precise semantic interpretation of the  $\varepsilon$ -symbol, his formal system does suggest certain properties which any interpretation must satisfy. First of all, the  $\varepsilon$ -formula and its deductive equivalent

$$(\varepsilon_0) \quad \exists xA \rightarrow A(\varepsilon xA)$$

must be valid under this interpretation. Secondly, the interpretation must assign a value to every term of the form  $\varepsilon xA$ , even when the formula  $A$  is not satisfiable. This second requirement follows from the fact that Hilbert's system includes a rule for the replacement of free variables by arbitrary  $\varepsilon$ -terms. A third requirement would be that the 'second  $\varepsilon$ -axiom schema'

$$(\varepsilon_2) \quad \forall x(A \leftrightarrow B) \rightarrow (\varepsilon xA = \varepsilon xB)$$

should be valid under this interpretation. Although formulae of this form are not taken as axioms in Hilbert's system, schema  $(\varepsilon_2)$  is a standard axiom schema in formalizations of set theory which incorporate the  $\varepsilon$ -symbol. We shall say that an interpretation of the  $\varepsilon$ -symbol is 'suitable' if it satisfies these three conditions.

The idea of using choice functions as an interpretation of this symbol was first investigated by Asser [1957] following a suggestion by Schröter. His investigations deal with three types of choice function. The first type is the one which we have used in §3.3. This interpretation is suitable in view of our results in that section.

Even if one agrees to interpret the  $\varepsilon$ -symbol in terms of a choice function, there are various ways of interpreting a 'null term'—that is, an  $\varepsilon$ -term  $\varepsilon xA$ , where  $A$  is unsatisfiable. Clearly, the interpretation of such a term depends on the entity, if any, which the choice function assigns to the null set. The choice function we have used (Asser's first type) assigns to the null set some arbitrary, but fixed member of the universe  $M$ . Two alternative definitions have been proposed: (i) Asser's second type of choice function, where  $\Phi$  is undefined on the null set; (ii) the choice function employed by Hermes [1965] where  $\Phi$  assigns the same value to the null set as it does to the universe, i.e.

$$(1) \quad \Phi(\emptyset) = \Phi(M).$$

Hermes' definition seems to have been prompted by the way Hilbert and Bernays initially define the  $\varepsilon$ -symbol in terms of the  $\eta$ -symbol. The  $\eta$ -symbol is formally introduced by means of the following  $\eta$ -rule ([1939], page 10).

'If a formula  $\exists xA$  is an axiom or is derivable, then  $\eta xA$  can be introduced as a term, and the formula  $A(\eta xA)$  can be taken as an initial formula', i.e. from  $\exists xA$ , one can infer  $A(\eta xA)$ .

Obviously, the  $\eta$ -symbol represents the 'indefinite article' in the same way that Russell's  $\iota$ -symbol represents the 'definite article'. The  $\varepsilon$ -symbol is then defined as follows (p. 11):

$$(2) \quad \varepsilon xA =_{\text{Df}} \eta x(\exists y[A]_y^x \rightarrow A).$$

From this definition it follows by the  $\eta$ -rule and the predicate calculus that any formula of the form

$$\exists xA \rightarrow A(\varepsilon xA)$$

is derivable. Furthermore, although  $\eta xA$  is a term only if  $\exists xA$  is derivable, any expression of the form  $\varepsilon xA$  is a term. Hilbert and Bernays then dispense with the  $\eta$ -symbol, and instead take the  $\varepsilon$ -symbol as a primitive and introduce the  $\varepsilon$ -formula as an axiom.

Although this method of introducing the  $\varepsilon$ -symbol is only a heuristic device, Asser (p. 65) and presumably Hermes see in it an indication of Hilbert's intended interpretation of a null term. For, suppose there is no  $x$  for which  $A$  holds, then the formula  $\exists y[A]_y^x \rightarrow A$  is true for all  $x$ . Consequently, using a choice function interpretation for both  $\eta$  and  $\varepsilon$ , it would follow from definition (2) that this choice function must assign the same value to the null set as it does to the universe.

If we define a model using Hermes' notion of a choice function, our results still hold subject to the following modifications.

In the definition of a logical closure the following additional condition is required:

$$L7. \quad (\varepsilon x(x = x) = \varepsilon x \neg(x = x)) \in C(X).$$

In the proof of theorem I.11 it is necessary to show that the function  $\Phi$  as defined on page 27 satisfies Hermes' condition

$$\Phi(\emptyset) = \Phi(M).$$

This follows immediately from L7 and the fact that  $\langle(x = x), x\rangle$  represents  $M$  and  $\langle\neg(x = x), x\rangle$  represents  $\emptyset$ . It is not difficult to show that  $C_s$  (under our new definition of a model) satisfies L7. Hence the Compactness Theorem still holds.

The second type of choice function which Asser considers is that which is undefined on the null set. It is clear from his results that this concept of choice function is better suited as an interpretation of the  $\eta$ -symbol than of the  $\varepsilon$ -symbol, since no value is assigned to a null term and thus our second requirement for a 'suitable' interpretation fails.

Asser points out that the  $\varepsilon$ -calculus of Hilbert and Bernays is not complete under his first interpretation of the  $\varepsilon$ -symbol. In an attempt to find an interpretation under which their  $\varepsilon$ -calculus is complete, Asser defines a third type of choice function, which is a very complicated modification of the first.

'Wir werden nun zeigen, daß es tatsächlich möglich ist, den Begriff der Auswahlfunktion so zu fassen, daß die zugehörige Interpretation dem formalen Ansatz von Hilbert adäquat ist. Allerdings ist dieser Begriff von Auswahlfunktion so kompliziert, daß sich seine Verwendung in der inhaltlichen Mathematik kaum empfiehlt.'<sup>1</sup>

Later (p. 65) Asser remarks that in view of the complexity of this third type of choice function, it is unlikely that this was Hilbert's intended interpretation of the  $\varepsilon$ -symbol.

We have chosen to interpret the  $\varepsilon$ -symbol in terms of Asser's first type of choice function for three reasons: (i) this interpretation is intuitively natural and simple to define (as opposed to Asser's third interpretation); (ii) this interpretation satisfies the three requirements we have given for a 'suitable' interpretation; (iii) under this interpretation the  $\varepsilon$ -calculus which is used in formalizing set theory (cf. Ackermann [1937-8] and Bourbaki [1954]) is complete. Although Hermes' interpretation is also 'suitable', we have not used it because it then becomes necessary to adjoin a new axiom, such as

$$\varepsilon x(x = x) = \varepsilon x \neg(x = x),$$

to the axioms of the  $\varepsilon$ -calculus in order to maintain completeness. There does not, in general, seem to be any good reason why such an axiom should be available in the  $\varepsilon$ -calculus.

However, it would be a mistake to state dogmatically that one particular interpretation of the  $\varepsilon$ -symbol is correct and all others are incorrect. One of the advantages of this symbol is its flexibility and indeterminacy. It is always possible, and often advantageous, to adjoin additional  $\varepsilon$ -axioms to a system, thereby making the designations of the  $\varepsilon$ -terms more definite. For example, Hilbert and Bernays have shown (pages 85-87) that by taking the formula (schema)

$$(1) \quad A(i) \rightarrow \varepsilon xA \neq i'$$

<sup>1</sup> Asser (1957), p. 59.

as an axiom schema of arithmetic, the principle of mathematical induction can be derived. (The symbol ' denotes the arithmetic successor function.) In this formalization of arithmetic, the  $\varepsilon$ -symbol can be interpreted as a least number operator, although this is not the only possible interpretation, since (1) still allows the  $\varepsilon$ -symbol a certain amount of indeterminacy. If (1) is replaced by the stronger axiom schema

$$(2) \quad A(t) \rightarrow \varepsilon x A \leq t,$$

then the  $\varepsilon$ -symbol is uniquely characterized as a least number operator (cf. Tait [1965]). We shall return to these applications of the  $\varepsilon$ -symbol to arithmetic in Chapter IV, §3.2.

## 6 Languages without identity

Although this book deals only with languages whose logical constants include the identity symbol =, it is easy to modify the definitions and proofs so that our main results also hold for languages without an identity symbol.

In the definition of a logical closure for a language without identity we replace conditions L5 and L6 by the following:

L8. If  $\forall z(A_1 \overset{x_1}{z} \leftrightarrow A_2 \overset{x_2}{z}) \in X$  and  $B_{\varepsilon x_1 A_1}^y \in X$ , then  $B_{\varepsilon x_2 A_2}^y \in C(X)$ , where  $B$  is any atom and  $y$  any variable which does not have a free occurrence in  $B$  within the scope of an  $\varepsilon$ -symbol.

Theorem I.11 still holds for languages without identity. In the proof a more complicated definition of the equivalence relation  $\sim$  is required. Observe that for any term  $s$ , one of the following conditions must hold: (i)  $s$  is an individual symbol, (ii)  $s$  is an  $\varepsilon$ -term, (iii)  $s$  is of the form  $gs_1 \dots s_n$ , where  $n \geq 0$ . The definition of  $s \sim t$  is by induction on the length of  $s$  as follows.

Case 1.  $s$  is an individual symbol  $a$ : Then  $s \sim t$  iff  $t$  is the same individual symbol.

Case 2.  $s$  is of the form  $\varepsilon x A$ : Then  $s \sim t$  iff  $t$  is of the form  $\varepsilon y B$  and for all  $r \in T_{\mathcal{L}}$ ,  $A_r^x \in X$  iff  $B_r^y \in X$ .

Case 3.  $s$  is of the form  $gs_1 \dots s_n$ : Then  $s \sim t$  iff  $t$  is of the form  $gt_1 \dots t_n$  and  $s_i \sim t_i$ , for each  $i = 1, \dots, n$ .

Lemma 1, that  $\sim$  is an equivalence relation, follows easily from the definition of  $\sim$ . Lemma 2 follows by L8, and Lemma 3 is immediate from the definitions. The rest of the proof of Theorem I.11 goes through unchanged.

## CHAPTER II

### FORMAL SYSTEMS

#### 1.1 Finitary reasoning

In Chapter I we formalized the intuitive notion of logical consequence by using the non-constructive techniques of set theory to define the semantic consequence relation  $\vDash$ . Although this approach has a certain abstract mathematical appeal, it is important, particularly in proving the consistency of mathematical theories, to find a more concrete definition of logical consequence. For this reason we now turn to the notion of deducibility in a formal system.

A formal system may be regarded as an array of uninterpreted symbols together with rules for manipulating these symbols. Consequently, in dealing with formal systems we can use a much weaker metatheory than that which was used in dealing with models. Throughout the present chapter and succeeding chapters nearly all our metatheoretic arguments will fall within the domain of what Hilbert calls *finitary reasoning* (*das finite Schließen*). Hilbert defines this type of reasoning as follows: (Hilbert and Bernays [1934], page 32, translation by Kneebone [1963] page 205):

'We shall always use the word "finitary" to indicate that the discussion, assertion, or definition in question is kept within the bounds of thorough-going producibility of objects and thorough-going practicability of processes, and may accordingly be carried out within the domain of concrete inspection.'

In other words our discussions will deal with concrete objects such as terms, formulae, and finite sequences of formulae. In order to prove that a certain concrete object exists we must exhibit that object or at least describe a procedure for finding or constructing such an object. For example a proof of a metalinguistic statement of the form 'for all  $x$ , there exists a  $y$  such that . . . ' is a finitary proof if it enables one to construct an appropriate  $y$  for any given  $x$ .

Throughout this chapter we shall use the basic facts about the natural numbers and in particular the principle of mathematical induction. When we do so our metatheory remains finitary, since a natural number may be regarded as a sequence of vertical strokes and if we have proved by induction that every natural number has a certain property, then for any given number  $n$ , the inductive proof provides a method for showing in a finite number of steps that  $n$  has that property. (See Hilbert and Bernays [1934].)

We shall continue to use the *language* of set theory as part of our meta-language. Thus, for example, we will speak of 'sets' of formulae. This use of set theoretic terminology is only a matter of convenience, and nearly all references to sets could be eliminated.

The only theorems of this chapter which are not proved by finitary techniques are Theorems II.1 and II.11 which deal with the completeness of formal systems. In these two cases a finitary proof is impossible since the very notion of completeness depends on the non-finitary notion of semantic consequence. Similarly, Exercise 3 at the end of §2, which deals with the soundness of the  $\varepsilon$ -calculus, cannot be proved by finitary reasoning.

## 1.2 General definitions

In general, we may say that a *formal system*  $\mathcal{F}$  for a vocabulary  $\mathcal{V}$  consists of certain concrete objects called *deductions*, or more precisely *deductions of  $A$  from  $X$* , where  $A$  is a formula of  $\mathcal{L}(\mathcal{V})$  and  $X$  is a set of formulae of  $\mathcal{L}(\mathcal{V})$ . For the formal systems which we are about to consider, the deductions of  $A$  from  $X$  are certain finite sequences of formulae whose last member is  $A$ . However, in Chapter V we deal with a formal system whose deductions are sequences of sequences of formulae. Although particular formal systems can be set up in a variety of different ways, the one important feature which is shared by every formal system is the existence of an effective procedure for determining whether or not a given array of formulae is a deduction of  $A$  from  $X$  in that system.

If there exists a deduction of  $A$  from  $X$  in  $\mathcal{F}$ , then  $A$  is said to be *deducible from  $X$  in  $\mathcal{F}$*  and we denote this by writing  $X \vdash_{\mathcal{F}} A$ . Thus the statement expressed by the notation  $X \vdash_{\mathcal{F}} A$  is an existential statement in the metalanguage, and a finitary proof of such a statement must provide a procedure for constructing a deduction of  $A$  from  $X$ . A deduction of  $A$  from  $\emptyset$  in  $\mathcal{F}$  is called a *proof* of  $A$  in  $\mathcal{F}$ . If there exists a proof of  $A$  in  $\mathcal{F}$ , then  $A$  is said to be a *theorem* of  $\mathcal{F}$ . A formal system  $\mathcal{F}$  for  $\mathcal{V}$  is *consistent* if  $\perp$  is not a theorem of  $\mathcal{F}$ , *sound* if  $X \vdash_{\mathcal{F}} A$  implies  $X \vDash A$ , and *complete* if  $X \vDash A$  implies  $X \vdash_{\mathcal{F}} A$ , for every  $X \subseteq F_{\mathcal{L}(\mathcal{V})}$  and every  $A \in F_{\mathcal{L}(\mathcal{V})}$ . (This definition of completeness applies only to formal systems which incorporate the  $\varepsilon$ -symbol. For a formal system  $\mathcal{F}$ , such as the predicate calculus, where the  $\varepsilon$ -symbol may not be used in a deduction, we say that  $\mathcal{F}$  is *complete* if  $X \vDash A$  implies  $X \vdash_{\mathcal{F}} A$  for every  $\varepsilon$ -free formula  $A$  of  $\mathcal{L}(\mathcal{V})$  and every set  $X$  of  $\varepsilon$ -free formulae of  $\mathcal{L}(\mathcal{V})$ .)

If  $\mathcal{F}$  is a formal system for  $\mathcal{V}$ , we define the *deductive closure*  $C$  for  $\mathcal{F}$  as follows. For any set  $X$  of formulae in  $\mathcal{L}(\mathcal{V})$

$$C(X) = \{A : X \vdash_{\mathcal{F}} A\}.$$

The following theorem, which depends on the Satisfiability Theorem, provides a useful method for proving the completeness of a formal system.

**THEOREM II.1.** *For any vocabulary  $\mathcal{V}$  and any formal system  $\mathcal{F}$  for  $\mathcal{V}$ , if the deductive closure  $C$  for  $\mathcal{F}$  is a logical closure for  $\mathcal{L}(\mathcal{V})$ , then  $\mathcal{F}$  is complete.*

*Proof (non-finitary).* Suppose  $X \vDash A$ , where  $X \subseteq F_{\mathcal{L}(\mathcal{V})}$  and  $A \in F_{\mathcal{L}(\mathcal{V})}$ . Then by Theorem I.14,  $A \in C(X)$ . Hence  $X \vdash_{\mathcal{F}} A$ , and  $\mathcal{F}$  is complete.

## 1.3 Axioms and rules of inference

Let  $\mathcal{V}$  be any vocabulary. A particular formal system  $\mathcal{F}$  for  $\mathcal{V}$  is often defined by specifying certain formulae of  $\mathcal{L}(\mathcal{V})$  as *axioms* and by prescribing certain *rules of inference*, i.e., rules which determine effectively whether a given formula of  $\mathcal{L}(\mathcal{V})$  'follows from' other given formulae of  $\mathcal{L}(\mathcal{V})$ . The formal concept of a deduction in  $\mathcal{F}$  is defined in the obvious way. For any set  $X$  of formulae of  $\mathcal{L}(\mathcal{V})$  and any formula  $A$  of  $\mathcal{L}(\mathcal{V})$ , a *deduction of  $A$  from  $X$  in  $\mathcal{F}$*  is a sequence  $\langle A_1, \dots, A_n \rangle$  of formulae of  $\mathcal{L}(\mathcal{V})$  such that  $A_n$  is  $A$  and for each  $i = 1, \dots, n$ ,  $A_i$  is an axiom, or  $A_i$  is a member of  $X$ , or  $A_i$  follows by some rule of inference from some preceding members of the sequence.

In order that  $\mathcal{F}$  be effectively defined it is necessary to give an effective definition of the axioms of  $\mathcal{F}$ . Since in most cases we wish to specify infinitely many formulae as axioms, it is impossible to list all the axioms. However, we can specify a finite number of *forms* and then say that every axiom must be of one of these forms. The forms themselves are called *axiom schemata*. For example, we may say that any formula which has the form

$$A \rightarrow B \rightarrow A$$

is an axiom. Then the metalinguistic expression ' $A \rightarrow B \rightarrow A$ ' is an axiom schema. Any formula which has this form is called an *instance* of the axiom schema.

We shall now use the above approach to define a formal system called the  $\varepsilon$ -calculus for  $\mathcal{V}$ .

## 2 The $\varepsilon$ -calculus for $\mathcal{V}$

Let  $\mathcal{V}$  be any vocabulary. The  $\varepsilon$ -calculus for  $\mathcal{V}$ , which we shall denote by  $\varepsilon(\mathcal{V})$ , is defined as follows.

The *axioms* of  $\varepsilon(\mathcal{V})$  are all formulae of  $\mathcal{L}(\mathcal{V})$  which are instances of the following axiom schemata:

- P1  $A \rightarrow B \rightarrow A$   
 P2  $(A \rightarrow B \rightarrow C) \rightarrow (A \rightarrow B) \rightarrow A \rightarrow C$   
 P3  $(\neg A \rightarrow \neg B) \rightarrow (B \rightarrow A)$   
 P4  $(A \rightarrow \perp) \rightarrow \neg A$



P5  $(A \wedge B) \rightarrow A$

P6  $(A \wedge B) \rightarrow B$

P7  $A \rightarrow B \rightarrow (A \wedge B)$

Q1  $\forall xA \rightarrow \neg \exists x \neg A$

Q2  $\neg \forall xA \rightarrow \exists x \neg A$

E1  $(s = t \wedge A_s^x) \rightarrow A_t^x$

E2  $\forall z(A_z^x \leftrightarrow B_z^y) \rightarrow \varepsilon xA = \varepsilon yB$

P8  $\neg(A \vee B) \rightarrow \neg A$

P9  $\neg(A \vee B) \rightarrow \neg B$

P10  $\neg A \rightarrow \neg B \rightarrow \neg(A \vee B)$

Q3  $\neg \exists xA \rightarrow \neg A(t)$

Q4  $\exists xA \rightarrow A(\varepsilon xA)$

E3  $t = t$

*Restriction:* In axiom schema E1,  $A$  is an atom and  $x$  any variable which does not have a free occurrence in  $A$  within the scope of an  $\varepsilon$ -symbol.

In particular, any instance of axiom schemata P1–P10 is a *propositional axiom*, any instance of Q1–Q4 is a *quantificational axiom*, and any instance of E1–E3 is an *equality axiom*. Furthermore, an instance of P1 is called a *P1-axiom*, an instance of P2 a *P2-axiom*, etc.

There is one rule of inference for  $\varepsilon(\mathcal{V})$ . A formula  $B$  follows by *modus ponens from A and C* if and only if  $C$  is of the form  $A \rightarrow B$ . This rule can be expressed schematically as follows.

$$\text{modus ponens: } \frac{A, A \rightarrow B}{B}$$

The deductions in the  $\varepsilon$ -calculus for  $\mathcal{V}$  are now defined in the usual way. For any set  $X$  of formulae of  $\mathcal{L}(\mathcal{V})$  and any formula  $A$  of  $\mathcal{L}(\mathcal{V})$  a *deduction*  $\mathcal{D}$  of  $A$  from  $X$  in  $\varepsilon(\mathcal{V})$  is any finite sequence  $\langle A_1, \dots, A_n \rangle$  of formulae of  $\mathcal{L}(\mathcal{V})$  such that  $A_n$  is  $A$  and for each  $i = 1, \dots, n$  at least one of the following conditions holds:

- (i)  $A_i$  is an axiom of  $\varepsilon(\mathcal{V})$ ,
- (ii)  $A_i$  is a member of  $X$ ,
- (iii)  $A_i$  follows by modus ponens from  $A_j$  and  $A_k$ , for some  $j, k < i$ .

If  $A_i$  is a member of  $X$ , then  $A_i$  is called an *assumption* of  $\mathcal{D}$ , and if  $A_i$  satisfies condition (i) but not conditions (ii) or (iii), then  $A_i$  is said to be *used as an axiom* in  $\mathcal{D}$ . In other words a formula in  $\mathcal{D}$  is used as an axiom if its presence in  $\mathcal{D}$  can be justified only by (i).

If there exists a deduction of  $A$  from  $X$  in  $\varepsilon(\mathcal{V})$ , we say that  $A$  is *deducible* from  $X$  in  $\varepsilon(\mathcal{V})$  and denote this fact by writing  $X \vdash_{\varepsilon(\mathcal{V})} A$ . To simplify the notation we write  $X, B_1, \dots, B_n \vdash_{\varepsilon(\mathcal{V})} A$  instead of  $X \cup \{B_1, \dots, B_n\} \vdash_{\varepsilon(\mathcal{V})} A$  and  $\vdash_{\varepsilon(\mathcal{V})} A$  instead of  $\emptyset \vdash_{\varepsilon(\mathcal{V})} A$ .

Since  $\mathcal{V}$  is an arbitrary vocabulary, the above definition of the formal system  $\varepsilon(\mathcal{V})$  determines a whole class of formal systems, one for each vocab-

ulary  $\mathcal{V}$ . Throughout this chapter we shall let  $\mathcal{V}$  be a fixed, but arbitrary vocabulary, and we shall often speak of ‘the  $\varepsilon$ -calculus’ instead of ‘the  $\varepsilon$ -calculus for  $\mathcal{V}$ ’ and write  $X \vdash_{\varepsilon} A$ , instead of  $X \vdash_{\varepsilon(\mathcal{V})} A$ . Furthermore, when it is understood that we are referring to deductions in the  $\varepsilon$ -calculus (for  $\mathcal{V}$ ) we shall write simply  $X \vdash A$ .

Recall that a *proof* in a formal system  $\mathcal{F}$  is a deduction from the empty set, and that  $A$  is a *theorem* of  $\mathcal{F}$  if there exists a proof of  $A$ . We now give an example of a proof in  $\varepsilon(\mathcal{V})$ .

**THEOREM II.2.** *For any formula  $A$  of  $\mathcal{L}(\mathcal{V})$ , the formula  $A \rightarrow A$  is a theorem of  $\varepsilon(\mathcal{V})$ .*

*Proof.* The following sequence of formulae constitutes a proof of  $A \rightarrow A$ .

- (1)  $A \rightarrow (A \rightarrow A) \rightarrow A$  P1-axiom
- (2)  $(A \rightarrow (A \rightarrow A) \rightarrow A) \rightarrow (A \rightarrow A \rightarrow A) \rightarrow A \rightarrow A$  P2-axiom
- (3)  $(A \rightarrow A \rightarrow A) \rightarrow A \rightarrow A$  modus ponens from (1) and (2)
- (4)  $A \rightarrow A \rightarrow A$  P1-axiom
- (5)  $A \rightarrow A$  modus ponens from (4) and (3)

**THEOREM II.3.** *Let  $X$  and  $Y$  be any sets of formulae of  $\mathcal{L}(\mathcal{V})$  and  $A, B_1, \dots, B_n$  any formulae of  $\mathcal{L}(\mathcal{V})$ . Then:*

- (i) *If  $A$  is an axiom of  $\varepsilon(\mathcal{V})$  or a member of  $X$ , then  $X \vdash A$ .*
- (ii) *If  $Y, B_1, \dots, B_n \vdash A$ , and  $X \vdash B_i$  for each  $i = 1, \dots, n$ , then  $X \cup Y \vdash A$ .*
- (iii) *If  $X \subseteq Y$  and  $X \vdash A$ , then  $Y \vdash A$ .*
- (iv) *If  $X \vdash A$ , then there exists a finite subset  $X'$  of  $X$  such that  $X' \vdash A$ .*

*Proof.* Parts (i), (iii), and (iv) follow immediately from the definition of a deduction. To prove part (ii) it is sufficient to consider the case where  $n = 1$ , since the general case then follows by induction. Let  $\langle A_1, \dots, A_m \rangle$  be a deduction of  $B_1$  from  $X$  and let  $\langle C_1, \dots, C_n \rangle$  be a deduction of  $A$  from  $Y \cup \{B_1\}$ . Then the sequence  $\langle A_1, \dots, A_m, C_1, \dots, C_n \rangle$  is a deduction of  $A$  from  $X \cup Y$ .

### EXERCISES

1. Using only Theorem II.3, prove that the deductive closure  $C$  for  $\varepsilon(\mathcal{V})$  is a finitary closure operation on the set of formulae of  $\mathcal{L}(\mathcal{V})$ .
2. Prove that every propositional axiom of  $\varepsilon(\mathcal{V})$  is a tautology.
3. Prove that  $\varepsilon(\mathcal{V})$  is sound, i.e., for any set  $X$  of formulae of  $\mathcal{L}(\mathcal{V})$  and any formula  $A$  of  $\mathcal{L}(\mathcal{V})$  prove that  $X \vdash_{\varepsilon(\mathcal{V})} A$  implies  $X \models A$ .

### 3 Derived rules of inference

In general, it is very impractical to prove that there exists a deduction of  $A$  from  $X$  by actually displaying the appropriate sequence of formulae (as we

did in proving Theorem II.2). It is far more convenient to have at our disposal certain basic rules which assert that a deduction of  $A$  from  $X$  can be formed from certain known deductions. Such rules are called *derived rules of inference*. For example, Theorem II.3(ii) asserts a useful rule of this type. The following theorem provides two rather obvious derived rules of inference.

**THEOREM II.4.** *Let  $X$  and  $Y$  be any sets of formulae. Then:*

- (i)  $\rightarrow$ -elimination rule: *If  $X \vdash A \rightarrow B$ , then  $X, A \vdash B$ .*
- (ii) MP rule: *If  $Y \vdash B_1 \rightarrow \dots \rightarrow B_n \rightarrow A$ , and  $X \vdash B_i$  for each  $i = 1, \dots, n$ , then  $X \cup Y \vdash A$ .*

*Proof.* (i) Let  $\langle A_1, \dots, A_n \rangle$  be a deduction of  $A \rightarrow B$  from  $X$ . Then the sequence  $\langle A_1, \dots, A_n, A, B \rangle$  is a deduction of  $B$  from  $X \cup \{A\}$ , since  $B$  follows by modus ponens from  $A$  and  $A_n$ .

(ii) Since  $Y \vdash B_1 \rightarrow \dots \rightarrow B_n \rightarrow A$ , then  $Y, B_1, \dots, B_n \vdash A$  by  $n$  applications of the  $\rightarrow$ -elimination rule. Hence  $X \cup Y \vdash A$  by Theorem II.3(ii).

In the next section we shall prove the converse of part (i). Although this result is commonly called the Deduction Theorem, we shall refer to it as the  $\rightarrow$ -introduction rule in order to emphasize its role as a derived rule of inference.

### 3.1 The $\rightarrow$ -introduction rule (Deduction Theorem)

In mathematics one commonly proves a statement of the form 'if  $A$ , then  $B$ ' by taking  $A$  as an assumption and deducing  $B$  from  $A$ . The following theorem can be regarded as a formal justification of this method.

**THEOREM II.5** (The  $\rightarrow$ -introduction rule). *If  $X, A \vdash B$ , then  $X \vdash A \rightarrow B$ .*

*Proof.* Let  $\langle A_1, \dots, A_n \rangle$  be a deduction of  $B$  from  $X \cup \{A\}$ . We shall prove by induction that for each  $i$ ,  $X \vdash A \rightarrow A_i$ , thus proving  $X \vdash A \rightarrow B$ , since  $A_n$  is  $B$ .

*Case 1.*  $A_i$  is an axiom or a member of  $X$ : Hence

$$\begin{array}{ll} X \vdash A_i & \text{Theorem II.3(i)} \\ \vdash A_i \rightarrow A \rightarrow A_i & \text{P1-axiom} \\ X \vdash A \rightarrow A_i & \text{MP} \end{array}$$

*Case 2.*  $A_i$  is the formula  $A$ :

$$\begin{array}{ll} \vdash A \rightarrow A_i & \text{Theorem II.2} \\ X \vdash A \rightarrow A_i & \text{Theorem II.3(iii)}. \end{array}$$

*Case 3.*  $A_i$  follows from  $A_j$  and  $A_k$  by modus ponens, where  $j, k < i$ : Then  $A_k$  is the formula  $A_j \rightarrow A_i$ . Hence

### §3.1 THE $\rightarrow$ -INTRODUCTION RULE (DEDUCTION THEOREM) 43

$$\begin{array}{ll} X \vdash A \rightarrow A_j \rightarrow A_i & \text{induction hypothesis} \\ X \vdash A \rightarrow A_j & \text{induction hypothesis} \\ \vdash (A \rightarrow A_j \rightarrow A_i) \rightarrow (A \rightarrow A_j) \rightarrow A \rightarrow A_i & \\ & \text{P2-axiom} \\ X \vdash A \rightarrow A_i & \text{MP} \end{array}$$

**COROLLARY** (The syllogism rule). *If  $X \vdash A_1 \rightarrow A_2$ ,  $X \vdash A_2 \rightarrow A_3, \dots$ , and  $X \vdash A_{n-1} \rightarrow A_n$ , then  $X \vdash A_1 \rightarrow A_n$ .*

*Proof.* It is sufficient to prove the rule for the case where  $n = 3$ , since the general case then follows by induction on  $n$ .

Assume  $\vdash A_1 \rightarrow A_2$  and  $X \vdash A_2 \rightarrow A_3$ . Then

$$\begin{array}{ll} X, A_1 \vdash A_2 & \rightarrow\text{-elimination} \\ X, A_1 \vdash A_3 & \text{MP} \\ X \vdash A_1 \rightarrow A_3 & \rightarrow\text{-introduction} \end{array}$$

### 3.2 Some theorems of the $\varepsilon$ -calculus

Our derived rules of inference can now be used to show that formulae of certain standard forms are theorems of the  $\varepsilon$ -calculus.

**THEOREM II.6.** *For any formulae  $A$  and  $B$ :*

- (i)  $\vdash \neg A \rightarrow A \rightarrow B$ ,
- (ii)  $\vdash \neg \neg A \rightarrow A$ ,
- (iii)  $\vdash A \rightarrow \neg \neg A$ .

*Proof.*

$$\begin{array}{ll} \text{(i)} \vdash \neg A \rightarrow \neg B \rightarrow \neg A & \text{P1-axiom} \\ \vdash (\neg B \rightarrow \neg A) \rightarrow (A \rightarrow B) & \text{P3-axiom} \\ \vdash \neg A \rightarrow A \rightarrow B & \text{syllogism} \\ \text{(ii)} \vdash \neg \neg A \rightarrow \neg A \rightarrow \neg \neg \neg A & \text{part (i)} \\ \vdash (\neg A \rightarrow \neg \neg \neg A) \rightarrow (\neg \neg A \rightarrow A) & \text{P3-axiom} \\ \vdash \neg \neg A \rightarrow \neg \neg A \rightarrow A & \text{syllogism} \\ \neg \neg A \vdash A & \rightarrow\text{-elimination (twice)} \\ \vdash \neg \neg A \rightarrow A & \rightarrow\text{-introduction} \\ \text{(iii)} \vdash \neg \neg \neg A \rightarrow \neg A & \text{part (ii)} \\ \vdash (\neg \neg \neg A \rightarrow \neg A) \rightarrow (A \rightarrow \neg \neg A) & \text{P3-axiom} \\ \vdash A \rightarrow \neg \neg A & \text{MP} \end{array}$$

**THEOREM II.7** (Contrapositive rules).

- (i) *If  $X \vdash \neg A \rightarrow \neg B$ , then  $X \vdash B \rightarrow A$ .*
- (ii) *If  $X \vdash \neg A \rightarrow B$ , then  $X \vdash \neg B \rightarrow A$ .*
- (iii) *If  $X \vdash A \rightarrow \neg B$ , then  $X \vdash B \rightarrow \neg A$ .*
- (iv) *If  $X \vdash A \rightarrow B$ , then  $X \vdash \neg B \rightarrow \neg A$ .*

*Proof.* Use axiom schema P3, Theorem II.6(ii), (iii), the MP rule, and the syllogism rule.

**THEOREM II.8.** For any conjunctive formula  $\alpha$ :

- (i)  $\vdash \alpha \rightarrow \alpha_1$ ,
- (ii)  $\vdash \alpha \rightarrow \alpha_2$ ,
- (iii)  $\vdash \alpha_1 \rightarrow \alpha_2 \rightarrow \alpha$ .

*Proof.* If  $\alpha$  is of the form  $A \wedge B$  or  $\neg(A \vee B)$ , then (i), (ii), and (iii) follow by the propositional axioms P5, P6, and P7, or P8, P9, and P10 respectively. Suppose  $\alpha$  is of the form  $\neg(A \rightarrow B)$ .

- (i)  $\vdash \neg A \rightarrow A \rightarrow B$       Theorem II.6(i)  
 $\vdash \neg(A \rightarrow B) \rightarrow A$       contrapositive
- (ii)  $\vdash B \rightarrow A \rightarrow B$       P1-axiom  
 $\vdash \neg(A \rightarrow B) \rightarrow \neg B$       contrapositive
- (iii)  $A, A \rightarrow B \vdash B$       by modus ponens  
 $A \vdash (A \rightarrow B) \rightarrow B$        $\rightarrow$ -introduction  
 $A \vdash \neg B \rightarrow \neg(A \rightarrow B)$       contrapositive  
 $\vdash A \rightarrow \neg B \rightarrow \neg(A \rightarrow B)$        $\rightarrow$ -introduction

**THEOREM II.9** (The  $f$ -rules).

- (i)  $X \vdash \neg A$  iff  $X, A \vdash f$ ,
- (ii)  $X \vdash A$  iff  $X, \neg A \vdash f$ .

*Proof.*

- (i) Assume  $X \vdash \neg A$   
 $\vdash \neg A \rightarrow A \rightarrow f$       Theorem II.6(i)  
 $X \vdash A \rightarrow f$       MP  
 $X, A \vdash f$        $\rightarrow$ -elimination
- Assume  $X, A \vdash f$   
 $X \vdash A \rightarrow f$        $\rightarrow$ -introduction  
 $\vdash (A \rightarrow f) \rightarrow \neg A$       P4-axiom  
 $X \vdash \neg A$       MP
- (ii)  $X \vdash A$  iff  $X \vdash \neg \neg A$       Theorem II.6(ii), (iii) and MP  
       iff  $X, \neg A \vdash f$       part (i)

#### 4 Completeness of the $\varepsilon$ -calculus

So far everything which we have proved about the  $\varepsilon$ -calculus for  $\mathcal{V}$  depends only on the facts that formulae of the form P1–P10 are axioms and that modus ponens is the one and only rule of inference. We shall now make use of the quantificational axioms.

**THEOREM II.10.** For any universal formula  $\gamma$ , any term  $t$ , and any existential formula  $\delta$ :

- (i)  $\vdash \gamma \rightarrow \gamma(t)$ ,
- (ii)  $\vdash \delta \rightarrow \delta(\varepsilon\delta)$ .

*Proof.* If  $\gamma$  is of the form  $\neg\exists xA$ , then  $\gamma \rightarrow \gamma(t)$  is a Q3-axiom, and if  $\delta$  is of the form  $\exists xA$ , then  $\delta \rightarrow \delta(\varepsilon\delta)$  is a Q4-axiom. Suppose  $\gamma$  is  $\forall xA$  and  $\delta$  is  $\neg\forall xA$ .

- (i)  $\vdash \forall xA \rightarrow \neg\exists x\neg A$       Q1-axiom  
 $\vdash \neg\exists x\neg A \rightarrow \neg\neg A(t)$       Q3-axiom  
 $\vdash \neg\neg A(t) \rightarrow A(t)$       Theorem II.6(ii)  
 $\vdash \forall xA \rightarrow A(t)$       syllogism
- (ii)  $\vdash \neg\forall xA \rightarrow \exists x\neg A$       Q2-axiom  
 $\vdash \exists x\neg A \rightarrow \neg A(\varepsilon x\neg A)$       Q4-axiom  
 $\vdash \neg\forall xA \rightarrow \neg A(\varepsilon x\neg A)$       syllogism

**THEOREM II.11** (The Completeness Theorem). For any vocabulary  $\mathcal{V}$  the  $\varepsilon$ -calculus for  $\mathcal{V}$  is complete.

*Proof (non-constructive).* By Theorem II.1, in order to prove that  $\varepsilon(\mathcal{V})$  is complete it is sufficient to prove that its deductive closure  $C$  is a logical closure operation. That  $C$  is a finitary closure operation follows from Theorem II.3. The 'logical' properties L1, L2, L3, and L4 follow from Theorems II.9, II.8, II.10(i), and II.10(ii), respectively, using the MP rule. Properties L5 and L6 follow easily from the equality axioms E1 and E2.

Notice that our proof of the Completeness Theorem does not make use of the fact that formulae of the form

$$E3 \quad t = t$$

are regarded as axioms of the  $\varepsilon$ -calculus. Although axioms of this form are superfluous in the  $\varepsilon$ -calculus, it is convenient to include them since they are needed in proving that the E2-axioms, i.e., formulae of the form

$$E2 \quad \forall z(A_z^x \leftrightarrow B_z^y) \rightarrow \varepsilon xA = \varepsilon yB$$

can be eliminated from proofs of  $\varepsilon$ -free formulae.

#### EXERCISE

Prove constructively that if  $X \vdash A$ , then there exists a deduction of  $A$  from  $X$  in which no formula of the form  $t = t$  is used as an axiom. (See the proof of Lemma 1(ix), page 26.)

#### 5 The Tautology Theorem

Our proof of the Completeness Theorem is a good example of a non-constructive existence proof, since we prove that there exists a deduction of  $A$

from  $X$  without actually describing how such a deduction can be formed. We now give a completely constructive proof of a much weaker completeness result—namely, that if  $A$  is a tautological consequence of  $B_1, \dots, B_n$ , then there exists a deduction of  $A$  from  $\{B_1, \dots, B_n\}$ .

We first establish some additional derived rules of inference.

**THEOREM II.12.** *Let  $X$  be any set of formulae,  $\alpha$  any conjunctive formula, and  $\beta$  any disjunctive formula. Then:*

- (i) Contradiction rule: if  $f$  is a member of  $X$  or if there exists some formula  $A$  such that both  $A$  and  $\neg A$  are members of  $X$ , then  $X \vdash f$ .
- (ii)  $\neg\neg$ -rule: if  $X, A \vdash f$ , then  $X, \neg\neg A \vdash f$ .
- (iii)  $\alpha$ -rule: if  $X, \alpha_1, \alpha_2 \vdash f$ , then  $X, \alpha \vdash f$ .
- (iv)  $\beta$ -rule: if  $X, \beta_1 \vdash f$  and  $X, \beta_2 \vdash f$ , then  $X, \beta \vdash f$ .

*Proof.*

(i) If  $f$  is a member of  $X$ , then  $X \vdash f$  by Theorem II.3(i). Suppose  $A$  and  $\neg A$  are members of  $X$ . Since  $A \vdash A$ , then  $A, \neg A \vdash f$  by the  $f$ -rule. Hence  $X \vdash f$  by Theorem II.3(iii).

(ii) Assume  $X, A \vdash f$   
 $X \vdash \neg A$   $f$ -rule (i)  
 $X, \neg\neg A \vdash f$   $f$ -rule (ii)

(iii) Assume  $X, \alpha_1, \alpha_2 \vdash f$   
 $\alpha \vdash \alpha_1$  Theorem II.8(i) and  $\rightarrow$ -elimination  
 $\alpha \vdash \alpha_2$  Theorem II.8(ii) and  $\rightarrow$ -elimination  
 $X, \alpha \vdash f$  Theorem II.3(ii)

(iv) Assume  $X, \beta_1 \vdash f$  and  $X, \beta_2 \vdash f$ , where  $\beta_1$  and  $\beta_2$  are the disjunctive components of some disjunctive formula  $\beta$ . By the duality principle (page 16) the contrary of  $\beta$  is a conjunctive formula  $\alpha$  such that  $\alpha_1$  and  $\beta_1$  are contradictory and  $\alpha_2$  and  $\beta_2$  are contradictory. Hence

$X \vdash \alpha_1$   $f$ -rule  
 $X \vdash \alpha_2$   $f$ -rule  
 $\vdash \alpha_1 \rightarrow \alpha_2 \rightarrow \alpha$  Theorem II.8(iii)  
 $X \vdash \alpha$  MP  
 $X, \beta \vdash f$   $f$ -rule

A finite set of formulae  $\{A_1, \dots, A_n\}$  is said to be *truth functionally invalid* if for every truth assignment  $\psi$  for  $A_1 \wedge \dots \wedge A_n$  there exists an  $A_i$  such that  $\psi(A_i) = 0$ .

**THEOREM II.13.** *If the set  $\{A_1, \dots, A_n\}$  is truth functionally invalid, then  $A_1, \dots, A_n \vdash f$ .*

*Proof.* Let  $m$  be the sum of the lengths of the  $A_i$ . The proof is by induction on  $m$ .

*Case 1. Each of the  $A_i$  is either a molecule or the negation of a molecule:* We shall prove that since  $\{A_1, \dots, A_n\}$  is truth functionally invalid, then either (i) some  $A_i$  is the formula  $f$ , or (ii) some  $A_i$  is the negation of some  $A_j$ . It then follows that  $A_1, \dots, A_n \vdash f$  by the contradiction rule. Suppose neither (i) nor (ii) holds. We can then define a truth assignment  $\psi$  for  $A_1 \wedge \dots \wedge A_n$  as follows. For each  $A_i$ , if  $A_i$  is a molecule let  $\psi(A_i) = 1$ , and if  $A_i$  is the negation of a molecule  $B_i$  let  $\psi(B_i) = 0$ . It then follows that for each  $i$ ,  $\psi(A_i) = 1$ , which contradicts the assumption that  $\{A_1, \dots, A_n\}$  is truth functionally invalid.

*Case 2. At least one of the  $A_i$  is neither a molecule nor the negation of a molecule:* We may assume that such a formula is  $A_1$ . Then one of the following three cases must hold.

*Case 2a.  $A_1$  is of the form  $\neg\neg B$ :* In this case  $\{B, A_2, \dots, A_n\}$  must be truth functionally invalid since for any  $\psi$ , if  $\psi(\neg\neg B) = 0$ , then  $\psi(B) = 0$ . By the induction hypothesis  $B, A_2, \dots, A_n \vdash f$ , and therefore  $\neg\neg B, A_2, \dots, A_n \vdash f$  by the  $\neg\neg$ -rule.

*Case 2b.  $A_1$  is a conjunctive formula  $\alpha$ :* In this case  $\{\alpha_1, \alpha_2, A_2, \dots, A_n\}$  must be truth functionally invalid since for any  $\psi$ , if  $\psi(\alpha) = 0$ , then  $\psi(\alpha_1) = 0$  or  $\psi(\alpha_2) = 0$ . By the induction hypothesis  $\alpha_1, \alpha_2, A_2, \dots, A_n \vdash f$ , and therefore  $\alpha, A_2, \dots, A_n \vdash f$  by the  $\alpha$ -rule.

*Case 2c.  $A_1$  is a disjunctive formula  $\beta$ :* In this case both  $\{\beta_1, A_2, \dots, A_n\}$  and  $\{\beta_2, A_2, \dots, A_n\}$  are truth functionally invalid since for any  $\psi$ , if  $\psi(\beta) = 0$ , then  $\psi(\beta_1) = 0$  and  $\psi(\beta_2) = 0$ . Hence  $\beta, A_2, \dots, A_n \vdash f$  by the induction hypothesis and the  $\beta$ -rule.

Notice that the above theorem holds for any formal system which satisfies the four rules that make up Theorem II.12.

**THEOREM II.14 (The Tautology Theorem).** *Every tautology is a theorem of the  $\varepsilon$ -calculus.*

*Proof.* Let  $A$  be a tautology. Then  $\{\neg A\}$  is truth functionally invalid and by Theorem II.13,  $\neg A \vdash f$ . Hence  $\vdash A$  by the  $f$ -rule.

**COROLLARY.**

- (i) *If  $A$  is a tautological consequence of  $B_1, \dots, B_n$ , then  $B_1, \dots, B_n \vdash A$ .*
- (ii) *The tautology rule: if  $A$  is a tautological consequence of  $B_1, \dots, B_n$  and  $X \vdash B_1, \dots, X \vdash B_n$ , then  $X \vdash A$ .*

*Proof.* Part (i) follows by the Tautology Theorem and the  $\rightarrow$ -elimination rule, and part (ii) by the Tautology Theorem and the MP rule.

## 6 The consistency of the $\varepsilon$ -calculus

Recall that a formal system  $\mathcal{F}$  is *consistent* if  $f$  is not a theorem of  $\mathcal{F}$ . Since  $f$  is not a valid formula, the soundness of the  $\varepsilon$ -calculus (cf. Exercise 3,

page 41) implies its consistency. However this consistency proof is somewhat unsatisfactory since it involves the non-finitary notion of validity. For this reason, we shall now give a completely finitary proof of the consistency of the  $\varepsilon$ -calculus.

**THEOREM II.15.** *For any vocabulary  $\mathcal{V}$ , the  $\varepsilon$ -calculus for  $\mathcal{V}$  is consistent.*

*Proof.* For any quasi-formula  $A$  of  $\mathcal{L}(\mathcal{V})$ , we define the formula  $g(A)$  as follows by induction on the length of  $A$ .

- (i) If  $A$  is of the form  $Pt_1, \dots, t_n$ , then  $g(A)$  is  $P$ , where  $P$  is now regarded as a 0-place predicate symbol.
- (ii) If  $A$  is of the form  $s = t$ , then  $g(A)$  is  $\neg f$ .
- (iii) If  $A$  is the formula  $f$ , then  $g(A)$  is  $f$ .
- (iv) If  $A$  is of the form  $\neg B$ , or  $B * C$ , where  $*$  is  $\wedge$ ,  $\vee$ , or  $\rightarrow$ , then  $g(A)$  is  $\neg g(B)$ , or  $g(B) * g(C)$ , respectively.
- (v) If  $A$  is of the form  $\exists xB$  or  $\forall xB$ , then  $g(A)$  is  $g(B)$ .

In other words,  $g(A)$  is obtained from  $A$  by first erasing all the quasi-terms in  $A$  and all occurrences of the symbols  $\forall x$  and  $\exists x$ , and then replacing each occurrence of  $=$  by  $\neg f$ . It is easy to see that if  $A$  is an axiom of  $\varepsilon(\mathcal{V})$ , then  $g(A)$  is a tautology, and if  $g(A)$  and  $g(A \rightarrow B)$  are tautologies, then  $g(B)$  is a tautology. Consequently, if  $A$  is a theorem of the  $\varepsilon$ -calculus for  $\mathcal{V}$ , then  $g(A)$  is a tautology. Since  $g(f)$  is not a tautology, then  $f$  is not a theorem, and therefore  $\varepsilon(\mathcal{V})$  is consistent.

### 7.1 Some derived rules for operating with quantifiers

Theorem II.10 states that all formulae of the following forms are theorems of the  $\varepsilon$ -calculus: (i)  $\forall xA \rightarrow A(t)$ , (ii)  $\neg \exists xA \rightarrow \neg A(t)$ , (iii)  $\neg \forall xA \rightarrow \neg A(\varepsilon x \neg A)$ , and (iv)  $\exists xA \rightarrow A(\varepsilon x A)$ . These results provide the following derived rules of inference for the introduction and elimination of quantifiers.

**THEOREM II.16.** *Let  $X$  be any set of formulae. Then:*

- (i)  $\forall$ -elimination rule: if  $X \vdash \forall xA$ , then  $X \vdash A(t)$  for any term  $t$ .
- (ii)  $\exists$ -introduction rule: if there exists a term  $t$  such that  $X \vdash A(t)$ , then  $X \vdash \exists xA$ .
- (iii)  $\forall$ -introduction rule: if  $X \vdash A(\varepsilon x \neg A)$ , then  $X \vdash \forall xA$ .
- (iv)  $\exists$ -elimination rule: if  $X \vdash \exists xA$ , then  $X \vdash A(\varepsilon x A)$ .

The proof follows immediately by Theorem II.10 and the tautology rule.

Often the specified  $\varepsilon$ -terms in the  $\forall$ -introduction rule and the  $\exists$ -elimination rule are rather complicated expressions. For example, if we know that  $X \vdash \exists x \exists y Pxy$  and we then make two applications of the  $\exists$ -elimination rule we obtain the following complicated result

$$X \vdash P \varepsilon x \exists y Pxy \varepsilon y P \varepsilon x \exists y Pxyy.$$

### §7.1 SOME DERIVED RULES FOR OPERATING WITH QUANTIFIERS 49

In order to avoid such complicated expressions it is convenient to use metalinguistic symbols, such as the letters  $s$  and  $t$ , to denote certain specified  $\varepsilon$ -terms. For example, we can prove  $\vdash \exists x \forall y Pxy \rightarrow \forall y \exists x Pxy$  as follows. Let  $X$  be the set  $\{\exists x \forall y Pxy\}$ . Then

$X \vdash \exists x \forall y Pxy$	Theorem II.3(i)
$X \vdash \forall y Pxy$	$\exists$ -elimination, $s$ is $\varepsilon x \forall y Pxy$
$X \vdash Pst$	$\forall$ -elimination, $t$ is $\varepsilon y \neg \exists x Pxy$
$X \vdash \exists x Pxt$	$\exists$ -introduction
$X \vdash \forall y \exists x Pxy$	$\forall$ -introduction
$\vdash \exists x \forall y Pxy \rightarrow \forall y \exists x Pxy$	$\rightarrow$ -introduction

### EXERCISE

Prove  $\vdash \exists x \exists y Pxy \rightarrow \exists y \exists x Pxy$  by using the letters  $s$  and  $t$  to denote the appropriate  $\varepsilon$ -terms.

An alternative way of avoiding complicated  $\varepsilon$ -terms is by employing the following rules of inference in which individual symbols are used in effect as abbreviations for arbitrary  $\varepsilon$ -terms.

**THEOREM II.17.** *Let  $X$  be any set of formulae,  $B$  any formula,  $A$  any quasi-formula which contains no free variable other than  $x$ , and  $a$  any individual symbol which does not appear in  $A$ ,  $B$ , or any member of  $X$ . Then:*

- (i) Substitution rule: if  $X \vdash A_a^x$ , then  $X \vdash A_t^x$ , for any term  $t$ .
- (ii) Generalization rule: if  $X \vdash A_a^x$ , then  $X \vdash \forall x A$ .
- (iii)  $\exists$ -rule: if  $\vdash A_a^x \rightarrow B$ , then  $X \vdash \exists x A \rightarrow B$ .
- (iv)  $\forall$ -rule: if  $X \vdash B \rightarrow A_a^x$ , then  $X \vdash B \rightarrow \forall x A$ .

*Proof.*

(i) Let  $\langle A_1, \dots, A_n \rangle$  be a deduction of  $A_a^x$  from  $X$ . For each  $i = 1, \dots, n$ , let  $A_i'$  be the formula obtained from  $A_i$  by replacing each occurrence of  $a$  in  $A_i$  by the term  $t$ . Since  $a$  does not occur in  $A$  and since  $A_n$  is  $A_a^x$ , then  $A_n'$  is  $A_t^x$ . It is easy to see that the sequence  $\langle A_1', \dots, A_n' \rangle$  constitutes a deduction of  $A_t^x$  from  $X$ . For, if  $A_i$  is an axiom, then  $A_i'$  is an axiom of the same form, if  $A_i$  is a member of  $X$ , then  $A_i'$  is  $A_i$ , and if  $A_i$  follows by modus ponens from  $A_j$  and  $A_k$ , then  $A_i'$  follows by that rule from  $A_j'$  and  $A_k'$ .

- (ii) Assume  $X \vdash A_a^x$ . Then  $X \vdash A(\varepsilon x \neg A)$  by part (i). Hence  $X \vdash \forall x A$  by the  $\forall$ -introduction rule.
- (iii) Assume  $X \vdash A_a^x \rightarrow B$ . Since  $a$  does not appear in  $B$ , the proof of part (i) yields  $X \vdash A(\varepsilon x A) \rightarrow B$ . Therefore  $X \vdash \exists x A \rightarrow B$  by the syllogism rule and the Q4-axiom  $\exists x A \rightarrow A(\varepsilon x A)$ .
- (iv) The proof of (iv) is similar to that of (iii).

The proof of the above theorem depends on the crucial fact that when an individual symbol  $a$  is replaced by some term  $t$  in an axiom, the resulting formula is an axiom of the same form. The converse assertion is not in general true—that is, replacing a term  $t$  by an individual symbol  $a$  in an axiom does not necessarily yield another axiom of the same form. For example, the term  $t$  may be one of the specified  $\varepsilon$ -terms in a Q4-axiom or an E2-axiom, or it may contain the specified terms in a Q3- or Q4-axiom. The fact that the converse does not hold provides one of the major difficulties in proving the eliminability of the  $\varepsilon$ -symbol (Hilbert's Second  $\varepsilon$ -Theorem) as we shall see in the next chapter.

In view of these remarks we can now explain the motivation behind the restrictions which we imposed on the E1-axioms. Recall that a formula is an E1-axiom if it is of the form

$$(s = t \wedge A_s^x) \rightarrow A_t^x,$$

provided that  $A$  is an atom and  $x$  is a variable which does not have a free occurrence in  $A$  within the scope of an  $\varepsilon$ -symbol. This second restriction guarantees that if any  $\varepsilon$ -term is replaced by some other term  $t$  in an E1-axiom, the resulting formula is still an E1-axiom. Consequently in our proof of the Second  $\varepsilon$ -Theorem, the E1-axioms present no difficulties. (The reason for restricting  $A$  to an atom will be seen in Chapter V.)

In spite of these restrictions on the E1-axioms, the desired results concerning the identity symbol are deducible in the  $\varepsilon$ -calculus. In particular, we shall prove presently that any formula of the form

$$(s = t \wedge A_s^x) \rightarrow A_t^x$$

is a theorem of the  $\varepsilon$ -calculus.

We now use Theorem II.17 to establish another useful derived rule of inference.

**THEOREM II.18** (The distribution rule). *Let  $X$  be any set of formulae,  $A$  and  $B$  any quasi-formulae, and  $a$  any individual symbol not appearing in  $A$ ,  $B$ , or any member of  $X$ . If  $X \vdash A_a^x \leftrightarrow B_a^y$ , then*

- (i)  $X \vdash \forall xA \leftrightarrow \forall yB$ ,
- (ii)  $X \vdash \exists xA \leftrightarrow \exists yB$ ,
- (iii)  $X \vdash \varepsilon xA = \varepsilon yB$ .

*Proof.* (i) Assume  $X \vdash A_a^x \leftrightarrow B_a^y$ . Then

$$\begin{array}{ll} X \vdash A_a^x \rightarrow B_a^y & \text{tautology rule} \\ X \vdash \forall xA \rightarrow A_a^x & \text{Theorem II.10(i)} \\ X \vdash \forall xA \rightarrow B_a^y & \text{syllogism rule} \\ X \vdash \forall xA \rightarrow \forall yB & \forall\text{-rule.} \end{array}$$

Similarly,  $X \vdash \forall yB \rightarrow \forall xA$ . Hence  $X \vdash \forall xA \leftrightarrow \forall yB$  by the tautology rule.

(ii) The proof of (ii) is similar to that of (i).

(iii) Assume  $X \vdash A_a^x \leftrightarrow B_a^y$ . Let  $z$  be any variable which is free for  $x$  in  $A$  and free for  $y$  in  $B$ . Since  $[A_z^x]^z$  is  $A_a^x$  and  $[B_z^y]^z$  is  $B_a^y$ , then  $X \vdash [A_z^x]^z \leftrightarrow [B_z^y]^z$ . Hence  $X \vdash \forall z(A_z^x \leftrightarrow B_z^y)$  by the generalization rule, and  $X \vdash \varepsilon xA = \varepsilon yB$  by axiom schema E2 and the MP rule.

## 7.2 Substitution instances and universal closures

Let  $A$  be any quasi-formula and let  $x_1, \dots, x_n$  be the free variables in  $A$ . For any terms  $t_1, \dots, t_n$  the formula  $A_{t_1}^{x_1} \dots A_{t_n}^{x_n}$  is called a *substitution instance* of  $A$ , and the formula  $\forall x_1 \dots \forall x_n A$  is called a *universal closure* of  $A$ . Thus  $A$  has  $n!$  universal closures—one for each ordering of the variables in  $A$ . (If  $A$  has no free variables, then  $A$  is regarded as a substitution instance and a universal closure of itself.) For any  $X$ , we write  $X \vdash \forall[A]$  to denote that every universal closure of  $A$  is deducible from  $X$ .

**THEOREM II.19.** *Let  $X$  be any set of formulae and  $A$  any quasi-formula. Then:*

- (i) *If  $B$  is a universal closure of  $A$  and  $X \vdash B$ , then every substitution instance of  $A$  is deducible from  $X$ .*
- (ii) *If every substitution instance of  $A$  is deducible from  $X$ , then  $X \vdash \forall[A]$ .*

*Proof.* The proof of (i) follows by repeated application of the  $\forall$ -elimination rule and the proof of (ii) by repeated application of the  $\forall$ -introduction rule.

Theorem II.19(ii) has the following very useful application. Suppose we want to prove  $\vdash \forall[A]$  for any quasi-formula  $A$  of a certain given form. If every substitution instance of  $A$  is also of this form, then by the above theorem it is sufficient to prove that every formula of this form is a theorem. In other words we may assume that  $A$  is a formula and simply prove  $\vdash A$ . We shall use this technique in proving the following theorem.

**THEOREM II.20.** *Let  $B$  and  $C$  be any quasi-formulae,  $x$  any variable, and  $y$  any variable which is free for  $x$  in  $B$  and does not occur free in either  $B$  or  $C$ . Let  $Q$  denote either  $\forall$  or  $\exists$ . Then:*

- (i)  $\vdash \forall[QxB \leftrightarrow Qy[B]^y]$  and  $\vdash \forall[\varepsilon xB = \varepsilon y[B]^y]$ ;
- (ii)  $\vdash \forall[\neg QxB \leftrightarrow Q'y \neg B]$ , where  $Q'$  is  $\exists$  if  $Q$  is  $\forall$  and  $Q'$  is  $\forall$  if  $Q$  is  $\exists$ ;
- (iii)  $\vdash \forall[(QxB \vee C) \leftrightarrow Qy(B_y^x \vee C)]$ ;
- (iv)  $\vdash \forall[(C \vee QxB) \leftrightarrow Qy(C \vee B_y^x)]$ .

*Proof.* In each of the four parts of this theorem we want to prove that if  $A$  is a quasi-formula of a certain form, then  $\vdash \forall[A]$ . Since in each case every substitution instance of  $A$  is a formula of the same form as that of  $A$ , then by the above remark we may assume that  $A$  is a formula and simply prove  $\vdash A$ .

(i) Let  $a$  be any individual symbol not appearing in  $B$ . Since  $y$  is free for  $x$  in  $B$ , but not free in  $B$ , then  $[B_y^x]^y$  is  $B_a^x$ . Hence  $\vdash B_a^x \leftrightarrow [B_y^x]^y$  by the Tautology Theorem. The desired results now follow by the distribution rule.

(ii) If  $Q$  is  $\forall$ , then  $\neg QxB \leftrightarrow Q'\neg x\neg B$  is a tautological consequence of the  $Q$ -axioms  $\forall xB \rightarrow \neg \exists x\neg B$  and  $\neg \forall xB \rightarrow \exists x\neg B$ . Suppose  $Q$  is  $\exists$ . Let  $a$  be any individual symbol not appearing in  $B$ . Then

$\vdash B_a^x \leftrightarrow \neg \neg B_a^x$	Tautology Theorem
$\vdash \exists xB \leftrightarrow \exists x\neg \neg B$	distribution rule
$\vdash \neg \forall x\neg B \leftrightarrow \exists x\neg \neg B$	part (ii), where $Q$ is $\forall$
$\vdash \neg \exists xB \leftrightarrow \forall x\neg B$	tautology rule

(iii) We shall give a proof for the case where  $Q$  is  $\forall$ . The other case can be proved similarly. The proof hinges on the fact that for any term  $t$ ,  $[B_y^x \vee C]_t^y$  is  $B_t^x \vee C$ . Let  $t$  be the term  $\epsilon y \neg (B_y^x \vee C)$  and  $s$  the term  $\epsilon x \neg B$ .

$\vdash \forall y(B_y^x \vee C) \rightarrow B_s^x \vee C$	} Theorem II.10(i)
$\vdash \forall xB \rightarrow B_t^x$	
$\vdash B_s^x \rightarrow \forall xB$	} Theorem II.10(ii) and
$\vdash B_t^x \vee C \rightarrow \forall y(B_y^x \vee C)$	
$\vdash (\forall xB \vee C) \leftrightarrow \forall y(B_y^x \vee C)$	tautology rule

(iv) The proof is similar to that of (iii).

### 7.3 The equivalence rule

In this section we establish a derived rule of inference, the equivalence rule, which asserts in effect that within a given formula any expression may be replaced by an equivalent expression. Before giving a precise statement of this rule we first introduce the following unifying notation. Suppose  $E_1$  and  $E_2$  are any two quasi-terms or any two quasi-formulae. The notation  $E_1 \equiv E_2$  is used to denote the quasi-formula  $E_1 = E_2$  if  $E_1$  and  $E_2$  are both quasi-terms, and the quasi-formula  $E_1 \leftrightarrow E_2$ , if they are both quasi-formulae.

**THEOREM II.21** (The equivalence rule). *Let  $X$  be any set of formulae,  $E$  and  $E'$  any two quasi-terms or any two quasi-formulae, and  $A$  any formula or term containing some specified occurrence of  $E$ . Let  $A'$  be the expression obtained from  $A$  by replacing this specified occurrence of  $E$  by  $E'$ . If  $A'$  is a term or formula and if  $X \vdash \forall [E \equiv E']$ , then  $X \vdash A \equiv A'$ .*

*Proof.* The proof is by induction on the length of  $A$ . (We may assume without loss of generality that  $X$  is a finite set.)

*Case 1.* The specified occurrence of  $E$  in  $A$  is  $A$  itself: Then  $A'$  is  $E'$  and  $X \vdash A \equiv A'$  by the hypothesis.

We now assume that Case 1 does not apply. Consequently one of the following cases must apply.

*Case 2.*  $A$  is of the form  $Ps_1 \dots s_n$  where  $P$  is an  $n$ -place function or predicate symbol: Then the specified occurrence of  $E$  in  $A$  must be contained within some  $s_i$ . Hence  $A'$  is of the form  $Ps_1 \dots s_{i-1}s'_i s_{i+1} \dots s_n$ , and  $X \vdash s_i = s'_i$  by the induction hypothesis. Let  $B$  be the quasi-formula

$$Ps_1 \dots s_n \equiv Ps_1 \dots s_{i-1}xs_{i+1} \dots s_n.$$

Then  $B_{s_i}^x$  is  $A \equiv A$  and  $B_{s'_i}^x$  is  $A'$ . We prove  $X \vdash B_{s_i}^x$ , as follows:

$X \vdash s_i = s'_i$	induction hypothesis
$\vdash B_{s_i}^x$	E3-axiom or Tautology Theorem
$X \vdash (s_i = s'_i \wedge B_{s_i}^x) \rightarrow B_{s'_i}^x$	E1-axiom
$X \vdash B_{s'_i}^x$	tautology rule

*Case 3.*  $A$  is of the form  $s = t$ : The proof is similar to that of Case 2.

*Case 4.*  $A$  is of the form  $\neg B$ : Then  $A'$  is of the form  $\neg B'$ , and  $X \vdash B \leftrightarrow B'$  by the induction hypothesis. Hence  $X \vdash \neg B \leftrightarrow \neg B'$  by the tautology rule.

*Case 5.*  $A$  is of the form  $B \wedge C$ ,  $B \vee C$ , or  $B \rightarrow C$ : The proof is similar to that of Case 4.

*Case 6.*  $A$  is of the form  $\exists xB$ ,  $\forall xB$ , or  $\epsilon xB$ : Then  $A'$  is of the form  $\exists xB'$ ,  $\forall xB'$ , or  $\epsilon xB'$ , respectively, where  $B'$  is obtained from  $B$  by replacing some occurrence of  $E$  in  $B$  by  $E'$ . Let  $a$  be some individual symbol not appearing in  $B$ ,  $B'$ , or any member of  $X$ . It is sufficient to prove  $X \vdash B_a^x \leftrightarrow B'_a^x$  since the desired result then follows by the distribution rule. If no free occurrence of  $x$  in  $B$  lies within the specified occurrence of  $E$ , then  $B_a^x$  contains this occurrence of  $E$  and  $B'_a^x$  is obtained from  $B_a^x$  by replacing this occurrence by  $E'$ . Hence  $\vdash B_a^x \leftrightarrow B'_a^x$  by the induction hypothesis. On the other hand, suppose a free occurrence of  $x$  in  $B$  does lie within the specified occurrence of  $E$  in  $B$ . Then  $B'_a^x$  is obtained from  $B_a^x$  by replacing an occurrence of  $E_a^x$  by  $E'_a^x$ . Since  $X \vdash \forall [E \equiv E']$ , then  $X \vdash \forall [E_a^x \equiv E'_a^x]$  by the  $\forall$ -elimination rule, and therefore by the induction hypothesis  $X \vdash B_a^x \leftrightarrow B'_a^x$ .

**THEOREM II.22.** *Any formula of the form  $(s = t \wedge A_s^x) \rightarrow A_t^x$  is a theorem of the  $\epsilon$ -calculus.*

*Proof.* Let  $n$  be the number of free occurrences of  $x$  in  $A$ . Then  $A_t^x$  is obtained from  $A_s^x$  by replacing  $n$  occurrences of  $s$  in  $A_s^x$  by  $t$ . Since  $s = t \vdash s = t$ , then by  $n$  applications of the equivalence rule and the tautology rule we have  $s = t \vdash A_s^x \leftrightarrow A_t^x$ . (If  $n = 0$ , then  $A_s^x \leftrightarrow A_t^x$  is a tautology.) Consequently,  $\vdash s = t \rightarrow (A_s^x \leftrightarrow A_t^x)$  by the  $\rightarrow$ -introduction rule, and  $\vdash (s = t \wedge A_s^x) \rightarrow A_t^x$  by the tautology rule.

### 7.4 Rule of relabelling bound variables

Let  $A$  be any formula. If some well-formed part of  $A$  of the form  $\epsilon xB$ ,  $\exists xB$ , or  $\forall xB$  is replaced by  $\epsilon y[B]_y^x$ ,  $\exists y[B]_y^x$ , or  $\forall y[B]_y^x$ , respectively, where  $y$  is not free in  $B$  and is free for  $x$  in  $B$ , then the resulting formula is said to be obtained from  $A$  by an *admissible relabelling of a bound variable*. A formula  $A'$  is said to be a *variant* of  $A$  if there exists a sequence  $A_1, \dots, A_n$  of formulae such that  $A_1$  is  $A$ ,  $A_n$  is  $A'$  and for each  $i = 2, \dots, n$ ,  $A_i$  is obtained from  $A_{i-1}$  by an admissible relabelling of a bound variable.

THEOREM II.23 *If  $A'$  is a variant of  $A$ , then  $\vdash A \leftrightarrow A'$ .*

*Proof.* Clearly, it is sufficient to prove the theorem for the case where  $A'$  is obtained from  $A$  by a single admissible relabelling of a bound variable. In this case  $\vdash A \leftrightarrow A'$  by the equivalence rule, since by Theorem II.20(i)  $\vdash \forall[\varepsilon x B = \varepsilon y[B]_y^x]$ ,  $\vdash \forall[\exists x B \leftrightarrow \exists y[B]_y^x]$ , and  $\vdash \forall[\forall x B \leftrightarrow \forall y[B]_y^x]$ , if  $y$  is free for  $x$  in  $B$  and not free in  $B$ .

COROLLARY (The rule of relabelling bound variables). *If  $X \vdash A$  and  $A'$  is a variant of  $A$ , then  $X \vdash A'$ .*

### 8 Prenex formulae

A *prenex formula*  $A$  is a formula of the form  $Q_1 x_1 \dots Q_n x_n B$  where  $n \geq 0$ , each  $Q_i$  is either  $\exists$  or  $\forall$ , the  $x_i$  are all distinct variables, and  $B$  is an elementary quasi-formula. (Recall that a quasi-formula is elementary if the symbols  $\forall$ ,  $\exists$ , and  $\varepsilon$  do not occur in it.) The expression  $Q_1 x_1 \dots Q_n x_n$  is called the *prefix* of  $A$  and the quasi-formula  $B$  is called the *matrix* of  $A$ . Since our definition of a prenex formula includes the possibility that the prefix is empty, it follows that any elementary formula  $A$  is a prenex formula, and in this case the matrix of  $A$  is  $A$  itself.

THEOREM II.24. *For any  $\varepsilon$ -free formula  $A$ , there exists a prenex formula  $A'$  such that  $\vdash A \leftrightarrow A'$ .*

*Proof.* We first convert  $A$  into a formula  $A_1$  which contains no occurrences of the symbols  $\wedge$  or  $\rightarrow$ . This can be done by replacing those quasi-formulae in  $A$  of the form  $B \wedge C$  by  $\neg(\neg B \vee \neg C)$  and those of the form  $B \rightarrow C$  by  $\neg B \vee C$ . Since  $\vdash \forall[(B \wedge C) \leftrightarrow \neg(\neg B \vee \neg C)]$  and  $\vdash \forall[(B \rightarrow C) \leftrightarrow (\neg B \vee C)]$  by the Tautology Theorem and Theorem II.19(ii), then  $\vdash A \leftrightarrow A_1$  by repeated applications of the equivalence rule. We can now convert  $A_1$  into a prenex formula  $A'$  by successively replacing those quasi-formulae in  $A_1$  of the form  $\neg Qx B$  by  $Q'x \neg B$  and those of the form  $Qx B \vee C$  or  $C \vee Qx B$  by  $Qy(B_y^x \vee C)$  or  $Qy(C \vee B_y^x)$ , respectively, where  $y$  is some variable which does not occur free in  $B$  or  $C$  and which is free for  $x$  in  $B$ . By Theorem II.20 and by repeated applications of the equivalence rule, we have  $\vdash A_1 \leftrightarrow A'$ . Hence  $\vdash A \leftrightarrow A'$  by the tautology rule.

### EXERCISE

Let  $A$  be the formula  $\forall x \exists y Pxy \rightarrow \exists y \forall x Pxy$ . Convert  $A$  into a prenex formula  $A'$  such that  $\vdash A \leftrightarrow A'$  and the prefix of  $A'$  is  $\exists x_1 \exists x_2 \forall x_3 \forall x_4$ .

### 9 The addition of new function symbols

We know by the  $\forall$ -elimination rule and the generalization rule that

$$\vdash \forall y B \quad \text{iff} \quad X \vdash B_y^a$$

provided that  $a$  does not occur in  $B$  or in any member of  $X$ . The following theorem is an interesting and useful extension of this result.

THEOREM II.25. *Let  $X$  be any set of formulae of  $\mathcal{L}(\mathcal{V})$  and  $A$  any formula of  $\mathcal{L}(\mathcal{V})$  of the form  $\exists x_1 \dots \exists x_n \forall y B$ , where  $n \geq 0$  and the variables  $x_1, \dots, x_n$ , and  $y$  are all distinct. Let  $\mathcal{V}_1$  be the vocabulary obtained from  $\mathcal{V}$  by adjoining a new  $n$ -place function symbol  $g$ . Then*

$$X \vdash_{\varepsilon(\mathcal{V})} \exists x_1 \dots \exists x_n \forall y B \quad \text{iff} \quad X \vdash_{\varepsilon(\mathcal{V}_1)} \exists x_1 \dots \exists x_n [B]_{g x_1 \dots x_n}^y$$

*Proof.* To simplify the notation we shall write  $\vdash$  instead of  $\vdash_{\varepsilon(\mathcal{V})}$  and  $\vdash_1$  instead of  $\vdash_{\varepsilon(\mathcal{V}_1)}$ . As a further notational simplification we shall prove the theorem for the case where  $n = 1$ . However, the method of proof is completely general.

First, assume  $X \vdash \exists x \forall y B$ . Since every formula of  $\mathcal{L}(\mathcal{V})$  is a formula of  $\mathcal{L}(\mathcal{V}_1)$  and every axiom of  $\varepsilon(\mathcal{V})$  is an axiom of  $\varepsilon(\mathcal{V}_1)$ , then  $X \vdash_1 \exists x \forall y B$ . By the  $\exists$ -elimination rule this yields  $X \vdash_1 \forall y [B]_s^x$  where  $s$  is  $\varepsilon x \forall y B$ , and by the  $\forall$ -elimination rule we then get  $X \vdash_1 B_s^x y_s^x$ . However,  $B_s^x y_s^x$  is  $B_{g s}^x y_s^x$ . Hence the  $\exists$ -introduction rule yields  $X \vdash \exists x [B]_{g s}^y$ .

Secondly, let  $\langle A_1, \dots, A_n \rangle$  be a deduction of  $\exists x [B]_{g s}^y$  from  $X$  in  $\varepsilon(\mathcal{V}_1)$ . Let  $\varepsilon y' \neg B'$  be a quasi-term obtained from  $\varepsilon y \neg B$  by replacing each bound variable in  $\varepsilon y \neg B$  by some variable which does not occur (either free or bound) in any of the  $A_i$ . For each  $i$ , let  $A_i'$  be the formula obtained from  $A_i$  by replacing each quasi-term in  $A_i$  of the form  $gt$ , for some  $t$ , by the quasi-term  $[\varepsilon y' \neg B']_t^x$ . (Note that by our relabelling of the bound variables in  $\varepsilon y \neg B$ ,  $t$  is free for  $x$  in  $\varepsilon y' \neg B'$ .) We can now prove by induction that for each  $i$ ,  $X \vdash A_i'$ . For, if  $A_i$  is a member of  $X$ , then  $A_i'$  is  $A_i$ . If  $A_i$  is an axiom of  $\varepsilon(\mathcal{V}_1)$ , other than an EI-axiom, then  $A_i'$  is an axiom of  $\varepsilon(\mathcal{V})$ . If  $A_i$  is an EI-axiom, then  $A_i'$  is a formula of  $\mathcal{L}(\mathcal{V})$  of the same form and therefore a theorem of  $\varepsilon(\mathcal{V})$  by virtue of Theorem II.22. Finally, if  $A_i$  follows by modus ponens, then so also does  $A_i'$ . Since  $A_n'$  is the formula  $\exists x [B]_{\varepsilon y' \neg B'}^y$ , we have  $X \vdash \exists x [B]_{\varepsilon y' \neg B'}^y$ . The rule of relabelling bound variables now yields  $X \vdash \exists x [B]_{\varepsilon y \neg B}^y$ , and the desired result  $X \vdash \exists x \forall y B$  follows by applications of the  $\exists$ -elimination rule,  $\forall$ -introduction, and  $\exists$ -introduction rules. (*Note:* it is implicit in the statement of the theorem that  $g x$  is free for  $y$  in  $B$ . Hence  $\varepsilon y \neg B$  is free for  $y$  in  $B$ , and therefore for any term  $s$ ,  $[B]_{\varepsilon y \neg B}^s$  is  $[B]_s^y \varepsilon y \neg [B]_s^x$ .)

THEOREM II.26. *Let  $X$  be any set of formulae of  $\mathcal{L}(\mathcal{V})$ ,  $C$  any formula of  $\mathcal{L}(\mathcal{V})$ , and  $A$  any formula of  $\mathcal{L}(\mathcal{V})$  of the form  $\forall x_1 \dots \forall x_n \exists y B$ , where  $n \geq 0$  and the variables  $x_1, \dots, x_n$  and  $y$  are all distinct. Let  $\mathcal{V}_1$  be the vocabulary obtained from  $\mathcal{V}$  by adjoining a new  $n$ -place function symbol  $g$ . Then*

$$X, \forall x_1 \dots \forall x_n \exists y B \vdash_{\varepsilon(\mathcal{V})} C \quad \text{iff} \quad X, \forall x_1 \dots \forall x_n [B]_{g x_1 \dots x_n}^y \vdash_{\varepsilon(\mathcal{V}_1)} C.$$



*Proof.* Using the same notational simplifications as above, we want to prove

$$X, \forall x \exists y B \vdash C \quad \text{iff} \quad X, \forall x [B]_{gx}^y \vdash_1 C.$$

By Theorem II.25, we have

$$X, \neg C \vdash \exists x \forall y \neg B \quad \text{iff} \quad X, \neg C \vdash_1 \exists x \neg [B]_{gx}^y.$$

However, Theorem II.20(ii) and the equivalence rule yield

$$\vdash \exists x \forall y \neg B \leftrightarrow \neg \forall x \exists y B$$

and

$$\vdash \exists x \neg [B]_{gx}^y \leftrightarrow \neg \forall x [B]_{gx}^y.$$

Hence, by the tautology rule

$$X, \neg C \vdash \neg \forall x \exists y B \quad \text{iff} \quad X, \neg C \vdash_1 \neg \forall x [B]_{gx}^y.$$

The desired result now follows by the  $\rightarrow$ -elimination,  $\rightarrow$ -introduction, and contrapositive rules.

Theorem II.26 provides a formal justification in terms of the  $\varepsilon$ -calculus of a type of reasoning which is commonly used in mathematics. Suppose that we are trying to prove some statement  $C$  and in the course of the proof we prove a statement of the form

(1) 'for all  $x$ , there exists a  $y$  such that  $B(x,y)$ '

where  $B(x,y)$  asserts some relationship between  $x$  and  $y$ . It is often convenient to have, for each  $x$ , a way of denoting some  $y$  such that  $B(x,y)$ . Since the notation must express the fact that  $y$  depends on  $x$ , we introduce a new function symbol  $g$  and say

(2) 'for all  $x$ ,  $B(x,g(x))$ '

thus using  $g(x)$  to denote an appropriate  $y$ . Theorem II.26 shows that if we can deduce  $C$  using statement (2), then we can deduce  $C$  directly from statement (1) without using the function  $g$ . Of course our justification of this line of reasoning makes use of the  $\varepsilon$ -symbol and the logical power of the  $\varepsilon$ -calculus, since in effect what we have done is to identify the expression  $g(x)$  with the quasi  $\varepsilon$ -term  $\varepsilon y B(x,y)$  and to use the fact that under this identification statements (1) and (2) are equivalent in the logic of the  $\varepsilon$ -calculus. However, once we have proved the Second  $\varepsilon$ -Theorem, it will follow that the above line of reasoning is justifiable even when one is using a more standard system of logic (i.e., the predicate calculus) which does not include the  $\varepsilon$ -symbol. (See Chapter III, §4.1.)

### 9.1 Skolem and Herbrand resolutions of prenex formulae

A prenex formula is  $\exists$ -prenex if the symbol  $\forall$  does not occur in its prefix and  $\forall$ -prenex if the symbol  $\exists$  does not occur in its prefix. In this section we shall assign to any prenex formula  $A$  a certain  $\exists$ -prenex formula, denoted by  $A_H$ , and a certain  $\forall$ -prenex formula, denoted by  $A_S$ . The formula  $A_H$  is

called the *Herbrand resolution* of  $A$ , and the formula  $A_S$  the *Skolem resolution* of  $A$ . Using Theorems II.25 and II.26, we can establish close relationships between  $A_H$  and  $A$  and between  $A_S$  and  $A$ .

Let  $\mathcal{V}$  be any vocabulary and  $A$  any prenex formula of  $\mathcal{L}(\mathcal{V})$ . The *Herbrand resolution*,  $A_H$ , of  $A$  is defined as follows. If  $A$  is already  $\exists$ -prenex, then  $A_H$  is  $A$ . Otherwise,  $A$  is of the form  $\exists x_1 \dots \exists x_n \forall y B$ , where  $n \geq 0$ . Let  $\mathcal{V}'$  be the vocabulary obtained by adjoining a new  $n$ -place function symbol  $g$  to  $\mathcal{V}$ . Let  $A'$  be the formula  $\exists x_1 \dots \exists x_n [B]_{gx_1 \dots x_n}^y$ . If  $A'$  is  $\exists$ -prenex, then  $A_H$  is  $A'$ . Otherwise, repeat the procedure by adding a new function symbol to  $\mathcal{V}'$  and forming the prenex formula  $A''$ . After a finite number of steps a  $\exists$ -prenex formula is obtained. This formula is  $A_H$ . The new function symbols which are added to  $\mathcal{V}$  in forming  $A_H$  are called *Herbrand functions*.

The Skolem resolution,  $A_S$ , of  $A$  is defined similarly. (Thus if  $A$  is of the form  $\forall x_1 \dots \forall x_n \exists y B$ , then  $A'$  is the formula  $\forall x_1 \dots \forall x_n [B]_{gx_1 \dots x_n}^y$ .) The new function symbols which are added to  $\mathcal{V}$  in forming  $A_S$  are called *Skolem functions*. The process whereby  $A$  is converted to  $A_S$  is often referred to as *symbolic resolution* of existential formulae.

For example, suppose  $A$  is the prenex formula

$$\forall x_1 \exists y_1 \forall x_2 \exists y_2 B,$$

where  $B$  is the matrix of  $A$ . Then the formula

$$\exists y_1 \exists y_2 [B]_{g_1 g_2}^{x_1 x_2}$$

is the Herbrand resolution of  $A$ , where  $g_1$  and  $g_2$  are used as Herbrand functions, and the formula

$$\forall x_1 \forall x_2 [B]_{h_1 x_1 h_2 x_1 x_2}^{y_1 y_2}$$

is the Skolem resolution of  $A$ , where  $h_1$  and  $h_2$  are used as Skolem functions.

**THEOREM II.27.** *Let  $A$  be any prenex formula of  $\mathcal{L}(\mathcal{V})$ ,  $X$  any set of formulae of  $\mathcal{L}(\mathcal{V})$ , and  $\mathcal{V}^*$  the vocabulary obtained from  $\mathcal{V}$  by adjoining the Herbrand functions used in forming  $A_H$ . Then  $X \vdash_{\varepsilon(\mathcal{V})} A$  iff  $X \vdash_{\varepsilon(\mathcal{V}^*)} A_H$ .*

*Proof.* The proof follows from Theorem II.25 by induction on the number of occurrences of  $\forall$  in the prefix of  $A$ .

**THEOREM II.28.** *Let  $A$  be any prenex formula of  $\mathcal{L}(\mathcal{V})$ ,  $C$  any formula of  $\mathcal{L}(\mathcal{V})$ ,  $X$  any set of formulae of  $\mathcal{L}(\mathcal{V})$ , and  $\mathcal{V}^*$  the vocabulary obtained from  $\mathcal{V}$  by adjoining the Skolem functions used in forming  $A_S$ . Then*

$$X, A \vdash_{\varepsilon(\mathcal{V})} C \quad \text{iff} \quad X, A_S \vdash_{\varepsilon(\mathcal{V}^*)} C.$$

*Proof.* The proof follows from Theorem II.26 by induction on the number of occurrences of  $\exists$  in the prefix of  $A$ .

## EXERCISE

Let  $A$  be any prenex formula of  $\mathcal{L}(\mathcal{V})$ . Prove (i)  $\vdash_{\varepsilon(\mathcal{V}^*)} A \rightarrow A_H$ , where  $\mathcal{V}^*$  is as in Theorem II.27, and (ii)  $\vdash_{\varepsilon(\mathcal{V}^*)} A_S \rightarrow A$ , where  $\mathcal{V}^*$  is as in Theorem II.28. Find a prenex formula  $A$  such that  $A_H \rightarrow A$  is invalid and therefore not a theorem of  $\varepsilon(\mathcal{V}^*)$ .

## 10 The elementary calculus

Let  $\mathcal{V}$  be any vocabulary. We shall now define two formal systems for  $\mathcal{V}$ , the *elementary calculus* and the *elementary calculus without identity*. The first of these will be denoted by  $EC(\mathcal{V})$  and the second by  $EC'(\mathcal{V})$ . Each may be regarded as a subsystem of  $\varepsilon(\mathcal{V})$ .

Recall that a formula of  $\mathcal{L}(\mathcal{V})$  is *elementary* if it does not contain the symbols  $\forall$ ,  $\exists$ , or  $\varepsilon$ . (Equivalently, a formula is elementary if it contains no variables.) The *axioms* of  $EC(\mathcal{V})$  are the elementary formulae of  $\mathcal{L}(\mathcal{V})$  which are instances of axiom schemata P1–P10, E1, and E3. For any set  $X$  of formulae of  $\mathcal{L}(\mathcal{V})$  and any formula  $A$  of  $\mathcal{L}(\mathcal{V})$ , a *deduction* of  $A$  from  $X$  in  $EC(\mathcal{V})$  is any sequence  $\langle A_1, \dots, A_n \rangle$  of elementary formulae of  $\mathcal{L}(\mathcal{V})$  such that  $A_n$  is  $A$  and for each  $i = 1, \dots, n$ ,  $A_i$  is an axiom of  $EC(\mathcal{V})$ , or  $A_i$  is a member of  $X$ , or  $A_i$  follows by modus ponens from  $A_j$  and  $A_k$  for some  $j, k < i$ . In other words a sequence of formulae of  $\mathcal{L}(\mathcal{V})$  is a deduction in  $EC(\mathcal{V})$  if and only if it is a deduction in  $\varepsilon(\mathcal{V})$  and each of its members is elementary. As usual, we write  $X \vdash_{EC(\mathcal{V})} A$  to denote that there exists a deduction of  $A$  from  $X$  in  $EC(\mathcal{V})$ . Notice that if  $X \vdash_{EC(\mathcal{V})} A$ , then  $A$  must be elementary. Since any deduction in  $EC(\mathcal{V})$  is automatically a deduction in  $\varepsilon(\mathcal{V})$ , then  $X \vdash_{EC(\mathcal{V})} A$  implies  $X \vdash_{\varepsilon(\mathcal{V})} A$ . We shall see later that the converse is also true, i.e.,  $X \vdash_{\varepsilon(\mathcal{V})} A$  implies  $X \vdash_{EC(\mathcal{V})} A$ , provided that  $X$  is a set of elementary formulae and  $A$  is elementary (see page 64).

The formal system  $EC'(\mathcal{V})$  is defined in a similar way except that the axioms of  $EC'(\mathcal{V})$  do not include instances of E1 and E3, and a deduction in  $EC'(\mathcal{V})$  must be a sequence of elementary, *identity-free* formulae.

Since, as usual,  $\mathcal{V}$  is an arbitrary vocabulary, we shall write  $EC$  instead of  $EC(\mathcal{V})$  and  $EC'$  instead of  $EC'(\mathcal{V})$ .

Many of the derived rules of inference which we have proved for the  $\varepsilon$ -calculus also apply to the elementary calculus (without identity). In particular we can prove exactly as for the  $\varepsilon$ -calculus that the  $f$ -rule, contradiction rule,  $\neg\neg$ -rule,  $\alpha$ -rule, and  $\beta$ -rule all hold in  $EC$  and  $EC'$  (assuming, of course, that the letters  $A$ ,  $\alpha$ , and  $\beta$  which appear in the statements of these rules now denote elementary (identity-free) formulae). Consequently, the proofs of Theorems II.13 and II.14 (the Tautology Theorem) can be used verbatim to prove the following theorem.

**THEOREM II.29.** *For any elementary (identity-free) formula  $A$ , if  $A$  is a tautology, then  $A$  is a theorem of the elementary calculus (without identity).*

A theorem of the elementary calculus is not necessarily a tautology, since for example,  $a = a$  is a theorem but not a tautology. Nonetheless, we have the following very useful theorem.

**THEOREM II.30.** *Every identity-free theorem of  $EC$  is a tautology.*

*Proof.* For any formula  $B$ , we write  $h(B)$  to denote the formula which is obtained from  $B$  by replacing every occurrence in  $B$  of a quasi-formula of the form  $s = t$  by  $f$  if  $s$  and  $t$  are not the same quasi-terms and by  $\neg f$  if they are the same.

Now let  $\langle A_1, \dots, A_n \rangle$  be a proof in  $EC$  of some identity-free formula  $A$ . We shall prove by induction that for each  $A_i$ ,  $h(A_i)$  is a tautology, thus proving that  $h(A_n)$ , i.e.  $A$ , is a tautology.

*Case 1.*  $A_i$  is a *propositional axiom*: Then  $h(A_i)$  is also a propositional axiom and therefore a tautology by Exercise 2, page 41.

*Case 2.*  $A_i$  is an *E3-axiom*: Then  $h(A_i)$  is  $\neg f$  which is a tautology.

*Case 3.*  $A_i$  is an *E1-axiom*: Thus  $A_i$  has the form  $(s = t \wedge B_s^x) \rightarrow B_t^x$ . If  $s$  and  $t$  are not the same terms, then  $h(A_i)$  has the form  $(f \wedge C_1) \rightarrow C_2$ , which is a tautology. On the other hand, if  $s$  and  $t$  are the same terms, then  $h(A_i)$  has the form  $(\neg f \wedge C) \rightarrow C$ , which is also a tautology.

*Case 4.*  $A_i$  follows by *modus ponens* from  $A_j$  and  $A_k$ , where  $j, k < i$ : Then  $h(A_i)$  follows from  $h(A_j)$  and  $h(A_k)$  by modus ponens. Hence by the induction hypothesis and Exercise 2, page 18,  $h(A_i)$  is a tautology.

Theorems II.29 and II.30 together imply that for any identity-free formula  $A$ ,  $A$  is a tautology iff  $\vdash_{EC} A$ . Furthermore these theorems provide the interesting, though perhaps not unexpected, information that any identity-free theorem of  $EC$  is a theorem of  $EC'$ . In other words, in the elementary calculus the identity symbol and the equality axioms are superfluous in proving identity-free theorems. We shall return to the general problem of proving the eliminability of the identity symbol in Chapter III, page 83.

## 11 The predicate calculus

We now turn to the standard formalization of logic, the predicate calculus. For any vocabulary  $\mathcal{V}$ , the predicate calculus for  $\mathcal{V}$ , which we shall denote by  $PC(\mathcal{V})$  will have as its deductions certain sequences of  $\varepsilon$ -free formulae. We want to define the notion of a deduction in  $PC(\mathcal{V})$  in such a way that all the derived rules of inference which hold for the  $\varepsilon$ -calculus hold also for the predicate calculus, except of course those rules, such as the  $\exists$ -elimination and  $\forall$ -introduction rules, in which certain  $\varepsilon$ -terms are specified. Since the quanti-

fier rules, such as the rule of generalization,  $\exists$ -rule, and  $\forall$ -rule, depend on the axiom schema

$$Q4 \quad \exists xA \rightarrow A(\epsilon xA)$$

and since instances of this schema cannot be used in deductions in the predicate calculus, we compensate for this deficiency by adopting the  $\exists$ -rule as one of the basic rules of inference of the system. Formally, we state this rule as follows:

A formula of the form  $\exists xA \rightarrow B$  follows by the  $\exists$ -rule from a formula  $C$ , if and only if  $C$  is of the form  $A_a^x \rightarrow B$ , where  $a$  does not occur in  $A$  or  $B$ .

The *axioms* of the predicate calculus for  $\mathcal{V}$  are the  $\epsilon$ -free instances of the axiom schemata P1–P10, Q1–Q3, E1, and E3. Thus the axioms of  $PC(\mathcal{V})$  are simply the  $\epsilon$ -free axioms of  $\mathcal{L}(\mathcal{V})$ .

In defining a deduction in  $PC(\mathcal{V})$  we are faced with the following problem. Suppose we were to proceed in the usual fashion by defining a deduction of  $A$  from  $X$  as any sequence  $\langle A_1, \dots, A_n \rangle$  of  $\epsilon$ -free formulae of  $\mathcal{L}(\mathcal{V})$ , such that  $A_n$  is  $A$  and for each  $i$ ,  $A_i$  is an axiom, or  $A_i$  is a member of  $X$ , or  $A_i$  follows by modus ponens or by the  $\exists$ -rule. Unfortunately, under this definition the  $\rightarrow$ -introduction rule (Deduction Theorem) does not hold in its full generality (cf. Mendelson [1964], pp. 60–61). One standard way of overcoming this deficiency is to stipulate that in each application of the  $\exists$ -rule the specified symbol  $a$  must not occur in any member of  $X$ . However, this restriction is too strong, for although the  $\rightarrow$ -introduction rule now holds, we lose the MP rule as well as the simple rule which asserts that if  $X \subseteq Y$  and  $X \vdash A$ , then  $Y \vdash A$ . Consequently, we adopt the following modified restriction on the  $\exists$ -rule. (Our definition of a deduction is due to Lyndon [1966]. For other suitable definitions as well as a full analysis of the problem see Montague and Henkin [1956].)

First of all we define a 'derivation' in  $PC(\mathcal{V})$ . For any finite set  $Y$  of formulae of  $\mathcal{L}(\mathcal{V})$  and any formula  $A$  of  $\mathcal{L}(\mathcal{V})$  a *derivation* of  $A$  from  $Y$  in  $PC(\mathcal{V})$  is any sequence  $\langle A_1, \dots, A_n \rangle$  of  $\epsilon$ -free formulae of  $\mathcal{L}(\mathcal{V})$ , where  $A_n$  is  $A$  and for each  $i = 1, \dots, n$  at least one of the following conditions holds:

- (i)  $A_i$  is an axiom of  $PC(\mathcal{V})$ ,
- (ii)  $A_i$  is a member of  $Y$ ,
- (iii)  $A_i$  follows by modus ponens from  $A_j$  and  $A_k$  for some  $j, k < i$ ,
- (iv)  $A_i$  follows by the  $\exists$ -rule from some  $A_j$ , where  $j < i$ , provided that the specified individual symbol in  $A_j$  does not occur in any member of  $Y$ .

(Notice that this definition of a derivation corresponds to the second unsuccessful definition of a deduction given above.) We now define a *deduction* of  $A$  from  $X$  in  $PC(\mathcal{V})$  as any derivation of  $A$  from some finite subset  $Y$  of  $X$ . As usual, if there exists a deduction of  $A$  from  $X$  in  $PC(\mathcal{V})$  we write  $X \vdash_{PC(\mathcal{V})} A$ , or simply  $X \vdash_{PC} A$  or  $X \vdash A$  when there is no possibility of

ambiguity. Notice that by defining a deduction in this way we still have the rule which asserts that if  $X \subseteq Y$  and  $X \vdash A$ , then  $Y \vdash A$ .

### 11.1 Derived rules of inference for the predicate calculus

Some of the derived rules of inference, such as the MP rule,  $\rightarrow$ -elimination rule, and  $\rightarrow$ -introduction rule, which were so easy to verify for the  $\epsilon$ -calculus, are more difficult to establish in the case of the predicate calculus. For example, the proof of the  $\rightarrow$ -elimination rule is complicated by the fact that if  $\langle A_1, \dots, A_n \rangle$  is a derivation of  $A \rightarrow B$  from  $X$  in PC, then the sequence  $\langle A_1, \dots, A_n, A, B \rangle$  is not necessarily a derivation of  $B$  from  $X \cup \{A\}$  since an individual symbol occurring in  $A$  may be involved in an application of the  $\exists$ -rule. To overcome this difficulty we need the following theorem.

**THEOREM II.31.** *If  $X$  and  $Y$  are any finite sets of formulae and  $\mathcal{D}$  is a derivation of  $A$  from  $X$ , then there exists a derivation  $\mathcal{D}'$  of  $A$  from  $X \cup Y$ .*

*Proof.* Let  $I$  be the (finite) collection of individual symbols which occur in the members of  $Y$ . We want to prove that there exists a derivation  $\mathcal{D}'$  of  $A$  from  $X$  in which no member of  $I$  is used in an application of the  $\exists$ -rule, since then  $\mathcal{D}'$  is a derivation of  $A$  from  $X \cup Y$ . The proof is by induction on the number  $m$  of members of  $I$  which are so used in  $\mathcal{D}$ . If  $m = 0$ , then  $\mathcal{D}$  itself is the required derivation. Suppose  $m > 0$ . Let  $a$  be some member of  $I$  which is involved in an application of the  $\exists$ -rule in  $\mathcal{D}$  and let  $b$  be an individual symbol not in  $I$  and not occurring in any member of  $X$  or any member of  $\mathcal{D}$ . Suppose  $\mathcal{D}$  is the sequence  $\langle A_1, \dots, A_n \rangle$ . Let  $\mathcal{D}'$  be the sequence  $\langle A_1', \dots, A_{n-1}', A_1, \dots, A_n \rangle$ , where each  $A_i'$  is obtained from  $A_i$  by replacing every occurrence of  $a$  by  $b$ . Then  $\mathcal{D}'$  is a derivation of  $A$  from  $X$  in which only  $m - 1$  members of  $I$  are involved in applications of the  $\exists$ -rule. For, if  $A_i$  follows from  $A_j$  in  $\mathcal{D}$  by an application of the  $\exists$ -rule which involves  $a$ , then the presence of  $A_i$  in  $\mathcal{D}'$  can be justified by applying the  $\exists$ -rule to  $A_j'$ . This application of the  $\exists$ -rule involves the new symbol  $b$ . The desired result now follows by induction.

Using this result we can prove the counterpart of Theorem II.3(ii). (The counterparts of parts (i), (iii), and (iv) are trivial.)

**THEOREM II.32.** *Let  $X$  and  $Y$  be any sets of formulae and  $A, B_1, \dots, B_n$  any  $\epsilon$ -free formulae. If  $Y, B_1, \dots, B_n \vdash A$  and  $X \vdash B_i$  for each  $i = 1, \dots, n$ , then  $X \cup Y \vdash A$ .*

*Proof.* It is sufficient to consider the case where  $n = 1$ . Thus we want to prove that if  $Y, B \vdash A$  and  $X \vdash B$ , then  $X \cup Y \vdash A$ . Let  $\mathcal{D}_1$  be a derivation of  $A$  from  $Y'$  and  $\mathcal{D}_2$  a derivation of  $B$  from  $X'$ , where  $Y'$  is a finite subset of  $Y \cup \{B\}$  and  $X'$  is a finite subset of  $X$ . By Theorem II.31, there exist derivations  $\langle A_1, \dots, A_m \rangle$  of  $B$  from  $X' \cup Y'$  and  $\langle B_1, \dots, B_n \rangle$  of  $A$  from  $X' \cup Y'$ .

Hence the sequence  $\langle A_1, \dots, A_m, B_1, \dots, B_n \rangle$  is a derivation of  $A$  from  $X' \cup (Y' \setminus \{B\})$  and therefore a deduction of  $A$  from  $X \cup Y$ .

### EXERCISE

Prove that the  $\rightarrow$ -elimination rule and MP rule hold for the predicate calculus, where of course the letters  $A, B, B_1, \dots, B_n$  now denote  $\varepsilon$ -free formulae.

**THEOREM II.33.** *For any  $\varepsilon$ -free formula  $A$ , if  $A$  is a tautology, then  $\vdash_{PC} A$ .*

*Proof.* For the purposes of this proof we write  $PC_0(\mathcal{V})$  to denote the formal system which is obtained from  $PC(\mathcal{V})$  by excluding the  $\exists$ -rule. Thus the deductions in  $PC_0(\mathcal{V})$  are simply the deductions in  $\varepsilon(\mathcal{V})$  in which every formula is  $\varepsilon$ -free. For this formal system one can prove the Tautology Theorem just as it was proved for the  $\varepsilon$ -calculus. Hence if  $A$  is an  $\varepsilon$ -free tautology, then  $\vdash_{PC_0} A$  and *a fortiori*  $\vdash_{PC} A$ .

The tautology rule now follows immediately using this theorem and the MP rule.

We shall use the following important result in the next chapter to prove the Second  $\varepsilon$ -Theorem.

**THEOREM II.34.** *If  $\vdash(\exists xA \rightarrow A_a^x) \rightarrow C$ , where  $a$  does not occur in  $A$  or  $C$ , then  $\vdash C$ .*

*Proof.* Assume  $\vdash(\exists xA \rightarrow A_a^x) \rightarrow C$ . Since  $A_a^x \rightarrow C$  is a tautological consequence of  $(\exists xA \rightarrow A_a^x) \rightarrow C$ , we have  $\vdash A_a^x \rightarrow C$  by the tautology rule, and hence  $\vdash \exists xA \rightarrow C$  by an application of the  $\exists$ -rule. However,  $C$  is a tautological consequence of  $(\exists xA \rightarrow A_a^x) \rightarrow C$  and  $\exists xA \rightarrow C$ . Hence  $\vdash C$  by the tautology rule.

### EXERCISES

1. Prove that the  $\rightarrow$ -introduction rule (Deduction Theorem) holds for the predicate calculus.
2. Show that the counterpart of Theorem II.17 holds for the predicate calculus.

### 11.2 The predicate calculus without identity

If there exists a deduction of  $A$  from  $X$  in  $PC(\mathcal{V})$  in which every formula is identity-free we write  $X \vdash_{PC(\mathcal{V})} A$ . Thus  $PC'(\mathcal{V})$  may be regarded as a formal system for  $\mathcal{V}$  whose axioms are the identity-free axioms of  $PC(\mathcal{V})$  and whose deductions are the identity-free deductions of  $PC(\mathcal{V})$ . This formal system for  $\mathcal{V}$  is called the *predicate calculus without identity*. Obviously, all the results

which we proved for  $PC(\mathcal{V})$  in the last section also hold for  $PC'(\mathcal{V})$ , assuming that we restrict our attention to identity-free formulae.

In the next chapter we shall prove that for any identity-free,  $\varepsilon$ -free formula  $A$ , if  $\vdash_{PC} A$ , then  $\vdash_{PC'} A$ . Notice that the technique which we used to prove the analogous result for the elementary calculus cannot be used for the case of the predicate calculus. For, suppose  $A$  is the axiom  $\neg \exists x(x = a) \rightarrow \neg a = a$ . Then using the technique employed in proving Theorem II.30, the identity-free formula which is assigned to  $A$  is the formula  $\neg \exists x f \rightarrow \neg \neg f$ . This latter formula is not a theorem of  $PC'$ .

### EXERCISE

Prove that the formula  $\neg \exists x f \rightarrow \neg \neg f$  is not a theorem of  $PC'$ . (*Hint:* use the technique employed in proving Theorem II.15.)

### 12 The formal superiority of the $\varepsilon$ -calculus

It should now be apparent that the  $\varepsilon$ -calculus is a much simpler and neater formalization of logic than is the predicate calculus.

The deductions in the  $\varepsilon$ -calculus can be defined in a simple straightforward way, and the basic derived rules of inference can be established with very little difficulty. On the other hand, we have seen that no matter how one defines a deduction in the predicate calculus certain complications arise since one must adopt an additional rule of inference for dealing with the quantifiers.

Furthermore, in the  $\varepsilon$ -calculus the derived rules of inference for the quantifiers can be expressed more easily and can be used more conveniently than in the predicate calculus. For example, although it is possible to formulate a derived rule of inference for the predicate calculus which is analogous to the  $\exists$ -introduction rule (cf. rule C in Mendelson [1964], page 74), this rule is subject to many tedious restrictions.

In spite of these formal advantages which are gained by using the  $\varepsilon$ -symbol, the question arises whether the use of such an indeterminate logical symbol is philosophically justified. In the next chapter we shall give the best possible justification for the use of the  $\varepsilon$ -symbol in logic by proving that any  $\varepsilon$ -free theorem of the  $\varepsilon$ -calculus is a theorem of the predicate calculus. In other words, the  $\varepsilon$ -symbol and the axioms associated with it can be eliminated from proofs of  $\varepsilon$ -free formulae.

1 Introduction

In Chapter II we defined three basic formal systems (for a given vocabulary  $\mathcal{V}$ ): the elementary calculus, the predicate calculus, and the  $\varepsilon$ -calculus. Loosely speaking, the predicate calculus is obtained from the elementary calculus by adjoining the quantifiers and the appropriate axioms and rule of inference for dealing with them, and the  $\varepsilon$ -calculus is obtained from the predicate calculus by adjoining the  $\varepsilon$ -symbol and the appropriate axioms for dealing with it (and by excluding the redundant  $\exists$ -rule). Thus by passing from the elementary calculus to the predicate calculus and from the predicate calculus to the  $\varepsilon$ -calculus we obtain stronger and stronger logical systems. One of the main objectives of this chapter is to show that these successive strengthenings of the elementary calculus are completely justified since in each case the new logical symbols and axioms are in a certain sense eliminable. To put it more precisely, we shall prove the following two theorems.

1. If  $A$  is an elementary formula of  $\mathcal{L}(\mathcal{V})$ ,  $X$  is a set of elementary formulae of  $\mathcal{L}(\mathcal{V})$ , and  $X \vdash_{PC(\mathcal{V})} A$ , then  $X \vdash_{\varepsilon C(\mathcal{V})} A$ .
2. If  $A$  is an  $\varepsilon$ -free formula of  $\mathcal{L}(\mathcal{V})$ ,  $X$  is a set of  $\varepsilon$ -free formulae of  $\mathcal{L}(\mathcal{V})$ , and  $X \vdash_{\varepsilon(\mathcal{V})} A$ , then  $X \vdash_{PC(\mathcal{V})} A$ .

The second of these two statements is known as Hilbert's Second  $\varepsilon$ -Theorem and the first is a special case of his First  $\varepsilon$ -Theorem. The full statement of the First  $\varepsilon$ -Theorem is as follows.

- 1'. If  $A$  is any prenex formula of  $\mathcal{L}(\mathcal{V})$ ,  $X$  is any set of prenex formulae of  $\mathcal{L}(\mathcal{V})$ , and  $X \vdash_{PC(\mathcal{V})} A$ , then  $Z \vdash_{\varepsilon C(\mathcal{V})} B_1 \vee \dots \vee B_n$ , where each member of  $Z$  is some substitution instance of the matrix of some member of  $X$  and each of the  $B_i$  is a substitution instance of the matrix of  $A$ .

We shall also see in this chapter that besides providing a formal justification for these extensions of the elementary calculus, the two  $\varepsilon$ -theorems can be used to prove some important results about the predicate calculus, such as Skolem's Theorem (Theorem III.12) and Herbrand's Theorem (Theorem III.14).

As in Chapter II, our proofs will be completely finitary in the sense that whenever we prove that a certain deduction exists, our method of proof will

provide a technique for constructing such a deduction. It should be pointed out that by using a non-constructive, model-theoretic argument the Second  $\varepsilon$ -Theorem follows trivially from the completeness of the predicate calculus. For if  $X \vdash_{\varepsilon} A$ , where  $X$  and  $A$  are  $\varepsilon$ -free, then  $X \vDash A$  by the soundness of the  $\varepsilon$ -calculus (cf. Exercise 3, page 41), and hence  $X \vdash_{PC} A$  by the completeness of the predicate calculus.

Throughout this chapter the vocabulary  $\mathcal{V}$  shall be some arbitrary, but fixed vocabulary.

2.1 The basic problem

Since  $\varepsilon$ -terms play the same role in the  $\varepsilon$ -calculus as individual symbols play in the predicate calculus, it would seem that one could prove the Second  $\varepsilon$ -Theorem by assigning an appropriate individual symbol to each  $\varepsilon$ -term in a deduction  $\mathcal{D}$  and then forming a new deduction  $\mathcal{D}'$  by simply replacing each  $\varepsilon$ -term by its assigned individual symbol and making the necessary alterations so that the  $\exists$ -rule can be used where the Q4-axioms were used in  $\mathcal{D}$ .

However, this procedure is complicated by the following problem. Suppose that  $A_i$  is an axiom of the  $\varepsilon$ -calculus,  $\varepsilon yB$  is some  $\varepsilon$ -term, and  $A_i'$  is the formula obtained from  $A_i$  by replacing each occurrence of  $\varepsilon yB$  by some term  $s$ . Then  $A_i'$  may fail to be an axiom, namely if one of the following cases holds:

- (i)  $A_i$  is the Q4-axiom  $\exists yB \rightarrow B(\varepsilon yB)$ , in which case  $A_i'$  is  $\exists yB \rightarrow B(s)$ ;
- (ii)  $A_i$  is an E2-axiom in which  $\varepsilon yB$  is one of the two specified  $\varepsilon$ -terms;
- (iii)  $A_i$  is a Q3-axiom or Q4-axiom and an occurrence of the specified term in that axiom (i.e.,  $t$  or  $\varepsilon xA$ ) lies within an occurrence of  $\varepsilon yB$ .

The third case is the most troublesome one. Consider the following example. Suppose  $A_i$  is the Q3-axiom

$$(1) \quad \neg \exists x(x = \varepsilon y(y = x)) \rightarrow \neg(t = \varepsilon y(y = t))$$

and  $\varepsilon yB$  is the term  $\varepsilon y(y = t)$ . Notice that an occurrence of the specified term  $t$  of this Q3-axiom lies within an occurrence of  $\varepsilon y(y = t)$ . Now if we replace every occurrence of  $\varepsilon y(y = t)$  in  $A_i$  by the term  $s$ , we obtain the formula

$$(2) \quad \neg \exists x(x = \varepsilon y(y = x)) \rightarrow \neg(t = s).$$

This formula is not an instance of axiom schema Q3.

Notice that case (iii) can only arise if  $A_i$  is of the form  $\neg \exists xA \rightarrow \neg A(t)$  or  $\exists xA \rightarrow A(\varepsilon xA)$  where a free occurrence of  $x$  in  $A$  lies within the scope of an  $\varepsilon$ -symbol. Thus, in the above example, where  $A$  is the quasi-formula

$$x = \varepsilon y(y = x),$$

the second free occurrence of  $x$  lies within the quasi  $\varepsilon$ -term  $\varepsilon y(y = x)$ . Formulae such as (1) above are examples of what we shall now call 'improper' formulae.

## 2.2 Proper and improper formulae

Let  $A$  be any term or formula. If  $t$  is a term which occurs in  $A$ , but is not  $A$  itself, then  $t$  is called a *subterm* of  $A$ . For any formula  $A$ , the *skeleton* of  $A$  is that formula which is obtained from  $A$  by successively replacing each subterm of  $A$  by the individual symbol  $a_1$  (starting with those subterms of maximal length). We shall denote the skeleton of  $A$  by  $A^+$ . A formula  $A$  is *proper* if its skeleton is  $\varepsilon$ -free, and *improper* if its skeleton is not  $\varepsilon$ -free.

For example, suppose  $A$  is the formula

$$\neg \exists x(x = \varepsilon y(y = s)) \rightarrow \neg(t = \varepsilon y(y = s)),$$

where  $s$  and  $t$  are any two terms. Then the skeleton of  $A$  is the formula

$$\neg \exists x(x = a_1) \rightarrow \neg(a_1 = a_1).$$

Since this latter formula is  $\varepsilon$ -free, then  $A$  is proper. On the other hand, suppose  $A$  is the formula

$$\neg \exists x(x = \varepsilon y(y = x)) \rightarrow \neg(t = \varepsilon y(y = t)).$$

Then the skeleton of  $A$  is the formula

$$\neg \exists x(x = \varepsilon y(y = x)) \rightarrow \neg(a_1 = a_1),$$

and therefore, in this case,  $A$  is improper.

Obviously, if  $A_i$  is a proper formula of the form  $\neg \exists x A \rightarrow \neg A(t)$  or of the form  $\exists x A \rightarrow A(\varepsilon x A)$ , then  $x$  does not have a free occurrence in  $A$  within the scope of an  $\varepsilon$ -symbol. Consequently, if  $A_i$  is a proper axiom of the  $\varepsilon$ -calculus and  $A_i'$  is obtained from  $A_i$  by replacing every occurrence of  $\varepsilon y B$  by some term  $s$ , then the troublesome case (iii) considered above cannot arise. Therefore, we have the following theorem.

**THEOREM III.1.** *Let  $\varepsilon y B$  be any  $\varepsilon$ -term,  $s$  any term, and  $A$  any axiom of the  $\varepsilon$ -calculus other than an E2-axiom or the Q-axiom  $\exists y B \rightarrow B(\varepsilon y B)$ . If  $A$  is a proper formula, then the formula obtained from  $A$  by replacing each occurrence of  $\varepsilon y B$  by  $s$  is an axiom.*

## 2.3 The $\varepsilon^*$ -calculus

We shall now use the new notion of a proper formula to define a certain formal system called the  $\varepsilon^*$ -calculus (for  $\mathcal{V}$ ). We then prove quite easily that the Second  $\varepsilon$ -Theorem holds for this weakened version of the  $\varepsilon$ -calculus.

The *axioms* of the  $\varepsilon^*$ -calculus (for  $\mathcal{V}$ ) are all *proper* formulae of  $\mathcal{L}(\mathcal{V})$  which are instances of axiom schemata P1–P10, Q1–Q4, E1, and E3. (Note that we exclude E2-axioms.) A *deduction* of  $A$  from  $X$  in the  $\varepsilon^*$ -calculus (for  $\mathcal{V}$ ) is any sequence  $\langle A_1, \dots, A_n \rangle$  of proper formulae (of  $\mathcal{L}(\mathcal{V})$ ) such that  $A_n$  is  $A$  and for each  $i = 1, \dots, n$  either  $A_i$  is an axiom, or  $A_i$  is a member of

$X$ , or  $A_i$  follows by modus ponens from some  $A_j$  and  $A_k$ , where  $j, k < i$ . In other words, a deduction in the  $\varepsilon^*$ -calculus is a deduction in the  $\varepsilon$ -calculus in which every formula is proper and no instances of axiom schema E2 are used as axioms. As usual we write  $X \vdash_{\varepsilon^*} A$  to denote that there exists a deduction of  $A$  from  $X$  in the  $\varepsilon^*$ -calculus.

Obviously the arguments used in proving the  $\rightarrow$ -introduction rule,  $\rightarrow$ -elimination rule, and tautology rule for the  $\varepsilon$ -calculus apply equally well in the case of the  $\varepsilon^*$ -calculus. Furthermore Theorem II.17 can be proved for the  $\varepsilon^*$ -calculus just as it was proved for the  $\varepsilon$ -calculus, since if  $A_i$  is a proper formula and  $A_i'$  is the formula obtained from  $A_i$  by replacing each occurrence of  $a$  by some term  $t$ , then  $A_i'$  is also proper. Consequently, the  $\exists$ -rule holds for the  $\varepsilon^*$ -calculus. In fact, the only derived rules of inference which hold for the  $\varepsilon$ -calculus but not for the  $\varepsilon^*$ -calculus are those rules, such as the distribution rule, rule of relabelling bound variables, and equivalence rule, which depend on axiom schema E2. Even so, these rules still hold for the  $\varepsilon^*$ -calculus if one restricts one's attention to  $\varepsilon$ -free formulae.

In view of the fact that the  $\exists$ -rule holds for the  $\varepsilon^*$ -calculus, we have the following theorem.

**THEOREM III.2.** *If  $X \vdash_{PC} A$ , then  $X \vdash_{\varepsilon^*} A$ .*

*Proof.* Assume  $X \vdash_{PC} A$ . Then there exists a derivation  $\langle A_1, \dots, A_n \rangle$  of  $A$  from  $Y$  in the predicate calculus, where  $Y$  is some finite subset of  $X$ . It follows trivially by induction that, for each  $i = 1, \dots, n$ ,  $Y \vdash_{\varepsilon^*} A_i$ . Hence  $X \vdash_{\varepsilon^*} A$ .

We now prove the converse of Theorem III.2. This result may be regarded as a weaker form of the Second  $\varepsilon$ -Theorem.

**THEOREM III.3.** *For any  $\varepsilon$ -free  $X$  and  $A$ , if  $X \vdash_{\varepsilon^*} A$ , then  $X \vdash_{PC} A$ .*

By appealing to the  $\rightarrow$ -introduction rule for the  $\varepsilon^*$ -calculus and the  $\rightarrow$ -elimination rule for the predicate calculus, the proof of Theorem III.3 can be reduced to proving the following lemma.

**LEMMA.** *If  $\mathcal{D}$  is a proof in the  $\varepsilon^*$ -calculus of some  $\varepsilon$ -free formula  $C$ , then  $\mathcal{D}$  can be converted into a proof of  $C$  in the predicate calculus.*

*Proof.* The proof is by induction on the number  $n$  of distinct  $\varepsilon$ -terms occurring in  $\mathcal{D}$ . If  $n = 0$ , then  $\mathcal{D}$  is already a proof in the predicate calculus.

Suppose  $n > 0$ . Let  $\varepsilon y B$  be an  $\varepsilon$ -term occurring in  $\mathcal{D}$  such that the length of  $\varepsilon y B$  is less than or equal to the length of every other  $\varepsilon$ -term occurring in  $\mathcal{D}$ . Hence no  $\varepsilon$ -terms occur in  $B$ . Let  $a$  be some individual symbol which does not occur in  $B$  or  $C$ , and let  $\mathcal{D}'$  be the sequence of formulae obtained from  $\mathcal{D}$  by replacing each occurrence of  $\varepsilon y B$  by  $a$ . If the formula

$$(1) \quad \exists y B \rightarrow B(\varepsilon y B)$$

is not used as an axiom in  $\mathcal{D}$ , then by Theorem III.1  $\mathcal{D}'$  is a proof of  $C$  in the  $\varepsilon^*$ -calculus. Since only  $n - 1$  distinct  $\varepsilon$ -terms occur in  $\mathcal{D}'$ , then by the induction hypothesis there exists a proof of  $C$  in the predicate calculus. Suppose on the other hand that the formula (1) is used as an axiom in  $\mathcal{D}$ . Then by Theorem III.1,  $\mathcal{D}'$  is a deduction in the  $\varepsilon^*$ -calculus of  $C$  from the formula

$$(2) \quad \exists y B \rightarrow B(a).$$

By the  $\rightarrow$ -introduction rule  $\mathcal{D}'$  can be converted into a proof  $\mathcal{D}''$  in the  $\varepsilon^*$ -calculus of the formula

$$(3) \quad (\exists y B \rightarrow B(a)) \rightarrow C.$$

Since  $C$  is  $\varepsilon$ -free,  $B$  contains no  $\varepsilon$ -terms, and (1) is proper, it follows that (3) is  $\varepsilon$ -free. Furthermore only  $n - 1$  distinct  $\varepsilon$ -terms occur in  $\mathcal{D}''$ . Consequently, by the induction hypothesis there exists a proof of (3) in the predicate calculus. By our choice of  $a$  and by Theorem II.34, this implies that there exists a proof of  $C$  in the predicate calculus.

#### 2.4 The usefulness of the $\varepsilon^*$ -calculus

The  $\varepsilon^*$ -calculus is considerably weaker than the  $\varepsilon$ -calculus since (i) the E2-axioms are excluded and (ii) only proper formulae may be used in a deduction. Nonetheless, this weaker system can play a very useful role in the study of mathematical logic. For example, in order to prove various results about the predicate calculus, such as the  $\rightarrow$ -introduction rule (Deduction Theorem), generalization rule, substitution rule, distribution rule, equivalence rule, rule of relabelling bound variables, etc., it is perhaps easier to show that these results hold in the  $\varepsilon^*$ -calculus (subject to certain conditions) and then use Theorems III.2 and III.3 to prove that they hold in the predicate calculus.

It has been suggested (Fraenkel and Bar-Hillel [1958] page 184) that the  $\varepsilon$ -symbol be used in teaching elementary logic. However, if a full analysis of the  $\varepsilon$ -calculus together with a proof of the Second  $\varepsilon$ -Theorem seems too ambitious for an introductory course in logic, one could easily restrict one's attention to the  $\varepsilon^*$ -calculus. In this case one would modify the rules of formation of the formal languages by stating that an expression of the form  $\varepsilon x A$  is well-formed provided that no variable other than  $x$  is free in  $A$ . In this way all the formulae of the language would be proper formulae. This is essentially the method which is used by Shoenfield [1967] (page 46) although the  $\varepsilon$ -symbol is not explicitly mentioned. Shoenfield shows that this approach provides a simple proof of the completeness of the predicate calculus. Furthermore, we shall see in a later section that in order to prove the First  $\varepsilon$ -Theorem one may use the  $\varepsilon^*$ -calculus and avoid the  $\varepsilon$ -calculus altogether.

Because of the exclusion of the E2-axioms from the  $\varepsilon^*$ -calculus, the equivalence rule does not hold for this system. However, we can still prove the following weaker result.

**THEOREM III.4** (The equivalence rule for the  $\varepsilon^*$ -calculus). *Let  $X$  be any set of proper formulae,  $E$  and  $E'$  any two quasi-terms or quasi-formulae, and  $A$  any proper formula or term containing some specified occurrence of  $E$ , provided that this occurrence does not lie within the scope of an  $\varepsilon$ -symbol in  $A$ . Let  $A'$  be the expression obtained from  $A$  by replacing this specified occurrence of  $E$  by  $E'$ . If  $A'$  is a proper formula or term and if  $X \vdash_{\varepsilon^*} \forall [E \equiv E']$ , then  $X \vdash_{\varepsilon^*} A \equiv A'$*

*Proof.* The proof is identical to the proof of Theorem II.21 except that we can omit the third part of Case 6, where  $A$  is of the form  $\varepsilon x B$ . In this way we avoid using E2-axioms.

**THEOREM III.5.** *Any proper formula of the form*

$$(s = t \wedge A_x^x) \rightarrow A_t^x$$

*is a theorem of the  $\varepsilon^*$ -calculus, provided that  $x$  does not have a free occurrence in  $A$  within the scope of an  $\varepsilon$ -symbol.*

*Proof.* See the proof of Theorem II.22.

**COROLLARY.** *Any  $\varepsilon$ -free formula of the form*

$$(s = t \wedge A_x^x) \rightarrow A_t^x$$

*is a theorem of the predicate calculus.*

#### 3.1 Subordination and rank

In view of Theorem III.3, to complete the proof of the Second  $\varepsilon$ -Theorem it is sufficient to prove that for any  $\varepsilon$ -free  $X$  and  $A$ , if  $X \vdash_{\varepsilon} A$ , then  $X \vdash_{\varepsilon^*} A$ . In other words we must prove the eliminability of improper formulae and E2-axioms from deductions of  $A$  from  $X$  in the  $\varepsilon$ -calculus, where  $A$  and  $X$  are  $\varepsilon$ -free. This is by far the most difficult part of the proof of the Second  $\varepsilon$ -Theorem, and in order to carry it out we must examine the structure of improper formulae in some detail.

The notion of an improper formula is closely related to Hilbert and Bernays' notion of the subordination of quasi  $\varepsilon$ -terms. Hilbert and Bernays [1939], page 23, define a quasi  $\varepsilon$ -term  $t$  to be subordinate (*untergeordnet*) to another quasi  $\varepsilon$ -term  $\varepsilon y B$  if and only if  $B$  contains  $t$ , and a free occurrence of  $y$  in  $B$  lies within  $t$ . Thus, for example, the quasi  $\varepsilon$ -term

$$\varepsilon x Pxy$$

is subordinate to the term

$$\varepsilon y (y = \varepsilon x Pxy).$$

We shall extend this notion of subordination to include subordination within quasi-formulae. If  $A$  is a quasi-formula of the form  $QyB$  or a quasi-term of the form  $\varepsilon yB$ , then a quasi  $\varepsilon$ -term  $t$  is said to be *subordinate to  $A$*  if and only if  $B$  contains  $t$ , and a free occurrence of  $y$  in  $B$  lies within  $t$ . Clearly, a formula  $QyB$  is improper if it contains a quasi  $\varepsilon$ -term which is subordinate to it. For example, the quasi  $\varepsilon$ -term

$$\varepsilon xPxy$$

is subordinate to the (improper) formula

$$\forall y(y = \varepsilon xPxy).$$

We shall now assign to every  $\varepsilon$ -term  $t$  a positive integer, called the 'rank' of  $t$ . We shall denote this number by  $rk(t)$ . Intuitively, the rank of  $t$  is a measure of the complexity of the subordination within  $t$ . Before giving a definition of the rank function, we list the four properties which this function must possess.

*Properties of rank:*

- R1.  $rk(t) \geq 1$ , and if there are no quasi  $\varepsilon$ -terms subordinate to  $t$ , then  $rk(t) = 1$ .
- R2. If  $t'$  is obtained from  $t$  by replacing every occurrence of some subterm of  $t$  by some other term, then  $rk(t') = rk(t)$ .
- R3. If  $t$  is of the form  $[t_1]_s^x$  for some term  $s$ , and  $t_1$  is subordinate to  $\varepsilon xA$ , then  $rk(\varepsilon xA) > rk(t)$ .
- R4. For any  $\varepsilon$ -terms  $\varepsilon xA$  and  $\varepsilon yB$ ,  $rk(\varepsilon z \neg(A_z^x \leftrightarrow B_z^y)) = \max \{rk(\varepsilon xA), rk(\varepsilon yB)\}$ .

In what follows, every argument concerning the rank of an  $\varepsilon$ -term depends only on properties R1–R4, and therefore any definition of the rank function which yields these four properties will suffice. The definition we use is that which is given by Hilbert and Bernays (vol. 2, p. 25).<sup>1</sup> For any quasi  $\varepsilon$ -term  $t$ ,  $rk(t)$  is defined as follows (by induction on the length of  $t$ ):

If there are no quasi  $\varepsilon$ -terms subordinate to  $t$ , then  $rk(t) = 1$ ; otherwise  $rk(t)$  is one greater than the maximal rank of the quasi  $\varepsilon$ -terms which are subordinate to  $t$ .

For example,  $rk(\varepsilon xPxy) = 1$  and  $rk(\varepsilon y(y = \varepsilon xPxy)) = 2$ .

<sup>1</sup> Property R4 is not needed in proving the Second  $\varepsilon$ -Theorem for an  $\varepsilon$ -calculus without axiom schema E2. To prove this weaker theorem the following simpler definition of the rank of  $t$  may be used. Let  $t^0$  be the term obtained from  $t$  by replacing every subterm of  $t$  by the symbol  $a_1$ . Then the *rank* of  $t$  is defined to be the number of occurrences of the  $\varepsilon$ -symbol in  $t^0$ . This definition clearly satisfies R1–R3.

### EXERCISE

Show that the above definition of rank satisfies properties R1–R4.

### 3.2 $r, T_r$ -deductions

Let  $\mathcal{D}$  be any deduction in the  $\varepsilon$ -calculus. An  $\varepsilon$ -term  $\varepsilon yB$  is a *Q-term* in  $\mathcal{D}$  if the formula

$$\exists yB \rightarrow B(\varepsilon yB)$$

or some formula of the form

$$\neg \exists yB \rightarrow \neg B(t)$$

is used as an axiom in  $\mathcal{D}$ . The term  $\varepsilon yB$  is an *E2-term* in  $\mathcal{D}$  if a formula of the form

$$\forall z(A_z^x \leftrightarrow B_z^y) \rightarrow \varepsilon xA = \varepsilon yB$$

or of the form

$$\forall z(B_z^y \leftrightarrow A_z^x) \rightarrow \varepsilon yB = \varepsilon xA$$

is used as an axiom in  $\mathcal{D}$ . For any non-negative integer  $r$ , a deduction  $\mathcal{D}$  in the  $\varepsilon$ -calculus is an  *$r$ -deduction* if every Q-term and every E2-term in  $\mathcal{D}$  has rank  $\leq r$ . Furthermore, if  $T_r$  is some finite collection of  $\varepsilon$ -terms of rank  $r$ , then an  *$r, T_r$ -deduction* is an  $r$ -deduction in which every E2-term of rank  $r$  is a member of  $T_r$ . Notice that since every  $\varepsilon$ -term has rank  $\geq 1$ , a 0-deduction is one in which no instances of axiom schemata Q3, Q4, or E2 are used as axioms, and a 1,  $\emptyset$ -deduction is one in which every Q-term has rank 1 and no instances of axiom schema E2 are used as axioms. To denote that there exists an  $r$ -deduction of  $A$  from  $X$  we write  $X \vdash^r A$ , and to denote that there exists an  $r, T_r$ -deduction of  $A$  from  $X$  we write  $X \vdash^{r, T_r} A$ .

It is easy to see that for any  $r$  and  $T_r$ , the  $r$ -deductions and the  $r, T_r$ -deductions satisfy the MP rule, tautology rule,  $f$ -rule,  $\rightarrow$ -introduction rule, and  $\rightarrow$ -elimination rule, since no Q-axioms or E2-axioms are used in verifying these rules. Thus for example

$$X, \neg A \vdash^{r, T_r} f \text{ iff } X \vdash^{r, T_r} A,$$

$$\text{and } X, A \vdash^{r, T_r} B \text{ iff } X \vdash^{r, T_r} A \rightarrow B.$$

In order to prove the  $\varepsilon$ -Theorems we shall prove the following two results:

- (1) For any  $\varepsilon$ -free  $X$  and any  $r > 1$ , if  $X \vdash^{r, \emptyset} f$ , then  $X \vdash^{r-1} f$ .
- (2) For any  $\varepsilon$ -free  $X$ , any  $r \geq 1$ , and any finite collection  $T_r$  of  $\varepsilon$ -terms of rank  $r$ , if  $X \vdash^{r, T_r} f$ , then  $X \vdash^{r, \emptyset} f$ .

Using these two results (and the  $f$ -rule) it then follows that for any  $\varepsilon$ -free  $X$  and  $\varepsilon$ -free  $A$ , if  $X \vdash_e A$ , then  $X \vdash^{1, \emptyset} A$ .



The proof of the two results depends on the following theorem concerning the way in which an axiom is affected when an  $\varepsilon$ -term occurring in that axiom is replaced by some other term.

**THEOREM III.6.** *Suppose  $\varepsilon yB$  is any  $\varepsilon$ -term,  $s$  any term,  $C$  any axiom, and  $C'$  the formula obtained from  $C$  by replacing each occurrence of  $\varepsilon yB$  by  $s$ . Then:*

- (i) *If  $C$  is a propositional axiom, Q1-axiom, Q2-axiom, E1-axiom, or E3-axiom, then  $C'$  is an axiom of the same form.*
- (ii) *If  $C$  is an E2-axiom and  $\varepsilon yB$  is neither of the two specified  $\varepsilon$ -terms in  $C$ , then  $C'$  is an E2-axiom.*
- (iii) *If  $C$  is the Q3-axiom  $\neg \exists xA \rightarrow \neg A(t)$ , where  $rk(\varepsilon xA) \leq rk(\varepsilon yB)$ , then  $C'$  is a Q3-axiom.*
- (iv) *If  $C$  is the Q4-axiom  $\exists xA \rightarrow A(\varepsilon xA)$ , where  $rk(\varepsilon xA) \leq rk(\varepsilon yB)$  and  $\varepsilon xA$  is not  $\varepsilon yB$ , then  $C'$  is a Q4-axiom.*

*Proof.* The proofs of (i) and (ii) are trivial. (For a discussion of the E1-axioms see page 50.)

(iii). We shall prove that  $C'$  is the Q3-axiom  $\neg \exists xA' \rightarrow \neg A'(t')$ , where  $A'$  and  $t'$  are obtained from  $A$  and  $t$  by replacing each occurrence of  $\varepsilon yB$  by  $s$ . Suppose this is not the case. Then  $\varepsilon yB$  must be of the form  $[p]^x$ , for some quasi-term  $p$  which is subordinate to  $\exists xA$ . Hence  $p$  is subordinate to  $\varepsilon xA$ , and by rank property R3 we have  $rk(\varepsilon yB) < rk(\varepsilon xA)$  which contradicts our assumption that  $rk(\varepsilon xA) \leq rk(\varepsilon yB)$ . In other words our assumption on the ranks of  $\varepsilon xA$  and  $\varepsilon yB$  rules out the type of situation which we illustrated on page 65, where  $C$  was the formula

$$\neg \exists x(x = \varepsilon y(y = x)) \rightarrow \neg(t = \varepsilon y(y = t))$$

and  $\varepsilon yB$  the term  $\varepsilon y(y = t)$ . Note that in this example  $\varepsilon yB$  is  $[\varepsilon y(y = x)]^x$ , and  $\varepsilon y(y = x)$  is subordinate to  $\exists x(x = \varepsilon y(y = x))$ .

(iv). To prove part (iv) we can use the same argument that was employed in (iii) to show that  $C'$  must be the Q4-axiom

$$\exists xA' \rightarrow A'(\varepsilon xA'),$$

where  $A'$  is obtained from  $A$  by replacing each occurrence of  $\varepsilon yB$  by  $s$ .

### 3.3 The Rank Reduction Theorem

**THEOREM III.7** (The Rank Reduction Theorem). *For any set  $X$  of  $\varepsilon$ -free formulae and any  $r > 1$ , if  $X \vdash^{r,0} f$ , then  $X \vdash^{r-1} f$ .*

*Proof.* Let  $\mathcal{D}$  be an  $r, \emptyset$ -deduction of  $f$  from  $X$ . The proof follows by induction on the number,  $p$ , of Q-terms of rank  $r$  in  $\mathcal{D}$ . If  $p = 0$ , then  $\mathcal{D}$  itself is an  $(r-1)$ -deduction of  $f$  from  $X$ . Suppose  $p > 0$ . Let  $\varepsilon yB$  be some Q-term of rank  $r$  in  $\mathcal{D}$  whose length is at least as great as that of every other Q-term of

rank  $r$  in  $\mathcal{D}$ . Let  $S$  be the collection of remaining  $p-1$  Q-terms of rank  $r$ . Thus  $\varepsilon yB$  is not contained in any member of  $S$ . For the purposes of this proof we write  $Y \vdash' A$  to denote that there exists an  $r, \emptyset$ -deduction of  $A$  from  $Y$  in which every Q-term of rank  $r$  is a member of  $S$ . It is sufficient to prove  $X \vdash' f$ , since we may then apply the induction hypothesis.

Suppose that  $\mathcal{D}$  is the sequence  $\langle A_1, \dots, A_m \rangle$ . For each  $i = 1, \dots, m$ , let  $A'_i$  be the formula obtained from  $A_i$  by replacing each occurrence of  $\varepsilon yB$  by  $B(\varepsilon yB)$  and, if  $B$  is of the form  $\neg C$ , by replacing each occurrence of  $\forall yC$  by  $C(\varepsilon y \neg C)$ , i.e.,  $C(\varepsilon yB)$ . Thus if  $A_i$  is the Q4-axiom  $\exists yB \rightarrow B(\varepsilon yB)$ , then  $A'_i$  is the tautology

$$(1) \quad B(\varepsilon yB) \rightarrow B(\varepsilon yB).$$

If  $B$  is of the form  $\neg C$  and  $A_i$  is the Q1-axiom  $\forall yC \rightarrow \neg \exists y \neg C$ , then  $A'_i$  is the tautology

$$(2) \quad C(\varepsilon y \neg C) \rightarrow \neg \neg C(\varepsilon y \neg C),$$

and if  $A_i$  is the Q2-axiom  $\neg \forall yC \rightarrow \exists y \neg C$ , then  $A'_i$  is the tautology

$$(3) \quad \neg C(\varepsilon y \neg C) \rightarrow \neg C(\varepsilon y \neg C).$$

Finally, if  $A_i$  is the Q3-axiom  $\neg \exists yB \rightarrow \neg B(t)$ , for some term  $t$ , then  $A'_i$  is the formula

$$(4) \quad \neg B(\varepsilon yB) \rightarrow \neg B(t).$$

If  $A_i$  is any other axiom, then  $A'_i$  is an axiom of the same form. (Using rank property R4 and the fact that every E2-term in  $\mathcal{D}$  has rank  $< r$ , verify that if  $A_i$  is an E2-axiom, then so is  $A'_i$ .) Finally if  $A_i$  is a member of  $X$ , then  $A'_i$  is  $A_i$ . For, since  $rk(\varepsilon yB) > 1$ , then  $\exists yB$  (and  $\forall yC$ ) are not  $\varepsilon$ -free. Therefore, since every member of  $X$  is  $\varepsilon$ -free, the members of  $X$  are unaffected.

Consequently, if we augment the sequence  $\langle A'_1, \dots, A'_m \rangle$  by including proofs of the tautologies (1), (2), and (3), and if we regard the formulae of the form (4) as assumptions, we obtain a deduction of  $f$  from

$$X \cup \{ \neg B(\varepsilon yB) \rightarrow \neg B(t_1), \dots, \neg B(\varepsilon yB) \rightarrow \neg B(t_n) \}$$

for some terms  $t_1, \dots, t_n$ . Furthermore by rank property R2 and our choice of  $\varepsilon yB$  with maximal length this deduction is an  $r, \emptyset$ -deduction such that every Q-term of rank  $r$  is still a member of  $S$ . Hence

$$(5) \quad X, \neg B(\varepsilon yB) \rightarrow \neg B(t_1), \dots, \neg B(\varepsilon yB) \rightarrow \neg B(t_n) \vdash' f.$$

Now using the tautologies

$$B(\varepsilon yB) \rightarrow \neg B(\varepsilon yB) \rightarrow \neg B(t_i)$$

and Theorem II.3(ii), (5) implies

$$(6) \quad X, B(\varepsilon yB) \vdash' f,$$

and therefore

$$(7) \quad X \vdash' \neg B(\varepsilon yB)$$

by the  $f$ -rule.

Now let  $\langle C_1, \dots, C_k \rangle$  be this  $r, \emptyset$ -deduction of  $\neg B(\varepsilon yB)$  from  $X$ . For each  $C_i$  and each term  $t_j$  let  $C_i^j$  be the formula obtained from  $C_i$  by replacing each occurrence of  $\varepsilon yB$  by  $t_j$ . Then for each  $j = 1, \dots, n$ , the sequence  $\langle C_1^j, \dots, C_k^j \rangle$  is an  $r, \emptyset$ -deduction of  $\neg B(t_j)$  from  $X$ , by Theorem III.6. Furthermore by our choice of  $\varepsilon yB$  with maximal length, every Q-term of rank  $r$  in this deduction is still a member of  $S$ . Hence, for each  $j = 1, \dots, n$

$$(8) \quad X \vdash' \neg B(t_j),$$

and consequently by the tautology rule

$$(9) \quad X \vdash' \neg B(\varepsilon yB) \rightarrow \neg B(t_j).$$

The desired result,  $X \vdash' f$ , now follows from (5) and (9) by Theorem II.3(ii). This completes the proof.

Notice that the above proof depends on two essential conditions: (i) that  $r > 1$ , and (ii) that  $X$  is  $\varepsilon$ -free. We impose these conditions in order to ensure that the replacements of  $\exists yB$  by  $B(\varepsilon yB)$ ,  $\forall yC$  by  $C(\varepsilon y \neg C)$ , and  $\varepsilon yB$  by  $t_j$  do not affect any members of  $X$ . In the following theorem we consider the special cases where  $r = 1$  and where  $X$  is a certain set of formulae which may contain the  $\varepsilon$ -symbol. This theorem will be used later as the central lemma in proving the First  $\varepsilon$ -Theorem.

Recall that a quasi-formula is elementary if it contains no occurrences of the symbols  $\forall, \exists$ , or  $\varepsilon$ .

**THEOREM III.8.** *Let  $Y$  be any set of elementary quasi-formulae and let  $X$  be any set of substitution instances of members of  $Y$ . (The members of  $X$  are not necessarily  $\varepsilon$ -free.) If  $X \vdash^{1,0} f$ , then there exists a set  $Z$  of substitution instances of members of  $Y$  such that  $Z \vdash_{EC} f$ .*

*Proof.* The proof is identical to that of the Rank Reduction Theorem except that we may no longer assume that the various replacement procedures used in this proof leave the members of  $X$  unaffected. However, since any member of  $X$  is a formula of the form

$$B_{t_1 \dots t_n}^{x_1 \dots x_n}$$

where  $B$  is an elementary quasi-formula, then any member of  $X$  which is affected is converted into another substitution instance of the same quasi-formula. Consequently, the proof of the Rank Reduction Theorem yields a 0-deduction  $\mathcal{D}$  of  $f$  from some set  $X'$  of substitution instances of members of

$Y$ . Now a 0-deduction has no Q-terms or E2-terms. Consequently no Q3, Q4, or E2 axioms are used in  $\mathcal{D}$ . If throughout  $\mathcal{D}$  we replace every quasi-formula of the form  $\forall yC$  by  $f$  and every quasi-formula of the form  $\exists xB$  by  $\neg f$  (starting with those of maximal length), the Q1-axioms and Q2-axioms are converted into tautologies and the other axioms are converted into new axioms of the same form. If we then replace every remaining  $\varepsilon$ -term by some individual symbol  $a$ , we obtain a deduction of  $f$  from a set  $Z$  of substitution instances of members of  $Y$ . Since every formula in this deduction is elementary, then  $Z \vdash_{EC} f$ .

### 3.4 The E2 Elimination Theorem

To complete the proof of the Second  $\varepsilon$ -Theorem we need to show that the E2-axioms are eliminable. This is the essence of the following theorem.

**THEOREM III.9 (The E2 Elimination Theorem).** *For any  $\varepsilon$ -free  $X$ , any  $r \geq 1$ , and any finite collection  $T_r$  of  $\varepsilon$ -terms of rank  $r$ , if  $X \vdash^{r, T_r} f$ , then  $X \vdash^{r, \emptyset} f$ .*

*Proof.* The proof is by induction on the number  $n$  of terms in  $T_r$ . If  $n = 0$ , there is nothing to prove. Suppose  $n > 0$ . Let  $\varepsilon xA$  be a member of  $T_r$  whose length is at least as great as that of every other term in  $T_r$ . Let  $S_r$  be the collection of remaining  $n - 1$  terms in  $T_r$ . By our choice of  $\varepsilon xA$  with maximal length it follows that  $\varepsilon xA$  does not occur in any member of  $S_r$ . To prove the theorem it is sufficient to prove  $X \vdash^{r, S_r} f$ , since we may then apply the induction hypothesis.

Let  $\mathcal{D}$  be any  $r, T_r$ -deduction of  $f$  from  $X$ . We may assume that  $\varepsilon xA$  is an E2-term in  $\mathcal{D}$ , since otherwise there would be nothing to prove. Secondly, we may assume that no E2-axiom of the form

$$\forall z(A_z^x \leftrightarrow A_z^y) \rightarrow \varepsilon xA = \varepsilon xA$$

is used as an axiom in  $\mathcal{D}$ , since any formula of this form follows by modus ponens from axiom schema P1 and axiom schema E3. Finally, we may assume that every E2-axiom, involving  $\varepsilon xA$ , which is used in  $\mathcal{D}$  is of the form

$$\forall z(A_z^x \leftrightarrow B_z^y) \rightarrow \varepsilon xA = \varepsilon yB,$$

i.e.,  $\varepsilon xA$  is on the left-hand side of the identity symbol. This assumption can easily be justified by observing that

$$\vdash^{r, \emptyset} \forall z(B_z^y \leftrightarrow A_z^x) \rightarrow \forall z(A_z^x \leftrightarrow B_z^y),$$

and

$$\vdash^{1,0} \varepsilon xA = \varepsilon yB \rightarrow \varepsilon yB = \varepsilon xA.$$

Let  $E_1, \dots, E_m$  be the E2-axioms, involving  $\varepsilon xA$ , which are used in  $\mathcal{D}$ . If these are regarded as assumptions, then  $\mathcal{D}$  is an  $r, S_r$ -deduction of  $f$  from  $X \cup \{E_1, \dots, E_m\}$ . Thus

$$X, E_1, \dots, E_m \vdash^{r, S_r} f.$$

Consequently, to prove  $X \vdash^{r, S_r} f$ , it is sufficient by virtue of Theorem II.3(ii) to prove for each  $i = 1, \dots, m$

$$(1) \quad X \vdash^{r, S_r} E_i.$$

Take any  $E_i$ . Let it be the formula

$$E_i \quad \forall z(A_z^x \leftrightarrow B_z^y) \rightarrow \varepsilon xA = \varepsilon yB.$$

Thus either (i)  $rk(\varepsilon yB) < r$  or (ii)  $rk(\varepsilon yB) = r$  and  $\varepsilon yB$  is a member of  $S_r$ . By the tautology rule and the  $f$ -rule, to prove (1) it is sufficient to prove

$$(2) \quad X, \forall z(A_z^x \leftrightarrow B_z^y) \vdash^{r, S_r} f.$$

The proof of (2) is as follows. Throughout the original deduction  $\mathcal{D}$  of  $f$  from  $X$  replace every occurrence of  $\varepsilon xA$  by  $\varepsilon yB$ . For any formula  $A_j$  which is used as an axiom in  $\mathcal{D}$ , if  $A_j$  is not one of the E2-axioms  $E_1, \dots, E_m$  and not the Q3-axiom  $\exists xA \rightarrow A(\varepsilon xA)$ , then by Theorem III.6, the formula obtained from  $A_j$  by replacing  $\varepsilon xA$  by  $\varepsilon yB$  is an axiom of the same form. Furthermore since  $X$  is  $\varepsilon$ -free, the members of  $X$  are unaffected by the replacement of  $\varepsilon xA$  by  $\varepsilon yB$ . Consequently, it follows that

$$(3) \quad X, \exists xA \rightarrow A(\varepsilon yB), E_1', \dots, E_m' \vdash^{r, S_r} f,$$

where for each  $j = 1, \dots, m$ ,  $E_j'$  is the formula obtained from  $E_j$  by replacing each occurrence of  $\varepsilon xA$  by  $\varepsilon yB$ . (Why is the deduction still an  $r, S_r$ -deduction?)

Now to prove (2) it is sufficient by virtue of (3) and Theorem II.3(ii) to prove that

$$(4) \quad \forall z(A_z^x \leftrightarrow B_z^y) \vdash^{r, \emptyset} \exists xA \rightarrow A(\varepsilon yB),$$

and for each  $j = 1, \dots, m$ ,

$$(5) \quad \forall z(A_z^x \leftrightarrow B_z^y) \vdash^{r, S_r} E_j'.$$

*Proof of (4):* Let  $Y$  be  $\{\forall z(A_z^x \leftrightarrow B_z^y)\}$ . Since  $\varepsilon xA$  and  $\varepsilon yB$  have rank  $\leq r$ , the term  $\varepsilon z \neg(A_z^x \leftrightarrow B_z^y)$  also has rank  $\leq r$  by rank property R4. Therefore these three terms may be used as Q-terms in the desired  $r, \emptyset$ -deduction of  $\exists xA \rightarrow A(\varepsilon yB)$  from  $Y$ . This deduction can be constructed as follows:

$$\begin{array}{ll} Y \vdash^{r, \emptyset} \forall z(A_z^x \leftrightarrow B_z^y) & \\ Y \vdash^{r, \emptyset} A(\varepsilon xA) \leftrightarrow B(\varepsilon xA) & \text{V-elimination} \\ Y \vdash^{r, \emptyset} A(\varepsilon yB) \leftrightarrow B(\varepsilon yB) & \text{V-elimination} \\ Y \vdash^{r, \emptyset} \exists xA \rightarrow A(\varepsilon xA) & \text{Q4-axiom} \\ Y \vdash^{r, \emptyset} \neg \exists yB \rightarrow \neg B(\varepsilon xA) & \text{Q3-axiom} \\ Y \vdash^{r, \emptyset} \exists yB \rightarrow B(\varepsilon yB) & \text{Q4-axiom} \\ Y \vdash^{r, \emptyset} \exists xA \rightarrow A(\varepsilon yB) & \text{tautology rule} \end{array}$$

*Proof of (5):* Suppose  $E_j$  is the formula

$$\forall u(A_u^x \leftrightarrow C_u^y) \rightarrow \varepsilon xA = \varepsilon vC.$$

Then  $E_j'$  is the formula

$$\forall u(A_u^x \leftrightarrow C_u^y) \rightarrow \varepsilon yB = \varepsilon vC',$$

where  $C'$  is obtained from  $C$  by replacing every occurrence of  $\varepsilon xA$  (if any) by  $\varepsilon yB$ . Recall that each of the terms  $\varepsilon yB$  and  $\varepsilon vC'$  satisfies the condition that either (i) its rank is less than  $r$  or (ii) its rank equals  $r$  and it is a member of  $S_r$ . Consequently by rank property R2 and the fact that  $\varepsilon xA$  does not occur in any member of  $S_r$ , the term  $\varepsilon vC'$  also satisfies this condition. Therefore the terms  $\varepsilon yB$  and  $\varepsilon vC'$  may be used as E2-terms in the desired  $r, S_r$ -deduction of  $E_j'$  from  $\forall z(A_z^x \leftrightarrow B_z^y)$ . This deduction can be constructed as follows. Let  $Y$  be the set

$$\{\forall z(A_z^x \leftrightarrow B_z^y), \forall u(A_u^x \leftrightarrow C_u^y)\}$$

and let  $t$  be the term  $\varepsilon w \neg(B_w^y \leftrightarrow C_w^y)$ , where  $w$  is some variable which is free for  $y$  in  $B$  and free for  $v$  in  $C'$ . Then

$$\begin{array}{ll} Y \vdash^{r, S_r} \forall z(A_z^x \leftrightarrow B_z^y) & \\ Y \vdash^{r, S_r} \forall u(A_u^x \leftrightarrow C_u^y) & \\ Y \vdash^{r, S_r} [A_z^x \leftrightarrow B_z^y]_t^z & \text{V-elimination} \\ Y \vdash^{r, S_r} [A_u^x \leftrightarrow C_u^y]_t^u & \text{V-elimination} \\ Y \vdash^{r, S_r} [B_w^y \leftrightarrow C_w^y]_t^w & \text{tautology rule} \\ Y \vdash^{r, S_r} \forall w(B_w^y \leftrightarrow C_w^y) & \text{V-introduction} \\ Y \vdash^{r, S_r} \forall w(B_w^y \leftrightarrow C_w^y) \rightarrow \varepsilon yB = \varepsilon vC' & \text{E2-axiom} \\ Y \vdash^{r, S_r} \varepsilon yB = \varepsilon vC' & \text{modus ponens} \\ \forall z(A_z^x \leftrightarrow B_z^y) \vdash^{r, S_r} E_j' & \rightarrow\text{-introduction} \end{array}$$

This completes the proof of Theorem III.9.

### 3.5 The First and Second $\varepsilon$ -Theorems

We have now established all the major results which are needed in proving the Second  $\varepsilon$ -Theorem. However, in order to prove the First  $\varepsilon$ -Theorem we still need the following simple lemma.

**LEMMA.** For any prenex formula  $A$ , there exists a substitution instance  $A'$  of the matrix of  $A$  such that  $\vdash^{1, \emptyset} A' \leftrightarrow A$ .

*Proof.* The proof depends on the following fact. If  $rk(\varepsilon yB) = 1$ , then

$$\begin{array}{ll} (1) & \vdash^{1, \emptyset} B(\varepsilon yB) \leftrightarrow \exists yB, \text{ and} \\ (2) & \vdash^{1, \emptyset} B(\varepsilon y \neg B) \leftrightarrow \forall yB. \end{array}$$

This fact is an obvious consequence of Theorem II.10. Now let  $A$  be the prenex formula

$$Q_1 x_1 \dots Q_n x_n B,$$

where  $B$  is its matrix. For each  $i = 1, \dots, n$ , we define the term  $t_i$  as follows:

$$t_1 \text{ is } \begin{cases} \varepsilon x_1 Q x_2 \dots Q x_n B & \text{if } Q_1 \text{ is } \exists \\ \varepsilon x_1 \neg Q x_2 \dots Q x_n B & \text{if } Q_1 \text{ is } \forall, \end{cases}$$

$$t_2 \text{ is } \begin{cases} \varepsilon x_2 Q x_3 \dots Q x_n [B]_{t_1}^{x_1} & \text{if } Q_2 \text{ is } \exists \\ \varepsilon x_2 \neg Q x_3 \dots Q x_n [B]_{t_1}^{x_1} & \text{if } Q_2 \text{ is } \forall, \end{cases}$$

and so on. Now let  $A'$  be the substitution instance  $B_{t_1}^{x_1} \dots B_{t_n}^{x_n}$  of  $B$ . Since  $B$  is  $\varepsilon$ -free (by the definition of a prenex formula), then  $rk(t_i) = 1$ , for each  $i$ . Consequently, using induction on  $n$  we can prove  $\vdash A' \leftrightarrow A$  from (1) and (2).

**THEOREM III.10 (The First  $\varepsilon$ -Theorem).** *If  $X$  is any set of prenex formulae,  $A$  is any prenex formula, and  $X \vdash_{PC} A$ , then  $Z \vdash_{EC} B_1 \vee \dots \vee B_n$ , where each member of  $Z$  is some substitution instance of the matrix of some member of  $X$  and each  $B_i$  is some substitution instance of the matrix of  $A$ .*

*Proof.* We shall first prove the theorem for the special case where  $A$  is the formula  $f$ .

Assume  $X \vdash_{PC} f$ . Then there exists a finite subset  $X_1$  of  $X$  such that  $X_1 \vdash_{PC} f$ . Let  $Y$  be the (finite) set consisting of the matrices of the members of  $X_1$ . By Theorem III.2 there exists a deduction of  $f$  from  $X_1$  in the  $\varepsilon^*$ -calculus, and therefore  $X_1 \vdash^{1,0} f$ . Now let  $X_2$  be the finite set of substitution instances of members of  $Y$  which correspond to each member of  $X_1$  according to the above lemma. Since  $X_2 \vdash^{1,0} C$  for each  $C$  in  $X_1$ , then by Theorem II.3(ii),  $X_2 \vdash^{1,0} f$ . Therefore by Theorem III.8 there exists a set  $Z$  of substitution instances of members of  $Y$  such that  $Z \vdash_{EC} f$ . This proves the theorem for the special case where  $A$  is the formula  $f$ .

Now suppose that  $A$  is any prenex formula of the form  $Q_1 x_1 \dots Q_k x_k C$ , where  $C$  is its matrix. Let  $B$  be the prenex formula  $Q_1' x_1 \dots Q_k' x_k \neg C$ , where as usual  $Q_i'$  is  $\exists$  if  $Q_i$  is  $\forall$ , and  $Q_i'$  is  $\forall$  if  $Q_i$  is  $\exists$ . Note that any substitution instance of  $B$  is the negation of the corresponding substitution instance of  $A$ . Since  $\vdash_{PC} A \leftrightarrow \neg B$ , then  $X \vdash_{PC} A$  implies  $X, B \vdash_{PC} f$ . Therefore by the above special case there exists a set  $Z$  such that  $Z \vdash_{EC} f$  and every member of  $Z$  is a substitution instance of the matrix of some member of  $X$  or a substitution instance of  $\neg C$ . Let  $\neg B_1, \dots, \neg B_n$  be the substitution instances of  $\neg C$  and let  $A_1, \dots, A_m$  be the remaining members of  $Z$ . Since

$$\neg B_1, \dots, \neg B_n, A_1, \dots, A_m \vdash_{EC} f,$$

repeated applications of the  $\alpha$ -rule (Theorem II.12) yield

$$\neg(B_1 \vee \dots \vee B_n), A_1, \dots, A_m \vdash_{EC} f,$$

and the desired result now follows by the  $f$ -rule.

**THEOREM III.11 (The Second  $\varepsilon$ -Theorem).** *For any  $\varepsilon$ -free  $X$  and any  $\varepsilon$ -free  $A$ , if  $X \vdash_\varepsilon A$ , then  $X \vdash_{PC} A$ .*

*Proof.* Since  $X \vdash_\varepsilon A$ , then  $X, \neg A \vdash_\varepsilon f$  by the  $f$ -rule. By the Rank Reduction Theorem and the  $E2$ -Elimination Theorem this implies that there exists a 1,  $\emptyset$ -deduction of  $f$  from  $X \cup \{\neg A\}$ , and therefore a 1,  $\emptyset$ -deduction  $\mathcal{D}$  of  $A$  from  $X$ . We want to convert the deduction  $\mathcal{D}$  into a deduction of  $A$  from  $X$  in the  $\varepsilon^*$ -calculus, since we can then apply Theorem III.3 to get a deduction of  $A$  from  $X$  in the predicate calculus. The only difficulty is that the formulae in  $\mathcal{D}$  are not necessarily proper. Consequently, throughout  $\mathcal{D}$  we replace every quasi  $\varepsilon$ -term which contains a free variable by the symbol  $a$  (starting with those quasi-terms of maximal length.) Since every  $Q$ -term in  $\mathcal{D}$  has rank 1, it is easy to see that this procedure does not damage any of the axioms and therefore provides a deduction of  $A$  from  $X$  in the  $\varepsilon^*$ -calculus.

#### 4.1 Skolem's Theorem

One of the most important consequences of the Second  $\varepsilon$ -Theorem is the counterpart of Theorem II.28 for the predicate calculus. This result is commonly known as Skolem's Theorem.

**THEOREM III.12 (Skolem's Theorem).** *Let  $A$  be any prenex formula of  $\mathcal{L}(\mathcal{V})$ ,  $C$  any  $\varepsilon$ -free formula of  $\mathcal{L}(\mathcal{V})$ ,  $X$  any set of  $\varepsilon$ -free formulae of  $\mathcal{L}(\mathcal{V})$ , and  $\mathcal{V}^*$  the vocabulary obtained from  $\mathcal{V}$  by adjoining the Skolem functions used in forming the Skolem resolution  $A_S$  of  $A$ . Then*

$$X, A \vdash_{PC(\mathcal{V})} C \text{ iff } X, A_S \vdash_{PC(\mathcal{V}^*)} C.$$

*Proof.* Use Theorem II.28 and the Second  $\varepsilon$ -Theorem. (Note: The half of the theorem which states that  $X, A \vdash_{PC(\mathcal{V})} C$  implies  $X, A_S \vdash_{PC(\mathcal{V}^*)} C$  is much 'weaker' than the other half and can be proved directly using only the axioms and rules of inference of the predicate calculus.)

Notice that Theorems II.25, II.26, and II.27 also hold for the predicate calculus by virtue of the Second  $\varepsilon$ -Theorem. The analogue of Theorem II.26 now provides us with a full justification (in terms of the predicate calculus) of the familiar 'rule of inference' used by mathematicians when they pass from the statement

$$\text{'for all } x, \text{ there exists a } y \text{ such that } B(x,y)\text{'}$$

to the statement

$$\text{'for all } x, B(x,g(x))\text{'}$$

or, to put it more colloquially, to the statement

'for each  $x$ , let  $g(x)$  be some  $y$  such that  $B(x,y)$ ',

provided that the function  $g$  does not appear in the statement  $B(x,y)$ , in the statement being proved, or in any of the assumptions on which the proof is based (cf. Chapter II, §9). We can state the formal justification of this rule as follows.

**THEOREM III.13.** *If  $X \vdash_{PC} \forall x_1 \dots \forall x_n \exists y B$  and  $Y, \forall x_1 \dots \forall x_n [B]_{g x_1 \dots x_n}^y \vdash_{PC} C$ , then  $X \cup Y \vdash_{PC} C$ , provided that  $g$  does not occur in  $C$ ,  $B$ , or any member of  $Y$ .*

*Proof.* By Theorem III.2, Theorem II.26, and the Second  $\varepsilon$ -Theorem we have  $Y, \forall x_1 \dots \forall x_n \exists y B \vdash_{PC} C$ . Hence  $X \cup Y \vdash_{PC} C$  by Theorem II.32.

Unfortunately, the restriction that  $g$  does not occur in any member of  $Y$  limits the applicability of this rule in ordinary mathematical arguments. In most branches of mathematics the axioms of set theory are used as basic assumptions. These axioms usually include certain formulae known as *axioms of replacement* (cf. p. 106). In order to apply the above rule concerning the eliminability of the new function symbol  $g$ , one must make sure that the set of assumptions  $Y$  does not include an axiom of replacement in which the symbol  $g$  occurs. For example, if one uses the fact that the values of  $g(x)$  form a set when  $x$  ranges over some set, then one is using an axiom of replacement which contains  $g$ . In this case the above rule can only be justified if the axiom of choice is included as one of the basic assumptions.

However, one does not need to appeal to the axiom of choice if one has proved the stronger statement

'for all  $x$ , there exists a *unique*  $y$  such that  $B(x,y)$ '.

In this case the function symbol  $g$  can be eliminated even if it is used in an axiom of replacement. The proof of this fact does not depend on the Second  $\varepsilon$ -Theorem, but rather on the eliminability of the  $\iota$ -symbol (cf. p. 100).

We shall return to these problems in Chapter IV when we consider formalizations of set theory based on the  $\varepsilon$ -calculus.

#### 4.2 Herbrand's Theorem

Herbrand's Theorem, as it was originally formulated (Herbrand [1930]), involves the complicated notions of 'properties  $B$  and  $C$  of order  $p$ '. In order to simplify both the statement and proof of this theorem various people have proved results which bear the title 'Herbrand's Theorem', but which are considerably weaker than Herbrand's original assertion. This situation, together with the relative inaccessibility of Herbrand's paper, has doubtless given rise to a certain amount of confusion about the exact nature of this

theorem. In this section we shall give a relatively simple formulation and proof of Herbrand's Theorem which incorporates most of the essential features of the original assertion.

The essence of Herbrand's Theorem can be described as follows. Let  $A$  be any identity-free, prenex formula. For each positive integer  $p$  we assign to  $A$  a particular elementary formula, called the ' $p$ -reduction' of  $A$ . We say that  $A$  is a ' $p$ -tautology' if the  $p$ -reduction of  $A$  is a tautology. The theorem then states that  $A$  is a theorem of the predicate calculus if and only if  $A$  is a  $p$ -tautology for some  $p > 0$ . This theorem is a very powerful one since it establishes a useful necessary and sufficient condition for provability in the predicate calculus. One of its main applications has been in solving various cases of the decision problem for the predicate calculus.

Before defining the  $p$ -reduction of  $A$ , we need the following definitions. For any term  $t$ , the *degree* of  $t$  is defined as follows by induction on its length:

1. If  $t$  has no subterms, the degree of  $t$  is 1.
2. If  $t$  has subterms, the degree of  $t$  is one greater than the maximal degree of all its subterms.

For example, the degree of any individual symbol or 0-place function symbol is 1, and the degree of the term  $g^2 a h^1 a$  is 3.

Let  $\mathcal{G}$  be any finite collection of individual symbols and function symbols. Then a  $\mathcal{G}$ -term is any term whose symbols are all members of  $\mathcal{G}$ .

Now let  $A$  be any prenex formula, let  $A_H$  be its Herbrand resolution, let  $B$  be the matrix of  $A_H$ , and let  $\mathcal{G}$  be the collection of individual symbols and function symbols occurring in  $B$ . (If  $B$  contains no individual or 0-place function symbols, we take  $a_1$  as an additional member of  $\mathcal{G}$ .) For any positive integer  $p$ , a  $p$ -substitution instance of  $B$ , is any substitution instance

$$B_{t_1 \dots t_n}^{x_1 \dots x_n}$$

of  $B$  where each  $t_i$  is a  $\mathcal{G}$ -term with degree  $\leq p$ . Thus for any  $p$ , there are finitely many  $p$ -substitution instances of  $B$ . The  $p$ -reduction of  $A$  is now defined as the disjunction,  $B_1 \vee \dots \vee B_m$ , of all the  $p$ -substitution instances of  $B$ . We say that  $A$  is a  $p$ -tautology if its  $p$ -reduction is a tautology. (To be exact, the  $p$ -reduction of  $A$  depends not only on  $A$  and  $p$ , but also on the order of its disjunctive parts,  $B_1, \dots, B_m$ , and on the particular choice of Herbrand functions used in forming  $A_H$ . However, for any two  $p$ -reductions of  $A$ , the one is a tautology if and only if the other is. Consequently, the notion of a  $p$ -tautology is well-defined.)

Clearly, if  $A$  is a  $p$ -tautology, then for any  $q \geq p$ ,  $A$  is also a  $q$ -tautology. For, if  $q \geq p$ , then every  $p$ -substitution instance of the matrix of  $A_H$  is also a  $q$ -substitution instance, and therefore the  $p$ -reduction of  $A$  is a 'sub-disjunction' of the  $q$ -reduction of  $A$ .

To illustrate the above definitions, consider the following example. Let  $A$  be the formula

$$\exists y \forall z (Py \rightarrow Pz),$$

where  $P$  is some 1-place predicate symbol. Then the Herbrand resolution,  $A_H$ , of  $A$  is

$$\exists y (Py \rightarrow Pgy),$$

where  $g$  is a 1-place function symbol. Thus we let  $\mathcal{G}$  be the set  $\{a_1, g\}$ . The only  $\mathcal{G}$ -term of degree 1 is  $a_1$ , the only  $\mathcal{G}$ -term of degree 2 is  $ga_1$ , etc. Consequently, the 1-reduction of  $A$  is

$$(1) \quad Pa_1 \rightarrow Pga_1$$

and the 2-reduction of  $A$  is

$$(2) \quad (Pa_1 \rightarrow Pga_1) \vee (Pga_1 \rightarrow Pgga_1).$$

Although (1) is not a tautology, (2) is a tautology. Hence  $A$  is a  $p$ -tautology for any  $p \geq 2$ .

We can now give the statement and proof of Herbrand's Theorem.

**THEOREM III.14 (Herbrand's Theorem).** *Let  $A$  be any identity-free prenex formula. Then:*

- (i) *If  $\vdash_{PC} A$ , then there exists a positive integer  $p$  such that  $A$  is a  $p$ -tautology.*
- (ii) *If  $A$  is a  $p$ -tautology for some positive integer  $p$ , then  $\vdash_{PC} A$ .*

*Proof.* (i) Assume  $\vdash_{PC} A$ . Let  $A_H$  be the Herbrand resolution of  $A$ ,  $B$  the matrix of  $A_H$ , and  $\mathcal{G}$  the collection of individual symbols and function symbols occurring in  $B$ . ( $\mathcal{G}$  also contains  $a_1$  if necessary.) Since  $\vdash_{PC} A$ , then  $\vdash_{\varepsilon} A$  by Theorem III.2, and therefore  $\vdash_{\varepsilon} A_H$  by Theorem II.27. By the First and Second  $\varepsilon$ -Theorems this implies that there exist substitution instances  $B_1, \dots, B_n$  of  $B$  such that  $\vdash_{EC} B_1 \vee \dots \vee B_n$ . Since  $A$  is identity-free, this implies by Theorem II.30 that  $B_1 \vee \dots \vee B_n$  is a tautology. If any of the substituted terms in  $B_1 \vee \dots \vee B_n$  are not  $\mathcal{G}$ -terms, replace every occurrence of such terms by some  $\mathcal{G}$ -term of degree 1. In this way we obtain substitution instances  $B_1', \dots, B_n'$  of  $B$  such that  $B_1' \vee \dots \vee B_n'$  is still a tautology. Let  $p$  be the maximal degree of all the substituted terms in  $B_1' \vee \dots \vee B_n'$ . (If there are no such terms let  $p = 1$ .) Since each  $B_i'$  is a  $p$ -substitution instance of  $B$ , then the tautology  $B_1' \vee \dots \vee B_n'$  is a sub-disjunction of the  $p$ -reduction of  $A$ . Hence  $A$  is a  $p$ -tautology.

(ii) Conversely, assume that  $A$  is a  $p$ -tautology for some  $p > 0$ . The knowledge of  $p$  enables us to form the  $p$ -reduction of  $A$ , which we shall denote by  $B_1 \vee \dots \vee B_n$ . Since  $B_1 \vee \dots \vee B_n$  is a tautology, then by Theorem II.29,  $\vdash_{EC} B_1 \vee \dots \vee B_n$  and therefore  $\vdash_{PC} B_1 \vee \dots \vee B_n$ . However, since each  $B_i$  is a substitution instance of the matrix of the  $\exists$ -prenex formula

$A_H$ , then  $B_i \vdash_{PC} A_H$  for each  $i = 1, \dots, n$ . Consequently, by repeated applications of the  $\beta$ -rule, we have  $B_1 \vee \dots \vee B_n \vdash_{PC} A_H$ , and therefore  $\vdash_{PC} A_H$ . By Theorem III.2, this implies that there exists a proof of  $A_H$  in the  $\varepsilon$ -calculus in which every formula is identity-free. Consequently, by Theorem II.27 and the Second  $\varepsilon$ -Theorem we have  $\vdash_{PC} A$ . This completes the proof of Herbrand's Theorem.

It should be mentioned that in Herbrand's original statement of the theorem  $A$  is an arbitrary formula rather than a prenex formula. Since any  $\varepsilon$ -free formula is equivalent to some prenex formula, our version of the theorem is no weaker than Herbrand's, except that Herbrand attempts to reveal the relationship between an arbitrary formula and the various prenex equivalents of that formula. However, Dreben, Andrews, and Aanderaa [1963] have discovered errors in Herbrand's proof. A corrected proof has been produced by Denton and Dreben [1969].

### EXERCISES

1. Using Herbrand's Theorem, prove that the formula  $\forall x \exists y Pxy \rightarrow \exists y \forall x Pxy$  is not a theorem of the predicate calculus. (See the exercise on page 54.)
2. Let  $A$  be any identity-free, prenex formula of the form

$$\forall x_1 \dots \forall x_n \exists y_1 \dots \exists y_n B,$$

where  $B$  is the matrix of  $A$  and no  $n$ -place function symbols occur in  $B$  for  $n > 0$ . Describe a decision procedure for determining whether or not  $A$  is a theorem of the predicate calculus without identity.

### 4.3 The eliminability of the identity symbol

Herbrand's Theorem can be used to prove that the identity symbol can be eliminated from proofs in the predicate calculus of formulae which do not themselves contain the identity symbol.

**THEOREM III.15.** *If  $A$  is an identity-free formula and  $\vdash_{PC} A$ , then  $\vdash_{PC} A$ .*

*Proof.* We may assume that  $A$  is a prenex formula since for any  $\varepsilon$ -free, identity-free formula  $A$  there exists an identity-free prenex formula  $A'$  such that  $\vdash_{PC} A \leftrightarrow A'$  (cf. the proof of Theorem II.24). Since  $\vdash_{PC} A$  and  $A$  is identity-free, then by the first half of Herbrand's Theorem there exists a positive integer  $p$  such that  $A$  is a  $p$ -tautology. By the second half of Herbrand's Theorem this implies  $\vdash_{PC} A$ .

The eliminability of the  $=$ -symbol does not hold for the  $\varepsilon$ -calculus since the E2-axioms are needed for proving theorems which do not themselves contain

the identity-symbol. For example, if  $P$ ,  $Q$ , and  $R$  are 1-place predicate symbols, the formula

$$(\forall z(Pz \leftrightarrow Qz) \wedge R\epsilon xPx) \rightarrow R\epsilon yQy$$

is a theorem of the  $\varepsilon$ -calculus by virtue of axioms E1 and E2, but there exists no identity-free proof of this formula.

### EXERCISE

Modify the axioms of the  $\varepsilon$ -calculus to obtain an equivalent formal system in which the identity symbol is eliminable. (*Hint*: see Chapter I, §6.)

## CHAPTER IV

# FORMAL THEORIES

### 1 Introduction

The objective of this chapter is to reveal the role which the  $\varepsilon$ -symbol can play in the study of formal theories. First of all, we shall explain how the formalists used the  $\varepsilon$ -symbol and the  $\varepsilon$ -Theorems in constructing finitary consistency proofs of various formal theories. Secondly, we shall see how the formulation of certain theories can be simplified if the  $\varepsilon$ -calculus rather than the predicate calculus is used as the underlying formal system. Lastly, we shall investigate the relationship between the  $\varepsilon$ -symbol and the axiom of choice in formal set theory.

A *formal theory*  $\mathcal{T}$  consists of a vocabulary  $\mathcal{V}$ , a formal system  $\mathcal{F}$  for  $\mathcal{V}$ , and a set  $\mathcal{A}$  of formulae of  $\mathcal{L}(\mathcal{V})$ . Thus we may regard a formal theory  $\mathcal{T}$  as an ordered triple  $\langle \mathcal{V}, \mathcal{F}, \mathcal{A} \rangle$ . The members of  $\mathcal{A}$  are called the (*non-logical*) *axioms of*  $\mathcal{T}$ ,  $\mathcal{V}$  is called the *vocabulary of*  $\mathcal{T}$ , and  $\mathcal{F}$  is called the *underlying formal system of*  $\mathcal{T}$ . Alternatively, we often say that  $\mathcal{T}$  is *based on*  $\mathcal{F}$ . By an abuse of language we often say a 'formula of  $\mathcal{T}$ ' instead of a 'formula of  $\mathcal{L}(\mathcal{V})$ '. Furthermore, if  $\mathcal{T}$  is based on the predicate calculus we say a 'formula of  $\mathcal{T}$ ' instead of an ' $\varepsilon$ -free formula of  $\mathcal{L}(\mathcal{V})$ '. Throughout this chapter most of the formal theories we deal with are based either on the predicate calculus or on the  $\varepsilon$ -calculus.

Let  $\mathcal{T}$  be some formal theory  $\langle \mathcal{V}, \mathcal{F}, \mathcal{A} \rangle$ . A formula  $A$  of  $\mathcal{L}(\mathcal{V})$  is a *theorem of*  $\mathcal{T}$  if and only if  $\mathcal{A} \vdash_{\mathcal{F}} A$ . The theory  $\mathcal{T}$  is *inconsistent* if  $\perp$  is a theorem of  $\mathcal{T}$ ; otherwise it is *consistent*. A theory  $\mathcal{T}'$  is an *extension* of a theory  $\mathcal{T}$  if every formula of  $\mathcal{T}$  is a formula of  $\mathcal{T}'$  and every theorem of  $\mathcal{T}$  is a theorem of  $\mathcal{T}'$ . Two theories are said to be *equivalent* if each is an extension of the other. An extension  $\mathcal{T}'$  of  $\mathcal{T}$  is an *inessential extension* of  $\mathcal{T}$  if every formula of  $\mathcal{T}$  which is a theorem of  $\mathcal{T}'$  is also a theorem of  $\mathcal{T}$ . Obviously if  $\mathcal{T}'$  is an inessential extension of  $\mathcal{T}$ , then  $\mathcal{T}$  is consistent if and only if  $\mathcal{T}'$  is consistent.

### 2 Finitary consistency proofs

Suppose  $\mathcal{T}$  is a theory based on the predicate calculus. If there exists a model  $\mathcal{M}$  which satisfies the set  $\mathcal{A}$  of axioms of  $\mathcal{T}$ , it follows that  $\mathcal{T}$  must be consistent. For, the soundness of the predicate calculus implies that every theorem of  $\mathcal{T}$  must be true in the model  $\mathcal{M}$ . Since the formula  $\perp$  is false in all

models, then  $f$  is not a theorem of  $\mathcal{T}$ , and  $\mathcal{T}$  is consistent. The objection to this type of consistency proof is that a very strong metatheory is required. For example, the cardinality of the model may be infinite, and in order to show that  $f$  is not a theorem of  $\mathcal{T}$  one must use non-constructive arguments to show that the axioms of the predicate calculus are true in this model. One of the main contributions of the formalists was in showing that for certain special theories this type of consistency proof can be carried out in a completely finitary way by appealing to the First  $\varepsilon$ -Theorem rather than to the soundness of the predicate calculus.

To illustrate the formalists' method we shall prove the consistency of the following simple theory S.

Let  $\mathcal{V}$  be the vocabulary consisting of the single 2-place predicate symbol  $<$ , where as usual we write  $s < t$  instead of  $<st$ , and let  $\mathcal{A}$  be the set consisting of the following five formulae:

- S1  $\forall x \neg(x < x),$   
 S2  $\forall x \forall y \forall z ((x < y \wedge y < z) \rightarrow x < z),$   
 S3  $\forall x \forall y (x < y \vee y < x \vee y = x),$   
 S4  $\forall x \exists y (x < y),$   
 S5  $\exists x \forall y (x = y \vee x < y).$

We define S to be the formal theory  $\langle \mathcal{V}, \text{PC}(\mathcal{V}), \mathcal{A} \rangle$ . Obviously, the model consisting of the set of positive integers with the usual ordering relation satisfies the set  $\mathcal{A}$ , and therefore by a non-finitary argument it follows that S is consistent. However, despite the fact that every model which satisfies  $\mathcal{A}$  has infinite cardinality we can still prove the consistency of S in a completely finitary way.

First of all, we modify S so that its axioms are all  $\forall$ -prenex formulae. Let  $\mathcal{V}'$  be the vocabulary obtained from  $\mathcal{V}$  by adjoining the 0-place function symbol  $c$  and the 1-place function symbol  $g$ , and let  $\mathcal{A}'$  be the set of formulae obtained from  $\mathcal{A}$  by replacing formulae S4 and S5 by

- S4'  $\forall x (x < gx),$   
 S5'  $\forall y (c = y \vee c < y).$

We define S' to be the formal theory  $\langle \mathcal{V}', \text{PC}(\mathcal{V}'), \mathcal{A}' \rangle$ . Since S4' and S5' are Skolem resolutions of S4 and S5, then by Skolem's Theorem, S' is an inessential extension of S, and therefore S' is consistent if and only if S is consistent.

Let  $Y$  be the set of matrices of the members of  $\mathcal{A}'$ . Thus  $Y$  consists of the following five quasi-formulae:

- S1''  $\neg(x < x),$   
 S2''  $(x < y \wedge y < z) \rightarrow x < z,$   
 S3''  $x < y \vee y < x \vee y = x,$

- S4''  $x < gx,$   
 S5''  $c = y \vee c < y.$

In order to prove the consistency of S', and therefore of S, we shall describe an effective procedure for assigning truth values to all elementary formulae of  $\mathcal{L}(\mathcal{V}')$  in such a way that every substitution instance of a member of  $Y$  has the value 1 ('true'). It will then follow by the First  $\varepsilon$ -Theorem that  $f$  is not a theorem of S'.

First of all, to any formula of the form  $s = t$  we assign the value 1 if  $s$  and  $t$  have the same length, and the value 0 if they do not. Similarly, to any formula of the form  $s < t$  we assign the value 1 if the length of  $s$  is less than the length of  $t$ , and the value 0 if it is not. Notice that one can effectively compare the lengths of any two expressions by successively crossing off the initial symbol from each of them until (at least) one of them is reduced to the empty expression. As usual, we assign the truth value 0 to the formula  $f$ . Since every atom of  $\mathcal{L}(\mathcal{V}')$  is either the formula  $f$  or a formula of the form  $s = t$  or  $s < t$ , then this assignment of values to the atoms can be extended to all elementary formulae by means of the truth functional interpretations of the propositional connectives.

We shall now show that, for any elementary formula  $B$ , if  $B$  is a theorem of S', then  $B$  has the truth value 1. By the First  $\varepsilon$ -Theorem, if  $B$  is a theorem of S', then there exists a set  $Z$  of substitution instances of the members of  $Y$  such that  $Z \vdash_{\text{EC}} B$ . However, it is easy to see that every member of  $Z$  must have the value 1, and furthermore every E1-axiom and every E3-axiom must have the value 1. Since the propositional axioms are tautologies and since the rule of modus ponens preserves the truth value 1, then  $B$  has the value 1. However  $f$  has the value 0. Consequently  $f$  is not a theorem of S', and therefore S' and S are consistent.

In general, we may describe the formalists' method of proving consistency as follows. Suppose  $\mathcal{T}$  is some theory based on the predicate calculus. By replacing each axiom of  $\mathcal{T}$  by some prenex equivalent of that axiom and then taking the Skolem resolutions of these prenex formulae and adjoining the new Skolem functions to the vocabulary, one obtains a theory  $\mathcal{T}'$  which is an inessential extension of  $\mathcal{T}$ . One then tries to find an effective assignment of truth values to the atomic formulae of  $\mathcal{T}'$  such that every E1-axiom, every E3-axiom, and every substitution instance of the matrices of the axioms of  $\mathcal{T}'$  has the value 1 under this assignment. If this can be done, it then follows that both  $\mathcal{T}'$  and  $\mathcal{T}$  are consistent. A detailed account of this method of proof is given in Hilbert and Bernays [1939], where the principal results are embodied in the Consistency Theorem (*Widerspruchsfreiheits-Theorem*, pages 36–37).

This method of proof actually establishes more than just the consistency of  $\mathcal{T}$ , since one shows that every elementary theorem of  $\mathcal{T}$  has an elementary



proof, i.e., one without quantifiers. A major objective of Hilbert's formalist programme was to justify the use of infinity in mathematics by proving that the non-finitary statements and methods can be eliminated from proofs of finitary (i.e. elementary) statements. Kreisel [1964], page 157, explains this aspect of Hilbert's programme as follows:

'... The *Consistency Problem* was associated with the problem of understanding the concept of infinity. He (Hilbert) sought such an understanding in understanding the use of transfinite machinery from a finitist point of view. And this he saw in the elimination of transfinite ( $\epsilon$ -) symbols from proofs of formulae not containing such symbols. He was convinced from the start that such an elimination was possible, and expressed it by saying that the problems of foundations were to be *removed* or that doubts were to be eliminated instead of saying that they were to be investigated.'

The classical logic of mathematics, as formalized by the predicate calculus, transcends the limits of finitary reasoning since it admits statements which refer to an infinite totality of objects. For example, if  $p$  is a prime number, the statement 'there exists a prime number which exceeds  $p$ ' is non-finitary, since it asserts the existence of a number having a certain property in the infinite totality of numbers which exceed  $p$ . On the other hand, the stronger statement 'there exists a prime number between  $p + 1$  and  $p! + 1$ ' is finitary since it can be expressed as a finite disjunction (cf. Hilbert [1926]). Furthermore, certain arguments which are used in classical logic and which can be formalized in the predicate calculus are unacceptable from the finitary point of view. For example, in classical logic one can prove that there exists a number which has a certain property by deducing a contradiction from the assumption that every number does not have this property.

Perhaps, the main significance of Hilbert's First  $\epsilon$ -Theorem is the following. Although classical logic, as formalized by the predicate calculus, contains certain non-finitary elements, any proof of a finitary statement can be converted into a finitary proof of that statement. It is in this sense that Hilbert justifies the use of the concept of infinity.

Unfortunately, this use of the First  $\epsilon$ -Theorem has a very limited range of applications. If the formal theory under consideration is at all complicated, the First  $\epsilon$ -Theorem may be inadequate for proving the eliminability of non-finitary elements from proofs of finitary statements. We shall consider such a theory in the following sections.

### 3 Formal Arithmetic

As we have already mentioned, one of the primary goals of Hilbert's formalist programme was to prove by finitary means the consistency of formal arithmetic, i.e., the theory which deals with the additive and multi-

plicative structure of the natural numbers. In the following sections we shall describe the way in which Hilbert hoped to use the  $\epsilon$ -symbol to prove this consistency result.

Arithmetic can be formalized as follows. Let  $\mathcal{V}$  be the vocabulary consisting of the 0-place function symbol  $\bar{0}$  (which designates the natural number 0), the 1-place function symbols  $'$  and  $g$  (which designate, respectively, the successor function and the predecessor function), and the 2-place function symbols  $+$  and  $\cdot$  (which designate addition and multiplication). As usual, we write  $(t)'$ ,  $(s + t)$ , and  $(s \cdot t)$  instead of  $'t$ ,  $+st$ , and  $\cdot st$ , and we shall omit parentheses whenever their omission gives rise to no ambiguity. Let  $\mathcal{A}$  be the set consisting of the following formulae:

N1	$\forall x \neg (\bar{0} = x')$
N2	$\forall x \forall y (x' = y' \rightarrow x = y)$
N3	$\forall x (x + \bar{0} = x)$
N4	$\forall x \forall y (x + (y)' = (x + y)')$
N5	$\forall x (x \cdot \bar{0} = \bar{0})$
N6	$\forall x \forall y (x \cdot (y)' = (x \cdot y) + x)$
N7	$\forall x (\bar{0} = x \vee (gx)' = x)$

and all universal closures of quasi-formulae of the form

$$I \quad A(\bar{0}) \rightarrow \forall x (A(x) \rightarrow A(x')) \rightarrow \forall x A(x).$$

The form I is called the *axiom schema of induction* and any universal closure of a quasi-formula of this form is called an *induction axiom*. *Formal arithmetic*, which we denote by  $N$ , is now defined as the formal theory  $\langle \mathcal{V}, PC(\mathcal{V}), \mathcal{A} \rangle$ . Actually, the function symbol  $g$  and axiom N7 are usually omitted in formalizations of arithmetic. We include them here only for technical reasons. Notice that formula N7 is the Skolem resolution of  $\forall x \exists y (\bar{0} = x \vee y' = x)$ . This latter formula can be deduced without using axiom N7.

Let  $\mathfrak{N}$  be the model  $\langle M, \Pi, \Phi \rangle$  for  $\mathcal{V}$ , where  $M$  is the set of natural numbers,  $\Pi$  assigns the above-mentioned interpretations to the symbols  $\bar{0}$ ,  $'$ ,  $g$ ,  $+$ , and  $\cdot$  and assigns the number 0 to every individual symbol, and  $\Phi$  is the least number operator, i.e., for any subset  $M_1$  of  $M$ , if  $M_1$  is non-empty,  $\Phi(M_1)$  is the least member of  $M_1$ , and if  $M_1$  is empty,  $\Phi(M_1)$  is 0. (For the interpretation of the predecessor function  $g$  we arbitrarily define the predecessor of 0 to be 0.) We shall refer to this model as the *standard model* of arithmetic. Since every axiom of  $N$  is true in  $\mathfrak{N}$ , then by the soundness of the predicate calculus every theorem of  $N$  is true in  $\mathfrak{N}$ .

Since it is possible to prove in  $N$  all the usual results about the addition and multiplication of natural numbers, the formalists felt that this theory was an adequate formalization of arithmetic and that a finitary proof of its consistency would justify the use of non-finitary statements in informal

arithmetic. However, Gödel [1931] proved two profound results which undermined the whole formalist programme. His first result, which implies that N is not an adequate formalization of arithmetic, was that if N is consistent then there exists an  $\varepsilon$ -free formula  $A$ , containing no individual symbols, such that  $A$  is true in the standard model  $\mathfrak{N}$  but is not a theorem of N. Furthermore, this deficiency cannot be rectified by adding new axioms to N. His second result, a corollary of the first, showed that any proof of the consistency of N must involve techniques or concepts which cannot be formalized within N. In particular, this result ruled out the possibility of proving the consistency of N by finitary reasoning.

Despite Gödel's results it is worth while examining Hilbert's unsuccessful attempt to find a finitary consistency proof of N since his methods can be used both to demonstrate the consistency of weakened versions of N and also to prove, using a non-finitary but limited metatheory, that N itself is consistent. Furthermore, his whole approach provides us with a clear analysis of the nature of the abstract and the concrete in mathematics.

### EXERCISES

1. Prove that the following formulae are theorems of N:
  - (i)  $\forall x \forall y (x + y = y + x)$ ,
  - (ii)  $\forall x \forall y (x \cdot y = y \cdot x)$ ,
  - (iii)  $\forall x \forall y \forall z (x \cdot (y + z) = (x \cdot y) + (x \cdot z))$ .
2. For any quasi-terms  $s$  and  $t$ , let  $s < t$  be an abbreviation of the quasi-formula  $\exists w (s + w' = t)$ , where  $w$  is some variable which does not occur free in  $s$  or  $t$ . Prove that with this definition of  $<$  the axioms S1–S5 of the theory S (page 86) are theorems of N.

(For solutions of these exercises see Mendelson [1964], pages 104–112.)

### 3.1 Numerals and numerically true formulae

The consistency of N would be established if we could prove that every theorem of N has a certain property which the formula  $f$  does not have. If we were to allow ourselves the luxury of using non-finitary reasoning, a suitable property would be 'truth in the standard model'. Of course, from our finitary point of view this property is far too abstract. However, we shall now show that if we restrict our attention to elementary formulae and replace the abstract notion of 'natural number' by the concrete notion of 'numeral', we can define in concrete terms an effective notion of 'numerical truth'. Having defined this notion, we can then try to prove that every elementary theorem of N is numerically true, thus establishing the consistency of N.

An expression of the language  $\mathcal{L}(\mathcal{V})$  is called a *numeral* if the only symbol occurring in that expression is the function symbol  $'$ . Thus, in particular,

the empty expression  $e$  is a numeral. We shall refer to this numeral as *zero*. For any numeral  $n$ , the *successor* of  $n$  is the numeral  $n'$  and the *predecessor* of  $n$  is the numeral obtained from  $n$  by removing the last occurrence of  $'$  in  $n$ . For any two numerals  $m$  and  $n$ , the *sum* of  $m$  and  $n$  is the numeral  $mn$ , and the *product* of  $m$  and  $n$  is the numeral obtained from  $n$  by replacing each occurrence of the symbol  $'$  in  $n$  by the numeral  $m$ . For example, the successor of the numeral zero is  $'$ , the predecessor of  $'''$  is  $''$ , the sum of  $''$  and  $'''$  is  $''''$ , and the product of  $'''$  and  $''$  is  $''''''$ .

To each  $\varepsilon$ -free term  $t$  of the language  $\mathcal{L}(\mathcal{V})$  we assign a unique numeral  $n$ , called the *numerical value* of  $t$ , as follows. To each individual symbol  $a$  and to the symbol  $\bar{0}$  we assign the numeral zero. If we have already assigned the numerals  $m$  and  $n$  to the terms  $s$  and  $t$  respectively, then we assign the successor of  $m$  to the term  $s'$ , the predecessor of  $m$  to the term  $gs$ , the sum of  $m$  and  $n$  to the term  $s + t$ , and the product of  $m$  and  $n$  to the term  $s \cdot t$ . Now to any  $\varepsilon$ -free formula of the form  $s = t$  we assign the truth value 1 if the numerals assigned to  $s$  and  $t$  are of the same length, i.e., if  $s$  and  $t$  have the same numerical values. Otherwise, we give it the truth value 0. For example, for any  $\varepsilon$ -free terms  $s$  and  $t$ , to the formula

$$s \cdot (t)' = (s \cdot t) + s$$

is assigned the truth value 1, and to the formula

$$\bar{0}'' \cdot \bar{0}''' = \bar{0}'' + \bar{0}'''$$

is assigned the truth value 0.

Since the vocabulary  $\mathcal{V}$  contains no predicate symbols, every elementary formula of  $\mathcal{L}(\mathcal{V})$  is built up by means of the propositional connectives from formulae of the form  $s = t$  and possibly the formula  $f$ . Consequently by using the above assignment of truth values to formulae of the form  $s = t$  and by giving the symbols  $f$ ,  $\neg$ ,  $\wedge$ ,  $\vee$ , and  $\rightarrow$  their usual truth functional interpretations, we can assign a unique truth value to any elementary formula  $B$ . If this truth value is 1, we say that  $B$  is *numerically true*. For example, an elementary formula of the form  $s' = t' \rightarrow s = t$  is numerically true, since regardless of what numerals are assigned to  $s$  and  $t$  the case where  $s' = t'$  has the truth value 1 and  $s = t$  has the truth value 0 can never arise. It is important to observe that this definition of numerical truth is an effective one. In other words, for any given elementary formula  $B$ , the above definitions describe an effective procedure for determining whether or not  $B$  is numerically true.

### EXERCISES

1. Prove that every  $\varepsilon$ -free substitution instance of the matrices of axioms N1–N7 is numerically true.

2. Let  $N_1$  be the theory whose axioms are the formulae N1–N7. Give a finitary proof of the consistency of this theory.

### 3.2 Formal arithmetic based on the $\varepsilon$ -calculus

We now turn to the problem of proving that every elementary theorem of  $N$  is numerically true. The exercises at the end of the last section reveal that the First  $\varepsilon$ -Theorem provides a simple solution to this problem if we exclude the induction axioms from  $N$ . However, when the induction axioms are included, the problem is much more difficult, since these axioms are not  $\forall$ -prenex formulae. Although we know by Skolem's Theorem that  $N$  has an inessential extension  $N'$  in which all the axioms are  $\forall$ -prenex formulae, in order to form  $N'$  we must replace each induction axiom by the Skolem resolution of some prenex equivalent of that axiom and adjoin the new Skolem functions to the vocabulary  $\mathcal{V}$ . Obviously, the theory  $N'$  which is obtained in this way is very complicated, and it seems unlikely that one could devise an effective notion of numerical truth for the elementary formulae of this theory.

In order to overcome this difficulty Hilbert and Bernays define a new formalization of arithmetic, which we shall denote by  $N_\varepsilon$ , which is based on the  $\varepsilon$ -calculus and in which the  $\varepsilon$ -symbol is used in effect as a least number operator. The formulation of the theory  $N_\varepsilon$  is based on the well-known fact that the principle of mathematical induction is deducible from the principle of the least number.

We can define the theory  $N_\varepsilon$  as follows. Let  $\mathcal{V}$  be the vocabulary used in defining the theory  $N$  and let  $\mathcal{A}_\varepsilon$  be the set of formulae of  $\mathcal{L}(\mathcal{V})$  which are instances of the following axiom schemata:

N1'	$\neg(\bar{0} = t')$
N2'	$s' = t' \rightarrow s = t$
N3'	$s + \bar{0} = s$
N4'	$s + (t)' = (s + t)'$
N5'	$s \cdot \bar{0} = \bar{0}$
N6'	$s \cdot (t)' = (s \cdot t) + s$
N7'	$\bar{0} = s \vee (gs)' = s$
$\varepsilon_1$	$\neg A(\varepsilon xA) \rightarrow \neg A(t)$
$\varepsilon_2$	$t' = \varepsilon xA \rightarrow \neg A(t)$

We now define  $N_\varepsilon$  to be the theory  $\langle \mathcal{V}, \varepsilon(\mathcal{V}), \mathcal{A}_\varepsilon \rangle$  where as usual,  $\varepsilon(\mathcal{V})$  denotes the  $\varepsilon$ -calculus for  $\mathcal{V}$ . A formula of the form  $\varepsilon_1$  is called an  $\varepsilon_1$ -axiom and one of the form  $\varepsilon_2$  an  $\varepsilon_2$ -axiom. Note that the  $\varepsilon_1$ -axioms are theorems of the  $\varepsilon$ -calculus by virtue of axiom schemata Q3 and Q4. However, it is convenient to include formulae of this form as axioms of the theory in order to prove the eliminability of the quantifiers.

Recall that the choice function  $\Phi$  of the standard model  $\mathfrak{N}$  was defined as the least number operator. It is easy to see that with this interpretation of the  $\varepsilon$ -symbol the  $\varepsilon_2$ -axioms are all true. Consequently all the axioms of  $N_\varepsilon$  are true in  $\mathfrak{N}$ , and therefore by the soundness of the  $\varepsilon$ -calculus every theorem of  $N_\varepsilon$  is also true in  $\mathfrak{N}$ .

However, the least number operator is not the only interpretation of the  $\varepsilon$ -symbol which would satisfy the  $\varepsilon_2$ -axioms. These axioms merely state, intuitively, that the number denoted by  $\varepsilon xA$  is not the successor of a number having the property  $A$ . For example, if  $A$  is the quasi-formula

$$\exists y(y \cdot \bar{0}'' = x),$$

i.e., 'x is an even number', then there is no way of knowing which even number  $\varepsilon xA$  designates, since the  $\varepsilon_2$ -axiom  $\varepsilon xA = t' \rightarrow \neg A(t)$  merely states the obvious fact that the even number designated by  $\varepsilon xA$  is not the successor of an even number.

Despite the fact that the  $\varepsilon_2$ -axioms do not uniquely characterize the  $\varepsilon$ -symbol as a least number operator, we shall now see that these axioms are still strong enough to provide us with the principle of mathematical induction.

**THEOREM IV.1.** *If  $B$  is any formula of  $\mathcal{L}(\mathcal{V})$  of the form*

$$A(\bar{0}) \rightarrow \forall x(A(x) \rightarrow A(x')) \rightarrow \forall xA(x)$$

*then  $\mathcal{A}_\varepsilon \vdash_\varepsilon B$ . Furthermore, if  $B$  is a proper formula and  $x$  does not have a free occurrence in  $A(x)$  within the scope of an  $\varepsilon$ -symbol, then  $\mathcal{A}_\varepsilon \vdash_\varepsilon B$ .*

*Proof.* Let  $s$  be the term  $\varepsilon x \neg A(x)$  and let  $X$  be the set

$$\mathcal{A}_\varepsilon \cup \{A(\bar{0}), \forall x(A(x) \rightarrow A(x'))\}.$$

By virtue of the  $\forall$ -introduction rule and the  $\rightarrow$ -introduction rule it is sufficient to prove  $X \vdash_\varepsilon A(s)$ . The proof is as follows, where we let  $t$  be the term  $gs$ .

- |      |   |  |
|------|---|--|
| (1)  | $X \vdash \bar{0} = s \vee t' = s$                          | axiom N7'                                    |
| (2)  | $X \vdash A(\bar{0})$                                       | member of $X$                                |
| (3)  | $X \vdash (\bar{0} = s \wedge A(\bar{0})) \rightarrow A(s)$ | Theorem II.22                                |
| (4)  | $X \vdash \bar{0} = s \rightarrow A(s)$                     | tautology rule from (2) and (3)              |
| (5)  | $X \vdash \forall x(A(x) \rightarrow A(x'))$                | member of $X$                                |
| (6)  | $X \vdash A(t) \rightarrow A(t')$                           | $\forall$ -elimination                       |
| (7)  | $X \vdash (t' = s \wedge A(t')) \rightarrow A(s)$           | Theorem II.22                                |
| (8)  | $X \vdash t' = s \rightarrow \neg \neg A(t)$                | $\varepsilon_2$ -axiom and definition of $s$ |
| (9)  | $X \vdash t' = s \rightarrow A(s)$                          | tautology rule from (6), (7), and (8)        |
| (10) | $X \vdash A(s)$   | tautology rule from (1), (4), and (9)        |

For the special case where  $B$  is a proper formula and  $x$  does not have a free occurrence in  $A(x)$  within the scope of an  $\varepsilon$ -symbol, the proof is the same except that steps (3) and (7) are justified by Theorem III.5.

Often in mathematics it is simpler to use the principle of the least number than to use mathematical induction. Similarly, because of the  $\varepsilon_2$ -axioms it is often easier to prove theorems in  $N_\varepsilon$  than it is in  $N$ . For example, there exists a very simple proof in  $N_\varepsilon$  of the formula  $\forall x \neg \neg(x' = x)$ . Let  $t$  be the term  $\varepsilon x \neg \neg(x' = x)$ . Then the formula  $t' = t \rightarrow \neg \neg \neg(t' = t)$  is an  $\varepsilon_2$ -axiom. By the tautology rule  $\neg(t' = t)$  is a theorem of  $N_\varepsilon$ , and by the  $\forall$ -introduction rule,  $\forall x \neg \neg(x' = x)$  is also a theorem. To prove that this formula is a theorem in  $N$  one must use axioms N1, N2, and an induction axiom.

Theorem IV.1 can now be used to prove that  $N_\varepsilon$  is an extension of  $N$ .

**THEOREM IV.2.** *If  $A$  is a theorem of  $N$ , then  $\mathcal{A}_\varepsilon \vdash_\varepsilon A$ , and therefore  $A$  is a theorem of  $N_\varepsilon$ .*

*Proof.* Let  $\mathcal{D}$  be a proof of  $A$  in  $N$ . Thus  $\mathcal{D}$  is a deduction in the predicate calculus of  $A$  from some finite set  $X$  where each member of  $X$  is an ( $\varepsilon$ -free) axiom of  $N$ . By Theorem III.3, there exists a deduction  $\mathcal{D}_1$  of  $A$  from  $X$  in the  $\varepsilon^*$ -calculus. It is easy to see that for each  $B_i$  in  $X$ ,  $\mathcal{A}_\varepsilon \vdash_\varepsilon B_i$ . For, if  $B_i$  is one of the axioms N1–N7 of  $N$ , then  $B_i$  follows by the  $\forall$ -introduction rule from an instance of the corresponding axiom schema in  $N_\varepsilon$ , and if  $B_i$  is an induction axiom, then  $\mathcal{A}_\varepsilon \vdash_\varepsilon B_i$  by the  $\forall$ -introduction rule and Theorem IV.1. Consequently  $\mathcal{A}_\varepsilon \vdash_\varepsilon A$  by Theorem II.3(ii).

### EXERCISES

1. Prove that axiom schema N1' is redundant in the theory  $N_\varepsilon$ . In other words prove that any formula of the form  $\neg(\bar{0} = 1')$  is deducible from the other axioms of  $N_\varepsilon$ . Using finitary reasoning prove that axiom N1 is not redundant in the theory  $N$ . (*Hint:* Use the techniques employed in proving Theorem II.15.)
2. Prove that  $N_\varepsilon$  is an inessential extension of  $N$ , or in other words, prove that for any  $\varepsilon$ -free formula  $A$ ,  $A$  is a theorem of  $N$  if and only if  $A$  is a theorem of  $N_\varepsilon$ . (*Hint:* Adjoin the  $\iota$ -symbol and the appropriate rules of inference to the theory  $N$  to form the theory  $N_\iota$  (cf. Hilbert and Bernays [1934]). By replacing every  $\varepsilon$ -term,  $\varepsilon xA$ , by the  $\iota$ -term 'the least  $x$  such that  $A$ ' show that every  $\varepsilon$ -free theorem of  $N_\varepsilon$  is a theorem of  $N_\iota$ . The desired result now follows by Hilbert and Bernays' proof of the eliminability of the  $\iota$ -symbol.) Why is the Second  $\varepsilon$ -Theorem inadequate for solving this problem?

### 3.3 Quantifier-free proofs

We now return to the problem of proving the consistency of  $N$ . In view of Theorem IV.2, in order to prove the consistency of  $N$  it is sufficient to prove the consistency of  $N_\varepsilon$ . This result in itself does not really simplify the problem,

since  $N_\varepsilon$  is also a rather complicated theory. However, we shall now show that in fact it is sufficient to prove the consistency of a quite simple sub-theory of  $N_\varepsilon$ .

A well-formed expression of  $\mathcal{L}(\mathcal{V})$  is *quantifier-free* if the symbols  $\forall$  and  $\exists$  do not occur in that expression. A *quantifier-free proof* is now defined to be a proof in  $N_\varepsilon$  in which every formula is quantifier-free and in which the rule of relabelling bound variables may be used as a basic rule of inference. More precisely, a quantifier-free proof of  $B$  is a sequence  $\langle A_1, \dots, A_n \rangle$  of quantifier-free formulae of  $\mathcal{L}(\mathcal{V})$  such that  $A_n$  is  $B$  and for each  $i = 1, \dots, n$  at least one of the following is true:

1.  $A_i$  is an instance of one of the axiom schemata P1–P10, E1, E3, N1'–N7',  $\varepsilon_1$ , or  $\varepsilon_2$ ;
2.  $A_i$  follows by modus ponens from  $A_j$  and  $A_k$  for some  $j, k < i$ ;
3.  $A_i$  is a variant of  $A_j$  for some  $j < i$ .

(Recall that  $A$  is a *variant* of  $B$  if  $A$  can be obtained from  $B$  by a succession of admissible relabellings of bound variables.) Strictly speaking, a quantifier-free proof of  $B$  is not a proof in  $N_\varepsilon$ , since the rule of relabelling bound variables is only a derived rule of inference in the  $\varepsilon$ -calculus. However, it is obvious that any quantifier-free proof of  $B$  can be converted into a proof of  $B$  in  $N_\varepsilon$ . On the other hand, a proof in  $N_\varepsilon$  of some quantifier-free formula  $B$  cannot necessarily be converted into a quantifier-free proof of  $B$ , since within a quantifier-free proof there are no formal counterparts of the E2-axioms. Nonetheless we can prove the following weaker result which is all that is needed to simplify the problem of proving the consistency of  $N$ .

**THEOREM IV.3.** *For any elementary formula  $B$ , if  $B$  is a theorem of  $N$ , then there exists a quantifier-free proof of  $B$ .*

*Proof.* Since  $B$  is a theorem of  $N$ , then by Theorem IV.2 there exists a deduction  $\mathcal{D}$  in the  $\varepsilon^*$ -calculus of  $B$  from  $\mathcal{A}_\varepsilon$ . We now eliminate the quantifiers from  $\mathcal{D}$  as follows.

First of all we replace every occurrence of  $\forall$  by  $\neg \exists \neg$ , thus converting every Q1-axiom and Q2-axiom into a tautology without damaging any of the other axioms, and without affecting  $B$  since  $B$  is elementary. (Recall that E2-axioms are not used in the  $\varepsilon^*$ -calculus.) In this way we obtain a proof  $\mathcal{D}_1$  of  $B$  in  $N_\varepsilon$  such that no formula in  $\mathcal{D}_1$  contains the symbol  $\forall$ .

The following procedure can now be used to eliminate every occurrence of the symbol  $\exists$ . We start with some quasi-formula of the form  $\exists xA$  in  $\mathcal{D}_1$  where  $A$  is quantifier-free. If  $\varepsilon xA$  is free for  $x$  in  $A$ , we replace every occurrence of  $\exists xA$  in  $\mathcal{D}_1$  by  $A(\varepsilon xA)$ . On the other hand, suppose that  $\varepsilon xA$  is not free for  $x$  in  $A$ . For example,  $A$  might be  $y = \varepsilon y(y = x)$ . Let  $\bar{A}$  be some variant of  $A$  such that  $\varepsilon x\bar{A}$  is free for  $x$  in  $\bar{A}$ . For example, if  $A$  is  $y = \varepsilon y(y = x)$ , let  $\bar{A}$

be  $y = \varepsilon z(z = x)$ . We may now replace every occurrence of  $\exists xA$  in  $\mathcal{D}_1$  by  $\bar{A}(\varepsilon x\bar{A})$ . Successive applications of this procedure will eventually yield a sequence  $\mathcal{D}_n$  of quantifier-free formulae whose last member is the original formula  $B$ . Every Q3-axiom in  $\mathcal{D}_n$  will have been converted into an  $\varepsilon_1$ -axiom (or a variant of an  $\varepsilon_1$ -axiom), every Q4-axiom will have been converted into a tautology (or a variant of a tautology), and the remaining axioms will have been converted into axioms of the same form (or variants of such axioms). Consequently we obtain a quantifier-free proof of  $B$ . This completes the proof.

The following example reveals why it is necessary in the above proof to ensure that  $\varepsilon xA$  is free for  $x$  in  $A$  before replacing  $\exists xA$  by  $A(\varepsilon xA)$ . Let  $A$  be  $y = \varepsilon y(y = x)$  and let  $A_1$  be  $t = \varepsilon y(y = x)$  where  $t$  is some quantifier-free term. Then the formula

$$t' = \varepsilon y \exists x A \rightarrow \neg \exists x A_1$$

is an  $\varepsilon_2$ -axiom. If we were to eliminate the symbol  $\exists$  from this axiom simply by replacing  $\exists xA$  by  $A(\varepsilon xA)$  and  $\exists xA_1$  by  $A_1(\varepsilon xA_1)$  the resulting formula

$$t' = \varepsilon y A(\varepsilon xA) \rightarrow \neg A_1(\varepsilon xA_1)$$

would not be an  $\varepsilon_2$ -axiom, since a free occurrence of  $y$  in  $\varepsilon xA$  becomes bound in  $A(\varepsilon xA)$  and therefore the formula  $\neg A_1(\varepsilon xA_1)$  is not of the form  $\neg[A(\varepsilon xA)]_s^y$  as it should be.

*Historical Note:* Theorem IV.3 is an improvement on the corresponding result proved by Hilbert and Bernays [1939] since their  $\varepsilon$ -axioms include all formulae of the form

$$\varepsilon_3 \quad s = t \rightarrow [\varepsilon xA]_s^y = [\varepsilon xA]_s^y$$

as well as the  $\varepsilon_1$ -axioms and  $\varepsilon_2$ -axioms. Clearly, such formulae are weaker versions of the E2-axioms.

### 3.4 The consistency of arithmetic

It now follows that in order to prove the consistency of arithmetic, as it is formalized by the theory N, it is sufficient to prove the following proposition.

**PROPOSITION I.** *For any elementary formula  $B$  of  $\mathcal{L}(\mathcal{V})$ , if there exists a quantifier-free proof of  $B$ , then  $B$  is numerically true.*

Once this proposition has been proved, it then follows by Theorem IV.3 that every elementary theorem of N is numerically true. Since the elementary formula  $f$  is not numerically true, then  $f$  is not a theorem of N, and N is consistent. We know by Gödel's results that the above proposition cannot be proved by finitary reasoning since we would then have a finitary proof of the consistency of arithmetic. However, Hilbert and Bernays [1939] show

that a finitary proof is possible if a certain condition is imposed on the  $\varepsilon_2$ -axioms. We shall now see what that condition is and what effect it has on the original theory N.

For any  $\varepsilon_1$ -axiom  $\neg A(\varepsilon xA) \rightarrow \neg A(t)$  or any  $\varepsilon_2$ -axiom  $t' = \varepsilon xA \rightarrow \neg A(t)$ , the term  $\varepsilon xA$  is said to *belong* to that axiom. The *rank* of an  $\varepsilon_2$ -axiom is now defined to be the rank of the  $\varepsilon$ -term belonging to it. It follows from our definition of rank (page 70) that the rank of an  $\varepsilon_2$ -axiom,

$$t' = \varepsilon xA \rightarrow \neg A(t),$$

equals 1 if and only if  $x$  does not have a free occurrence in  $A$  within the scope of an  $\varepsilon$ -symbol. We can now give an exact statement of the result which Hilbert and Bernays prove.

**PROPOSITION II.** *For any elementary formula  $B$  of  $\mathcal{L}(\mathcal{V})$ , if there exists a quantifier-free proof of  $B$  in which the rank of every  $\varepsilon_2$ -axiom equals 1, then  $B$  is numerically true.*

If we carry this restriction on the  $\varepsilon_2$ -axioms back to the original theory N, it follows that N is consistent provided that we define the induction axioms to be all universal closures of quasi-formulae of the form

$$A(\bar{0}) \rightarrow \forall x(A(x) \rightarrow A(x')) \rightarrow \forall xA(x)$$

where  $x$  does not have a free occurrence in  $A(x)$  within the scope of a quantifier. We shall refer to this weakened version of N as *restricted arithmetic*.

The proof of Proposition II is rather long and complicated, and instead of giving all the details we shall merely outline the basic ideas.

Let  $B$  be any elementary formula of  $\mathcal{L}(\mathcal{V})$  and let  $\mathcal{D}$  be a quantifier-free proof of  $B$  in which the rank of every  $\varepsilon_2$ -axiom equals 1. We would like to prove that  $B$  is numerically true. First of all, using essentially the same techniques which we used in proving the Rank Reduction Theorem we can eliminate from  $\mathcal{D}$  all the  $\varepsilon$ -terms with rank  $> 1$ . Consequently, we may assume that the rank of every  $\varepsilon$ -term occurring in  $\mathcal{D}$  equals 1. In order to prove that  $B$  is numerically true, we would like to replace each  $\varepsilon$ -term in  $\mathcal{D}$  by an appropriate  $\varepsilon$ -free term and then show that this total replacement (*Gesamtersetzung*) of  $\varepsilon$ -terms by  $\varepsilon$ -free terms converts every formula in  $\mathcal{D}$  into a numerically true formula. Since the instances of axiom schemata P1–P10, E1, E3, N1'–N7' are converted into numerically true formulae no matter what terms are used to replace the  $\varepsilon$ -terms, our only concern is in finding a total replacement of  $\varepsilon$ -terms which will convert the  $\varepsilon_1$ -axioms and  $\varepsilon_2$ -axioms into numerically true formulae. (Because of the rule of relabelling bound variables we must also make sure that any two  $\varepsilon$ -terms which are variants of each other are replaced by the same term. This presents no difficulty if we simply regard any two such terms as being the same term.)

To illustrate how such a total replacement can be found we shall consider

the special case where no  $\varepsilon$ -term in  $\mathcal{D}$  contains another  $\varepsilon$ -term as a subterm. (In the terminology of Hilbert and Bernays this is the case where the *degree* of every  $\varepsilon$ -term in  $\mathcal{D}$  equals 1.) Since the rank of every  $\varepsilon$ -term in  $\mathcal{D}$  equals 1, it follows from this additional restriction that if  $\varepsilon xA$  occurs in  $\mathcal{D}$ , then  $A$  is  $\varepsilon$ -free. In order to find a total replacement of  $\varepsilon$ -terms which will convert the  $\varepsilon_1$ -axioms and  $\varepsilon_2$ -axioms into numerically true formulae we proceed by trial and error. Let  $R_1$  be the total replacement whereby each  $\varepsilon$ -term is replaced by  $\bar{0}$ . Notice that every  $\varepsilon_2$ -axiom is thereby converted into a numerically true formula of the form

$$t' = \bar{0} \rightarrow \neg A(t).$$

If  $R_1$  works, then we are done. If not, then certain of the  $\varepsilon_1$ -axioms must be to blame. Let

$$\neg A(\varepsilon xA) \rightarrow \neg A(t)$$

be one of the offending  $\varepsilon_1$ -axioms. Thus  $R_1$  converts this axiom into the numerically false formula

$$\neg A(\bar{0}) \rightarrow \neg A(s),$$

where  $s$  is obtained from  $t$  by replacing any  $\varepsilon$ -terms in  $t$  by  $\bar{0}$ . Notice that the quasi-formula  $A$  is unaffected by the replacement  $R_1$  since  $A$  is  $\varepsilon$ -free. We now compute the numerical value of  $s$  and the truth values of the formulae

$$A(\bar{0}), A(\bar{0}'), A(\bar{0}''), \dots, A(\bar{0}^{(n)})$$

where  $n$  is the numerical value of  $s$ . Let  $A(r)$  be the first formula in the series which is numerically true. Such a formula exists since  $\neg A(\bar{0}) \rightarrow \neg A(s)$  is numerically false and therefore the formula  $A(\bar{0}^{(n)})$  is numerically true. We shall refer to the term  $r$  as the *minimal value* of  $\varepsilon xA$ . It is easy to see that if  $\varepsilon xA$  is replaced by  $r$ , then the  $\varepsilon_1$ -axioms and  $\varepsilon_2$ -axioms to which  $\varepsilon xA$  belongs are converted into numerically true formulae no matter what terms are used to replace the other  $\varepsilon$ -terms. Let  $R_2$  be the total replacement whereby  $\varepsilon xA$  (and all its variants) are replaced by  $r$  and all the other  $\varepsilon$ -terms are replaced by  $\bar{0}$ . If  $R_2$  works, we are done. If not, we calculate the minimal value of one of the offending  $\varepsilon$ -terms and then define a new total replacement  $R_3$  as before. After at most  $m + 1$  such attempts, where  $m$  is the number of  $\varepsilon$ -terms in  $\mathcal{D}$ , we arrive at a total replacement which converts every formula in  $\mathcal{D}$  into a numerically true formula, thus proving that  $B$  is numerically true.

Of course the above restriction on the 'degree' of the  $\varepsilon$ -terms in  $\mathcal{D}$  constitutes a very special case. In the general case the  $\varepsilon$ -free term which is used to replace one  $\varepsilon$ -term will depend on the terms which are used to replace the subterms of that term. However Hilbert and Bernays show that the basic ideas which we have just used in dealing with the special case can be applied to the general case to provide a total replacement which converts all the

formulae in  $\mathcal{D}$  into numerically true formulae, assuming that every  $\varepsilon$ -term in  $\mathcal{D}$  has rank = 1. In this way they prove Proposition II and therefore the consistency of restricted arithmetic.

The basic ideas underlying Hilbert and Bernays' proof were originally devised by Ackermann [1924] in his doctoral dissertation written under Hilbert. As we have already mentioned, this was the first published work in which the  $\varepsilon$ -symbol was used. It is interesting to point out that Ackermann's dissertation was intended to prove the consistency of analysis. However, at the time of publication an error was discovered which invalidated many of the results. In order to correct the error Ackermann introduced a footnote (page 9) which severely restricted the formal theory he was dealing with. The proof of the consistency of restricted arithmetic which is given in Hilbert and Bernays [1939] and which we have just described is based on a letter from Ackermann to Bernays in which Ackermann clarifies and develops the methods used in his dissertation. Other finitary proofs of the consistency of restricted arithmetic were discovered by von Neumann [1927] and Herbrand [1931]. (For further historical details and for English translations of many important papers and lectures by Hilbert, Bernays, von Neumann, Ackermann, and Herbrand, see van Heijenoort [1967].)

The formalists' attempts to find a finitary proof of the consistency of unrestricted arithmetic came to an end in 1931 with the publication of Gödel's famous paper. By modifying Gödel's argument, Hilbert and Bernays [1939], pages 324–340, show that no proof of the consistency of  $N$  can be formalized within  $N$ . Thus any consistency proof must in some way or other involve techniques which cannot be formalized in  $N$ . Gentzen [1936] and Ackermann [1940] have constructed such consistency proofs by using transfinite induction up to the first  $\varepsilon$ -number (the first ordinal  $\alpha$  such that  $\omega^\alpha = \alpha$ ).

Ackermann's proof is based on an extension of his earlier methods. By using transfinite induction he essentially shows that given any array of  $\varepsilon_1$ -axioms and  $\varepsilon_2$ -axioms it is possible to assign numerals to the  $\varepsilon$ -terms so that these axioms are all converted into numerically true formulae, thus proving Proposition I and the consistency of  $N$ . An exposition of Ackermann's proof is given by Wang [1963], pages 362–370.

More recently, Tait [1965] has used recursive function theory to formalize the elimination of  $\varepsilon$ -terms from quantifier-free formalizations of arithmetic. He proves that if  $S$  is arithmetic with induction up to some ordinal  $\xi$ , then the elimination of  $\varepsilon$ -terms from proofs in  $S$  can be achieved by using second order functionals defined by transfinite recursion up to  $\xi, \xi^\xi, \xi^{\xi^\xi}, \dots$  etc. This theorem sharpens the results of Ackermann, Hilbert, and Bernays by making more explicit the metatheory which is used in eliminating the  $\varepsilon$ -symbol.

Gentzen's proof [1936] of the consistency of arithmetic depends on the eliminability of a certain rule of inference, the cut rule, and does not involve

the  $\varepsilon$ -symbol. In many ways this method of proof is more straightforward than that of Ackermann. A simple version of this proof is given by Mendelson [1964], pages 258–270.

#### 4 Theories based on the $\varepsilon$ -calculus

The  $\varepsilon$ -calculus is often used instead of the predicate calculus in formalizing certain mathematical theories, particularly arithmetic and set theory. The advantages of using the  $\varepsilon$ -calculus in this way are many. We have already discussed how the  $\varepsilon$ -calculus can be used to prove the consistency of arithmetic. We shall now consider how the  $\varepsilon$ -symbol can simplify the actual formulation of a theory. It has been seen in Chapter II, §12 that the formulation of the underlying logic is simplified by the availability of the  $\varepsilon$ -symbol, since, for example, the  $\varepsilon$ -calculus provides simple derived rules of inference for the introduction and elimination of quantifiers. Three other advantages of using the  $\varepsilon$ -calculus as the basis of a theory are: (i) the  $\iota$ -symbol is superfluous, since its role is assumed by the  $\varepsilon$ -symbol; (ii) Skolem functions can be explicitly defined as  $\varepsilon$ -terms; (iii) the  $\varepsilon$ -symbol can be used to define certain entities and concepts whose intended interpretations are to some extent indefinite.

If one is to use the  $\varepsilon$ -calculus rather than the predicate calculus in formalizing some mathematical theory one would like to know how this formalization compares with the corresponding formalization based on the predicate calculus. Obviously, the Second  $\varepsilon$ -Theorem provides the following answer to this question.

**THEOREM IV.4.** *Let  $\mathcal{T}$  be the theory  $\langle \mathcal{V}, \text{PC}(\mathcal{V}), \mathcal{A} \rangle$  where every member of  $\mathcal{A}$  is  $\varepsilon$ -free, and let  $\mathcal{T}_\varepsilon$  be the theory  $\langle \mathcal{V}, \varepsilon(\mathcal{V}), \mathcal{A} \rangle$ . Then for any  $\varepsilon$ -free formula  $A$  of  $\mathcal{L}(\mathcal{V})$ ,  $A$  is a theorem of  $\mathcal{T}$  if and only if  $A$  is a theorem of  $\mathcal{T}_\varepsilon$ . In other words,  $\mathcal{T}_\varepsilon$  is an inessential extension of  $\mathcal{T}$ .*

*Proof.* If  $A$  is a theorem of  $\mathcal{T}$ , then by Theorem III.2,  $A$  is a theorem of  $\mathcal{T}_\varepsilon$ . Conversely, if  $A$  is a theorem of  $\mathcal{T}_\varepsilon$  then by the Second  $\varepsilon$ -Theorem  $A$  is a theorem of  $\mathcal{T}$ .

It is important to notice that the above theorem would not hold if we removed the condition that every member of  $\mathcal{A}$  is  $\varepsilon$ -free, since the Second  $\varepsilon$ -Theorem could no longer be used. We shall return to this important point when we discuss formalizations of set theory which are based on the  $\varepsilon$ -calculus.

We shall now look at some of the simplifications which the  $\varepsilon$ -calculus provides.

##### 4.1 The $\iota$ -symbol

It is often desirable to have within a formal theory  $\mathcal{T}$  some way of designating 'the unique  $x$  such that  $A$ '. The  $\iota$ -symbol was introduced for just this

reason. The formal treatment of this symbol is given by the following  $\iota$ -rule (cf. Hilbert and Bernays [1934] page 384):

If the formulae

- (1)  $\exists xA$
- (2)  $\forall x\forall y((A \wedge A_x^y) \rightarrow x = y)$

are theorems of a theory  $\mathcal{T}$ , then  $\iota xA$  is a term of  $\mathcal{T}$  and the formula

- (3)  $A(\iota xA)$

is a theorem of  $\mathcal{T}$ .

Formula (1) is called the *existence condition* and formula (2) the *uniqueness condition*.

If the theory  $\mathcal{T}$  is based on the  $\varepsilon$ -calculus, then the  $\iota$ -symbol and  $\iota$ -rule are superfluous since one may replace  $\iota xA$  by  $\varepsilon xA$ . In this way (3) follows from (1) by the  $\exists$ -elimination rule.

One objection to the above treatment of the  $\iota$ -symbol is that when the  $\iota$ -rule is adjoined the concept of a term becomes undecidable since there may be no way of knowing whether or not formulae (1) and (2) are theorems (cf. Bernays [1958], page 49). For this reason the following approach is often used. We write  $\exists! xA$  as an abbreviation for

$$\exists x(A \wedge \forall y(A_y^x \rightarrow x = y))$$

where  $y$  is not free in  $A$ . Thus  $\exists! xA$  may be read as 'there exists a unique  $x$  such that  $A$ '. Notice that  $\exists! xA$  is logically equivalent to the conjunction of formulae (1) and (2). We adopt the  $\iota$ -symbol as a new logical symbol of the language and we enlarge the rules of formation of the language so that for any quasi-formula  $A$  and any variable  $x$ , the expression  $\iota xA$  is a well-formed quasi-term. We then adjoin all formulae of the following forms as additional axioms of the theory:

- (4)  $\exists! xA \rightarrow A(\iota xA)$
- (5)  $\neg \exists! xA \rightarrow \iota xA = t,$

where  $t$  is some specified term of the language such as the symbol  $\bar{0}$  in arithmetic or the term denoting the empty set in set theory. Intuitively, these axioms say that if there exists a unique  $x$  such that  $A$ , then  $\iota xA$  designates that unique object, and if not, then  $\iota xA$  designates whatever  $t$  designates.

Once again, however, if our theory is based on the  $\varepsilon$ -calculus, there is no need to adjoin the  $\iota$ -symbol as a new logical symbol and adopt formulae (4) and (5) as additional axioms, since the quasi  $\iota$ -terms may be defined in terms of the  $\varepsilon$ -symbol as follows:

$$\iota xA =_{\text{Df}} \varepsilon x((\exists! xA \wedge A) \vee (\neg \exists! xA \wedge x = t)),$$

where  $t$  is the term in (5). In other words we regard  $\iota xA$  as the formal abbreviation of the expression on the right. With this definition of the  $\iota$ -symbol it is easy to prove that all formulae of the forms (4) and (5) above are theorems of the  $\varepsilon$ -calculus and hence of our theory. To see this observe that any formula of the form

$$(6) \quad \exists x((\exists!x A \wedge A) \vee (\neg \exists!x A \wedge x = t))$$

is a theorem of the  $\varepsilon$ -calculus, and therefore by the  $\exists$ -elimination rule and our definition of  $\iota xA$ , the formula

$$(7) \quad (\exists!x A \wedge A(\iota x A)) \vee (\neg \exists!x A \wedge \iota x A = t)$$

is a theorem. Formulae (4) and (5) follow from (7) by the tautology rule.

### EXERCISE

Prove that any formula of the form (6) is a theorem of the  $\varepsilon$ -calculus.

### 4.2 Skolem functions

In §2 of this chapter we saw that by replacing the axioms of a theory  $\mathcal{T}$  by their Skolem resolutions and by adjoining the new Skolem functions to the vocabulary one can sometimes find a proof of the consistency of  $\mathcal{T}$ . This elimination of the existential quantifiers from the axioms of  $\mathcal{T}$  also provides a more practical formulation of the theory itself. For example, in set theory instead of stating the power set axiom in the usual existential form,

$$(1) \quad \forall x \exists y \forall z (z \in y \leftrightarrow z \subseteq x),$$

it is preferable to state it in the form

$$(2) \quad \forall x \forall z (z \in \pi(x) \leftrightarrow z \subseteq x),$$

where the new function symbol  $\pi$  is taken as a primitive (cf. the system of Bernays [1958]). The availability of this symbol makes it possible to designate within the theory the power set of any given set. If, however, the theory is based on the  $\varepsilon$ -calculus, there is no need to take the formula (2) instead of (1) as an axiom, since we can define  $\pi(t)$  for any  $t$  as follows

$$\pi(t) =_{\text{Df}} \varepsilon y \forall z (z \in y \leftrightarrow z \subseteq t).$$

Under this definition of  $\pi$ , formula (2) follows from (1) by the  $\exists$ -elimination rule. Thus, in general, if a theory  $\mathcal{T}_\varepsilon$  is based on the  $\varepsilon$ -calculus, the axioms may be stated in the weaker existential form (or unresolved form) and the Skolem functions for the existential variables may be introduced by explicit definitions in terms of the  $\varepsilon$ -symbol. By Theorem IV.4 it follows that the introduction of these new function symbols does not essentially strengthen the theory, since any formula not containing these symbols which is a theorem

of  $\mathcal{T}_\varepsilon$  is a theorem of the corresponding theory  $\mathcal{T}$  which is based on the predicate calculus (assuming that the axioms of  $\mathcal{T}_\varepsilon$  are  $\varepsilon$ -free).

This method of eliminating existential quantifiers by means of  $\varepsilon$ -terms can be used to simplify not only the axioms of  $\mathcal{T}_\varepsilon$ , but also the theorems. For example, suppose the formula

$$(3) \quad \exists y \forall z (z \in y \leftrightarrow B)$$

is a theorem of some formalization of set theory, where  $B$  is a quasi-formula not containing a free occurrence of  $y$ . Formula (3) asserts that there exists a set  $y$  whose members are those entities which satisfy  $B$ . Using the terminology of Bourbaki (1954), we say ' $B$  is collectivizing in  $z$ '. It is convenient, though not necessary, to have some formal apparatus for designating this set. Such designations are possible if, for example, the primitive symbols of the theory include the 'comprehension operator'  $\hat{\ }^z$ , where the term  $\hat{z}B$  denotes the set whose members are the entities which satisfy  $B$ . If, however, the  $\varepsilon$ -symbol is available, the operator  $\hat{\ }^z$  need not be taken as a primitive, but can be defined explicitly as follows

$$(4) \quad \hat{z}B =_{\text{Df}} \varepsilon y \forall z (z \in y \leftrightarrow B),$$

where  $B$  is any quasi-formula and  $y$  any variable which does not occur free in  $B$ . Then by formula (3) and the  $\exists$ -elimination rule we obtain the formula

$$(5) \quad \forall z (z \in \hat{z}B \leftrightarrow B),$$

which asserts that the term  $\hat{z}B$  does designate the required set. It is interesting to note that definition (4) may be used for any quasi-formula  $B$ , even if  $B$  is not collectivizing in  $z$ . For example,  $B$  may be the quasi-formula  $\neg z \in z$ . This definition of  $\hat{z}B$  does not introduce any contradictions, such as Russell's paradox, since formula (5) depends on (3), i.e. on the fact that  $B$  is collectivizing in  $z$ . If  $B$  is not collectivizing in  $z$ , the expression  $\hat{z}B$  as defined by (4) is still a well-formed term of the language, but nothing very much can be said about it (cf. Bourbaki [1954] p. 63). In this case  $\hat{z}B$  is a 'null term' (see Chapter I, page 54).

We shall return to this subject in a later section (see page 107) where a formal system similar to that used by Bernays [1958] is presented in which the operator  $\hat{\ }^z$  is taken as a primitive symbol and Church's schema

$$\forall z (z \in \hat{z}B \leftrightarrow B)$$

is taken as an axiom schema.

### 4.3 Definability of indeterminate concepts

The above definition of the comprehension operator is used by both Bourbaki [1954], p. 63, and Ackermann [1937-8]. Clearly, the  $\iota$ -symbol



could be used instead of the  $\varepsilon$ -symbol in formulating this definition, since by the axiom of extensionality,

$$\forall x \forall y (x = y \leftrightarrow \forall z (z \in x \leftrightarrow z \in y)),$$

formula (3) satisfies the uniqueness condition. However, if an existential formula  $\exists x A$  does not satisfy the uniqueness condition, then in order to designate some object which satisfies  $A$ , one must use the  $\varepsilon$ -symbol. In this case the designation,  $\varepsilon x A$ , has a certain degree of indeterminacy since as we have observed before, nothing definite can be said about  $\varepsilon x A$  except that it satisfies  $A$  if anything does. It more than one entity satisfies  $A$ , then there is no way of knowing exactly which of these objects  $\varepsilon x A$  designates. Occasionally, it is desirable to define objects which have just this degree of indeterminacy. We shall now consider two such occasions.

In set theory, the concept of cardinal number is difficult to formulate explicitly since the intended interpretation of this concept is rather indefinite. All that we require of the definition is that the following formula be a theorem

$$(1) \quad \forall x \forall y (x \sim y \leftrightarrow \bar{x} = \bar{y}),$$

where  $\bar{x}$  and  $\bar{y}$  are, respectively, the cardinal numbers of  $x$  and  $y$ , and the expression  $x \sim y$  is an abbreviation of the assertion that there exists a one-to-one correspondence between  $x$  and  $y$ . Various definitions of  $=$  can be found in the literature, but probably the simplest is the following, which is used by Bourbaki and Ackermann:

$$(2) \quad \bar{t} =_{\text{Df}} \varepsilon z (z \sim t),$$

where  $t$  is any quasi term and  $z$  does not occur free in  $t$ . Using the fact that  $\sim$  is an equivalence relation, it is a simple matter to prove (1). The proof is as follows. Let  $s$  and  $t$  be any two terms. Since  $t \sim t$ , then  $\exists z (z \sim t)$ , and consequently,  $\varepsilon z (z \sim t) \sim t$ , i.e.

$$(3) \quad \bar{t} \sim t.$$

Similarly, we get

$$(4) \quad \bar{s} \sim s.$$

From (3) and (4) and the fact that  $\sim$  is an equivalence relation we obtain

$$(5) \quad \bar{s} = \bar{t} \rightarrow s \sim t.$$

On the other hand, the fact that  $\sim$  is an equivalence relation implies

$$(6) \quad s \sim t \rightarrow \forall z (z \sim s \leftrightarrow z \sim t).$$

By axiom schema E2, we obtain

$$(7) \quad \forall z (z \sim s \leftrightarrow z \sim t) \rightarrow \varepsilon z (z \sim s) = \varepsilon z (z \sim t).$$

Therefore, (6) and (7) yield

$$(8) \quad s \sim t \rightarrow \bar{s} = \bar{t}.$$

Consequently, from (5) and (8) we get

$$(9) \quad s \sim t \leftrightarrow \bar{s} = \bar{t}.$$

Formula (1) now follows from (9) by the  $\forall$ -introduction rule.

This definition of cardinal number is an essentially indeterminate one since nothing can be said about the set  $\bar{t}$  in (2) except that it is equivalent to  $t$  and that it equals the cardinal number of any set which is equivalent to  $t$ . It is just this degree of indeterminacy which we want the concept of cardinal number to possess. Ackermann [1937-8] has observed that for any equivalence relation  $\sim$ , definition (2) can be used to specify a representative element from each equivalence class of  $\sim$ . This method of designating a completely arbitrary representative of an equivalence class could have useful applications in the formulation of various mathematical theories, particularly in the introduction of 'definitions by abstraction' (cf. Beth [1959], pp. 91-95).

The basic indeterminacy of  $\varepsilon$ -terms is also used by Carnap [1961] to overcome certain difficulties which arise in the formulation of theories of empirical science. Such theories include certain terms, called the 'theoretical terms' or 'T-terms', which represent the 'unobservables' of the theory (e.g. 'temperature', 'electric field', etc.). The interpretation of these terms is provided by the postulates of the theory. These postulates are of two kinds, the 'theoretical postulates' ('T-postulates') and the 'correspondence postulates' ('C-postulates'). However, these postulates do not provide a complete interpretation of the T-terms 'because the scientist can always add further C-postulates (e.g., operational rules for T-terms) or T-postulates and thereby increase the specification of the meanings of the T-terms'. Because of the indeterminacy of these terms, the following problem arises: how can one give explicit definitions of the T-terms which satisfy the postulates without contributing anything new to the factual content of the theory? Carnap solves this problem by defining the T-terms as  $\varepsilon$ -terms, thereby obtaining just the intended degree of indeterminacy.

#### 4.4 The $\varepsilon$ -symbol and the axiom of choice

Finally, the indeterminate nature of the  $\varepsilon$ -symbol helps to explain the close connection which exists between this symbol and the axiom of choice. The axiom of choice differs from the other axioms of set theory in that these other axioms (e.g., the axioms of power set, pairing, replacement, etc.) not only assert the existence of a new set, but also specify the members of this set. The axiom of choice, on the other hand, merely asserts the existence of a selection set  $y$  for a given set  $x$  without actually specifying the members of  $y$ . Similarly, the quasi  $\varepsilon$ -term  $\varepsilon u (u \in w)$  expresses a choice function for the set  $x$  as the variable  $w$  ranges over  $x$ , but there is no way of knowing which member of  $w$  is being selected. For these reasons the  $\varepsilon$ -symbol and Q4 axioms

are often regarded as logical counterparts of the axiom of choice. However, it does not necessarily follow that the axiom of choice is derivable in a formalization of set theory which is based on the  $\varepsilon$ -calculus. For, suppose that  $\mathcal{T}$  is a set theory based on the predicate calculus, and  $\mathcal{T}_\varepsilon$  is the corresponding theory based on the  $\varepsilon$ -calculus, i.e.  $\mathcal{T}$  and  $\mathcal{T}_\varepsilon$  have the same set of axioms. Then by Theorem IV.4 (or the Second  $\varepsilon$ -Theorem), if the axiom of choice is not a theorem of  $\mathcal{T}$ , then it is not a theorem of  $\mathcal{T}_\varepsilon$ , even though this latter theory is based on the  $\varepsilon$ -calculus.

The intuitive explanation of why the  $\varepsilon$ -symbol and  $\varepsilon$ -axioms do not necessarily yield the axiom of choice is as follows. Although the quasi  $\varepsilon$ -term  $\varepsilon u(u \in w)$  can be used to make a simultaneous selection from each member of a given set  $x$ , it does not necessarily follow that there exists a set  $y$  consisting of these selected entities. The axiom of choice, on the other hand, does assert the existence of the selection set  $y$ . Wang [1955] has observed that if the formula

$$(1) \quad \forall x \exists y \forall z (z \in y \leftrightarrow \exists w (w \in x \wedge z = \varepsilon u(u \in w)))$$

is a theorem of  $\mathcal{T}_\varepsilon$ , then the  $\varepsilon$ -axioms do yield the axiom of choice. Clearly, formula (1) asserts that for any  $x$ , there exists a set  $y$  whose members are the selected entities from each member of  $x$ .

In most set theories, the axioms include the instances of a certain axiom schema, usually referred to as the axiom schema of replacement. If the theory is based on the  $\varepsilon$ -calculus, then the question of the deducibility of the axiom of choice usually hinges on whether the axioms include *all* the instances of this schema, or just those instances which are  $\varepsilon$ -free. In the former case, the  $\varepsilon$ -symbol does provide the axiom of choice, but in the latter case it does not.

For example, consider the set theory of Bourbaki [1954]. This system is based on an  $\varepsilon$ -calculus which is virtually equivalent to ours. (The differences are that the Greek letter  $\tau$ , instead of  $\varepsilon$ , is used for the selection operator, and the quantifiers are defined in terms of  $\tau$  instead of being taken as primitive symbols.) The theory has the following axiom schema:

$$S8 \quad \forall w \exists y \forall z (A \rightarrow z \in y) \rightarrow \forall x \exists y \forall z (z \in y \leftrightarrow \exists w (w \in x \wedge A)),$$

where  $A$  is any quasi-formula not containing free occurrences of  $x$  and  $y$ , and  $w$ ,  $x$ ,  $y$ , and  $z$  are all distinct. Letting  $A$  be the quasi-formula  $z = \varepsilon u(u \in w)$ , S8 yields the following theorem (cf. Bourbaki, p. 66, C53):

$$\forall x \exists y \forall z (z \in y \leftrightarrow \exists w (w \in x \wedge z = \varepsilon u(u \in w))).$$

Consequently, the axiom of choice is derivable in this system. On the other hand, if the axiom schema S8 were subject to the restriction that the quasi-formula  $A$  must be  $\varepsilon$ -free, then all the axioms would be  $\varepsilon$ -free, and by the Second  $\varepsilon$ -Theorem the axiom of choice would not be derivable in this system

(since it is known that this axiom is independent of the other axioms of a set theory based on the predicate calculus).

Wang [1955], pp. 66–67 points out another distinction between the  $\varepsilon$ -symbol and the axiom of choice:

‘There are also cases where, although the  $\varepsilon$ -rule would yield the desired result, the axiom of choice would not. For example, in the Zermelo theory we can infer “ $(x)R(x, \varepsilon y Rxy)$ ” from “ $(x)(\exists y)Rxy$ ” by the  $\varepsilon$ -rule, but we cannot infer “there exists  $f$ ,  $(x)R(x, f/x)$ ” from “ $(x)(\exists y)Rxy$ ” by the axiom of choice, on account of the absence of a universal set in Zermelo’s theory.’

In other words, the  $\varepsilon$ -symbol provides us with a ‘universal choice function’ which is defined on the class of all sets. The existence of such a function is normally provided by only the strongest forms of the axiom of choice (e.g., axiom E in Gödel [1940] and axiom  $A_\varepsilon$  in Bernays [1958]). We shall return to this subject in § 5.2.

In conclusion, if  $\mathcal{T}_\varepsilon$  is some theory based on the  $\varepsilon$ -calculus, then various simplifying definitions and processes can be formulated within  $\mathcal{T}_\varepsilon$ , which could not be formulated within the corresponding theory  $\mathcal{T}$ , based on the predicate calculus. However, if all the axioms of  $\mathcal{T}_\varepsilon$  are  $\varepsilon$ -free, then by the eliminability of the  $\varepsilon$ -symbol, every  $\varepsilon$ -free theorem of  $\mathcal{T}_\varepsilon$  is also a theorem of  $\mathcal{T}$ . Thus the introduction of the  $\varepsilon$ -symbol and the  $\varepsilon$ -axioms can simplify the formulation of a theory without enlarging its set of theorems in any essential way, and in particular, without introducing any inconsistency.

On the other hand, if the axioms of  $\mathcal{T}_\varepsilon$  include *all* the instances of some axiom schema, such as the axiom schema of replacement in set theory, then the Second  $\varepsilon$ -Theorem is not applicable and the theory  $\mathcal{T}_\varepsilon$  is not necessarily consistent relative to the consistency of  $\mathcal{T}$ . Nevertheless, we shall show in § 5.2 that in the case of set theory the relative consistency of  $\mathcal{T}_\varepsilon$  with respect to  $\mathcal{T}$  still holds. This provides a positive solution to the following problem, raised by Fraenkel and Bar-Hillel [1958], p. 185:

‘This relative consistency need no longer hold if the axioms of the theory do also contain  $\varepsilon$ -terms; indeed, the consistency of every set theory in which the axiom schema of comprehension (in any of its variants) is to hold also for conditions containing  $\varepsilon$ -terms relative to that in which this is not assumed, has not yet been proved.’

### 5.1 The predicate calculus with class operator and choice function

In § 4.2 we observed that in a set theory based on the  $\varepsilon$ -calculus the operator  $\hat{\ }^c$  can be defined in terms of the  $\varepsilon$ -symbol in such a way that the formula

$$(1) \quad \forall x (x \in \hat{x}A \leftrightarrow A)$$

is derivable from

$$\exists y \forall x (x \in y \leftrightarrow A).$$

It is often convenient, however, to take the symbol  $\wedge$  as a primitive and to adopt (1) as an axiom schema. In such a system the expressions of the form  $\hat{x}A$ , which we call class-terms, must be treated with some caution so as to avoid the paradoxes. One common way of doing this is to formulate the grammar of the language in such a way that class-terms may only appear on the right hand side of the membership symbol  $\in$  (cf. Bernays [1958] p. 48). Such a system can be further strengthened by taking as a primitive a choice function whose arguments are class terms. For example, Bernays [1958] takes the symbol  $\sigma$  as a primitive and adopts the following formulae as axioms:

$$\begin{aligned} A_{\sigma}' & a \in C \rightarrow \sigma(C) \in C, \\ A_{\sigma}'' & A \equiv B \rightarrow \sigma(A) = \sigma(B), \end{aligned}$$

where  $a$  is a free set variable, and  $A$ ,  $B$ , and  $C$  are free class variables. In Bernays' system of set theory, these two axioms provide a very strong form of the axiom of choice.

In this section we define a formal system,  $C(\mathcal{V})$ , which incorporates the class operator  $\wedge$  and the choice function  $\sigma$ , and investigate the relationship between this system and the  $\varepsilon$ -calculus.

We first define the *class-language*,  $\mathcal{L}_1(\mathcal{V})$ , which is determined by a given vocabulary  $\mathcal{V}$ . Let  $\mathcal{V}$  be any vocabulary. The set of symbols of  $\mathcal{L}_1(\mathcal{V})$  is the same as that of  $\mathcal{L}(\mathcal{V})$  except that we exclude the  $\varepsilon$ -symbol and include the symbols  $\in$ ,  $\wedge$ , and  $\sigma$  as logical symbols. The rules of formation for defining the quasi-terms and quasi-formula of the language are the same as for  $\mathcal{L}(\mathcal{V})$  (see page 11) except that we replace G8 (the rule for forming quasi  $\varepsilon$ -terms) by the following:

G8°. If  $t$  is a quasi-term,  $A$  a quasi-formula, and  $x$  any variable, then  $t \in \hat{x}A$  is a quasi-formula, and  $\sigma(\hat{x}A)$  is a quasi-term.

Any expression of the form  $\hat{x}A$  is called a *quasi class-term*. The quasi class-terms are not included among the quasi-terms of the language, and to avoid any confusion we shall refer to the quasi-terms of the language as *quasi set-terms*. As can be seen from the above rule a quasi class-term  $\hat{x}A$  can occur within a well-formed expression in only two possible contexts: (i)  $t \in \hat{x}A$  and (ii)  $\sigma(\hat{x}A)$ . As usual, we shall use the letters  $s$  and  $t$  to denote arbitrary quasi set-terms. The quasi class-terms will be denoted by the letters  $S$  and  $T$ . For any two quasi class-terms  $S$  and  $T$  we write

$$(S = T) \text{ for } \forall z (z \in S \leftrightarrow z \in T),$$

where  $z$  is some variable which does not occur free in  $S$  or  $T$ .

Free and bound occurrences of variables are defined in the usual way, where any occurrence of  $x$  in  $\hat{x}A$  is bound, and we define a *set-term*, *class-term*, and *formula* as a quasi set-term, quasi class-term, or quasi-formula, respectively, in which no variable occurs free.

We now define the formal system  $C(\mathcal{V})$ , called the *class calculus* for the vocabulary  $\mathcal{V}$ . The (*logical*) *axioms* of  $C(\mathcal{V})$  are all formulae of  $\mathcal{L}_1(\mathcal{V})$  which are instances of axiom schemata P1–P10, Q1–Q3, E1, E3 (see page 39) as well as the following three schemata:

$$\begin{aligned} \text{CS} & t \in \hat{x}A \leftrightarrow A_t^x, \\ \sigma 1 & \exists x (x \in S) \rightarrow \sigma(S) \in S, \\ \sigma 2 & S = T \rightarrow \sigma(S) = \sigma(T). \end{aligned}$$

The rules of inference of the class calculus are the same as for the predicate calculus, namely, the modus ponens rule and the  $\exists$ -rule. The deductions in  $C(\mathcal{V})$  are defined exactly as they were for the predicate calculus by first defining a derivation (see page 60). To denote that  $A$  is deducible from  $X$  in  $C(\mathcal{V})$  we write  $X \vdash_{C(\mathcal{V})} A$ , or just  $X \vdash_C A$ . The usual derived rules of inference can be established for  $C(\mathcal{V})$  as they were for  $PC(\mathcal{V})$ .

**THEOREM IV.5.** Any formulae of the following forms are theorems of the class calculus:

$$\begin{aligned} \text{(i)} & \exists x A \rightarrow A(\sigma(\hat{x}A)), \\ \text{(ii)} & \forall z (A_z^x \leftrightarrow B_z^y) \rightarrow \sigma(\hat{x}A) = \sigma(\hat{y}B). \end{aligned}$$

*Proof.* Throughout the proof we shall write  $\vdash$  instead of  $\vdash_C$ .

(i): Let  $a$  be some individual symbol not appearing in  $A$ .

$$\begin{array}{ll} \vdash a \in \hat{x}A \leftrightarrow A_a^x & \text{CS} \\ \vdash a \in \hat{x}A \rightarrow \exists x (x \in \hat{x}A) & \text{Q3 and P3} \\ \vdash \exists x (x \in \hat{x}A) \rightarrow \sigma(\hat{x}A) \in \hat{x}A & \sigma 1 \\ \vdash \sigma(\hat{x}A) \in \hat{x}A \leftrightarrow A(\sigma(\hat{x}A)) & \text{CS} \\ \vdash A_a^x \rightarrow A(\sigma(\hat{x}A)) & \text{tautology rule} \\ \vdash \exists x A \rightarrow A(\sigma(\hat{x}A)) & \exists\text{-rule.} \end{array}$$

(ii): Let  $a$  be some individual symbol not appearing in  $A$  or  $B$ . By axiom schemata Q1 and Q3,

$$\vdash \forall z (A_z^x \leftrightarrow B_z^y) \rightarrow (A_a^x \leftrightarrow B_a^y).$$

But by CS,

$$\begin{aligned} \vdash a \in \hat{x}A & \leftrightarrow A_a^x, \\ \vdash a \in \hat{y}B & \leftrightarrow B_a^y. \end{aligned}$$

Hence, by the tautology rule,

$$\vdash \forall z (A_z^x \leftrightarrow B_z^y) \rightarrow (a \in \hat{x}A \leftrightarrow a \in \hat{y}B).$$

By the  $\forall$ -rule and the definition of  $=$ ,

$$\vdash \forall z(A_z^* \leftrightarrow B_z^*) \rightarrow \hat{x}A = \hat{y}B.$$

Consequently, by  $\sigma 2$ ,

$$\vdash z(A_z^* \leftrightarrow B_z^*) \rightarrow \sigma(\hat{x}A) = \sigma(\hat{y}B).$$

The above theorem indicates that there is a close connection between the system  $C(\mathcal{V})$  and the  $\varepsilon$ -calculus. Of course, this relationship is not unexpected in view of our semantic interpretation of the  $\varepsilon$ -symbol as a choice function. In order to investigate this relationship more rigorously we define a transform operation which translates a formula of  $\mathcal{L}(\mathcal{V})$  into a formula  $\mathcal{L}_1(\mathcal{V})$  and another transform operation which translates a formula  $\mathcal{L}_1(\mathcal{V})$  into a formula of  $\mathcal{L}(\mathcal{V})$ .

For any formula  $B$  of  $\mathcal{L}(\mathcal{V})$  the  $C$ -transform of  $B$  is that formula of  $\mathcal{L}_1(\mathcal{V})$  which is obtained from  $B$  by replacing each quasi  $\varepsilon$ -term  $\varepsilon xA$  in  $B$  by  $\sigma(\hat{x}A)$ . Conversely, for any formula  $B$  of  $\mathcal{L}_1(\mathcal{V})$  the  $\varepsilon$ -transform of  $B$  is that formula of  $\mathcal{L}(\mathcal{V})$  which is obtained from  $B$  by eliminating the quasi class-terms in  $B$  as follows: (i) replace each expression of the form  $(t \in \hat{x}A)$  by  $A_x^*$  (if  $t$  is not free for  $x$  in  $A$ , relabel the bound variables in  $A$  so that it is free for  $x$ ); (ii) replace each expression of the form  $\sigma(\hat{x}A)$  by  $\varepsilon xA$ .

**THEOREM IV.6.** *The  $\varepsilon$ -transform of an instance of CS,  $\sigma 1$ , or  $\sigma 2$  is a theorem of the  $\varepsilon$ -calculus.*

*Proof.* (i): Clearly, the  $\varepsilon$ -transform of an instance of CS is a theorem of the  $\varepsilon$ -calculus since it has the form

$$\forall y(A \leftrightarrow A).$$

(ii): An instance of  $\sigma 1$  has the form

$$(1) \quad \exists x(x \in \hat{y}B) \rightarrow \sigma(\hat{y}B) \in \hat{y}B.$$

The  $\varepsilon$ -transform of (1) has the form

$$(2) \quad \exists x[B'_x]^y \rightarrow [B'_y]^y,$$

which by the Q4 axiom schema is clearly a theorem of the  $\varepsilon$ -calculus.

(iii): An instance of  $\sigma 2$  has the form

$$\forall z(z \in \hat{x}A \leftrightarrow z \in \hat{y}B) \rightarrow \sigma(\hat{x}A) = \sigma(\hat{y}B).$$

Clearly, by axiom schema E2, the  $\varepsilon$ -transform of this is a theorem of the  $\varepsilon$ -calculus.

**THEOREM IV.7.** (i) *If  $B$  is a theorem of the class calculus, then the  $\varepsilon$ -transform of  $B$  is a theorem of the  $\varepsilon$ -calculus.*

(ii) *If  $B$  is a theorem of the  $\varepsilon$ -calculus, then the  $C$ -transform of  $B$  is a theorem of the class calculus.*

*Proof.* (i): The proof follows immediately by induction on the length of the proof of  $B$  in  $C(\mathcal{V})$  since by Theorem IV.6 the  $\varepsilon$ -transform of every axiom of  $C(\mathcal{V})$  is a theorem of the  $\varepsilon$ -calculus, modus ponens is a rule of the  $\varepsilon$ -calculus, and the  $\exists$ -rule is a derived rule in the  $\varepsilon$ -calculus.

(ii): Similarly, the proof follows immediately by induction on the length of the proof of  $B$ , since the  $C$ -transform of every axiom of the  $\varepsilon$ -calculus is a theorem of  $C(\mathcal{V})$  by Theorem IV.5, and modus ponens is a rule of inference of  $C(\mathcal{V})$ .

**THEOREM IV.8** (Eliminability of  $\wedge$  and  $\sigma$ ). *Suppose  $X$  is a set of formulae of  $\mathcal{L}_1$  and  $A$  is a formula of  $\mathcal{L}_1$ , such that the symbols  $\wedge$  and  $\sigma$  do not appear in  $A$  or in any member of  $X$ . If  $X \vdash_C A$ , then  $X \vdash_{PC} A$ .*

*Proof.* Since  $X \vdash_C A$ , then there exist formulae  $B_1, \dots, B_n$  of  $X$  such that  $B_1 \rightarrow \dots \rightarrow B_n \rightarrow A$  is a theorem of  $C(\mathcal{V})$ . We shall denote this formula by  $B$ . Since  $B$  does not contain  $\wedge$  or  $\sigma$ , then the  $\varepsilon$ -transform of  $B$  is  $B$ , and therefore by Theorem IV.7  $B$  is a theorem of the  $\varepsilon$ -calculus. Finally by the Second  $\varepsilon$ -Theorem  $B$  is a theorem of the predicate calculus. Hence  $X \vdash_{PC} A$ .

The class calculus can be used as the logical basis for formalizing set theory. In this case the rules of formation of  $\mathcal{L}$  are extended to include quasi formulae of the form  $s \in t$ . (Alternatively, two distinct symbols  $\in$  and  $\eta$  can be used for set membership and class membership, respectively.) Although the symbol  $\sigma$  can have only quasi class-terms as its arguments, we can extend the use of this symbol in practice by writing

$$\sigma(t) \text{ for } \sigma(\hat{x}(x \in t)),$$

for any quasi set-term  $t$ , where  $x$  does not occur free in  $t$ . Using Theorem IV.5, it follows that for any set-terms  $s$  and  $t$ :

$$\begin{aligned} &\vdash \exists x(x \in t) \rightarrow \sigma(t) \in t, \\ &\vdash s = t \rightarrow \sigma(s) = \sigma(t). \end{aligned}$$

Although the symbol  $\sigma$  provides the system with a 'universal choice function', in view of Theorem IV.8 it does not necessarily follow that the axiom of choice is a theorem of set theory based on the class calculus. Clearly, the relationship between the symbol  $\sigma$  and the axiom of choice is similar to the relationship between the  $\varepsilon$ -symbol and the axiom of choice (cf. §4.4). If the axioms of the theory include those instances of the axiom schema of replacement which contain  $\sigma$ , then the axiom of choice is derivable. For, the formula

$$(1) \quad \forall x \exists y \forall z (z \in y \leftrightarrow \exists w (w \in x \wedge z = \sigma(w)))$$

together with  $\sigma_1$  and  $\sigma_2$  does yield the axiom of choice, and (1) is an immediate consequence of the unlimited axiom schema of replacement. In the set theory of Bernays [1958], the axioms  $A_{\sigma}'$  and  $A_{\sigma}''$  do yield the axiom of choice because the instances of his axiom schema A3 include formulae which contain the symbol  $\sigma$ .

**5.2 The relative consistency of set theory based on the  $\varepsilon$ -calculus**

We now turn to the problem raised at the end of § 4.4—namely, whether a set theory based on the  $\varepsilon$ -calculus with an unrestricted axiom schema of replacement is consistent relative to the consistency of the same theory based on the predicate calculus.

Let ZF be a set theory, based on the predicate calculus using set-variables only (i.e., no class-terms or class-variables), whose axioms consist of the axioms of extensionality, pairing, union-set, power-set, infinity, foundation, and the axiom schema of replacement. Let  $ZF_{\varepsilon}$  be the corresponding theory based on the  $\varepsilon$ -calculus, where an instance of the axiom schema of replacement may contain the  $\varepsilon$ -symbol. We wish to prove that if ZF is consistent, then  $ZF_{\varepsilon}$  is also consistent.

Let  $B_{\sigma}'$  be the system of Bernays [1958] with axioms  $A_{\sigma}'$  and  $A_{\sigma}''$ . Let  $B_{\sigma}$  be the system of Bernays where the axioms  $A_{\sigma}'$  and  $A_{\sigma}''$  are replaced by

$$A_{\sigma} \quad a \neq 0 \rightarrow \sigma(a) \in a.$$

(In  $B_{\sigma}$  the arguments of the primitive symbol  $\sigma$  are set-terms, and in  $B_{\sigma}'$  the arguments are class-terms.) Bernays has shown (pp. 200–207) that under a suitable definition of  $\sigma(A)$ , axioms  $A_{\sigma}'$  and  $A_{\sigma}''$  are derivable from  $A_{\sigma}$  and the axiom of foundation.

Let  $G'$  be the set theory of Gödel [1940] with axioms A, B, C, D, and E, and let  $G$  be the same theory without Axiom E. Axiom E, the axiom of choice, is the formula

$$\exists A(\text{Un}(A) \wedge \forall x(\neg \text{Em}(x) \rightarrow \exists y(y \in x \wedge \langle yx \rangle \in A))),$$

where

$$\begin{aligned} \text{Un}(A) &=_{\text{Df}} \forall u \forall v \forall w ((\langle vu \rangle \in A \wedge \langle wu \rangle \in A) \rightarrow v = w), \\ \text{Em}(x) &=_{\text{Df}} \forall u \neg (u \in x), \end{aligned}$$

and  $A$  is a class-variable.

Suppose that  $B$  is an  $\varepsilon$ -free theorem of  $ZF_{\varepsilon}$ . It can be shown that the C-transform of any axiom of  $ZF_{\varepsilon}$  is a theorem of  $B_{\sigma}'$ . Hence by Theorem IV.7(ii),  $B$  is a theorem of  $B_{\sigma}'$ . Consequently, using Bernays' definition of  $\sigma$ ,  $B$  is a theorem of  $B_{\sigma}$ . It can then be shown that  $B$  is a theorem of  $G'$ . Consequently, if  $G'$  is consistent, then  $ZF_{\varepsilon}$  is also consistent. But by Gödel's

proof [1940] of the relative consistency of the axiom of choice,  $G'$  is consistent if  $G$  is. Finally, using the method of Shoenfield [1954],  $G$  is consistent relative to the consistency of ZF. This completes the argument.

One further question remains. Is every  $\varepsilon$ -free theorem of  $ZF_{\varepsilon}$  also a theorem of ZF with the axiom of choice? Or, in other words, are the  $\varepsilon$ -axioms and the unrestricted axiom schema of replacement equivalent to the axiom of choice and the axiom schema of replacement? The answer to this problem seems to depend on whether the universal or the local version of the axiom of choice is used. These two versions of the axiom are defined by Lévy [1961] as follows. The local version of the axiom of choice is the formula

$$(1) \quad \forall x \exists / \forall y (y \in x \rightarrow y = 0 \vee / y \in y).$$

The universal axiom of choice is obtained by taking the formula

$$(2) \quad \forall x (x \neq 0 \rightarrow \sigma(x) \in x)$$

as an axiom and allowing the instances of the axiom schema of replacement to include the primitive symbol  $\sigma$ . Let  $ZF'$  be the theory obtained from ZF by adjoining the local version of the axiom of choice, and let  $ZF_{\sigma}$  be the theory obtained from ZF by adjoining the universal version. The above results can be used to show that the set of  $\varepsilon$ -free theorems of  $ZF_{\varepsilon}$  coincides with the set of  $\sigma$ -free theorems of  $ZF_{\sigma}$ . However, it is not known whether this set coincides with the set of theorems of  $ZF'$ . For a partial solution of this problem, see Lévy [1961].

THE CUT ELIMINATION THEOREM

1 The sequent calculus

In this chapter we define a new formal system called the sequent calculus. This system resembles Gentzen's system LK [1934-5] except that it incorporates the  $\varepsilon$ -symbol and the identity symbol. The main theorem of this chapter, the Cut Elimination Theorem can be regarded as an analogue of Gentzen's *Hauptsatz*. We shall use this theorem to provide new proofs of the  $\varepsilon$ -Theorems (except that in the case of the Second  $\varepsilon$ -Theorem we consider an  $\varepsilon$ -calculus in which the E2-axioms are not used). Although we are in effect only giving alternative proofs of some of the theorems of Chapter III, this new approach sheds some light on the relationship between Gentzen's *Hauptsatz* and Hilbert's  $\varepsilon$ -Theorems.

Throughout this chapter  $\mathcal{V}$  denotes some arbitrary vocabulary, and  $\mathcal{L}$  denotes the language determined by  $\mathcal{V}$ .

In the sequent calculus, the rules of inference apply not to formulae, but rather to more complicated formal objects called 'sequents'. A *sequent* of the language  $\mathcal{L}$  is an expression of the form  $A_1, \dots, A_n$  where the  $A_i$  are formulae of  $\mathcal{L}$ ,  $n \geq 0$ , and the symbol ',' (comma) is a formal separation symbol of the language  $\mathcal{L}$  (page 10). A formula *belongs to*, or *is a member of*, a sequent  $A_1, \dots, A_n$  if it is one of the  $A_i$ . A sequent is  $\varepsilon$ -free if every formula belonging to that sequent is  $\varepsilon$ -free. The capital Greek letters  $\Gamma, \Delta, \Theta$ , and  $\Lambda$  are used as syntactic variables for sequents. If  $\Gamma$  and  $\Delta$  denote two non-empty sequents, then clearly  $\Gamma, \Delta$  denotes a sequent. By an abuse of notation we shall also write  $\Gamma, \Delta$  when either  $\Gamma$  or  $\Delta$  is empty. If  $\Gamma$  is empty, then  $\Gamma, \Delta$  denotes the sequent  $\Delta$ , and vice versa when  $\Delta$  is empty. Notice that a formula is a sequent. This situation should not lead to any confusion, however. For any sequent  $\Delta$ , we write  $\Delta^*$  to denote the set of formulae which belong to  $\Delta$ . Thus if  $\Delta$  is  $A, B, A, A$ , then  $\Delta^* = \{A, B\}$ .

A non-empty sequent  $A_1, \dots, A_n$  has the same semantic interpretation as the formula  $A_1 \wedge \dots \wedge A_n$ . The empty sequent is given the truth value 1 (true). We shall not, however, be concerned with the semantics of sequents except as an intuitive guide to our understanding of the axioms and rules of inference of the system.

The sequent calculus deals with refutations rather than deductions. Intuitively, we can regard a refutation of  $\Delta$  as a formal demonstration of the

inconsistency of the set  $\Delta^*$ . Each axiom is an invalid sequent, and each rule of inference leads from invalid sequents to invalid sequents. When we have defined the notion of a refutation, we can then define a 'deduction' of  $A$  from  $X$  as a refutation of some sequent  $\neg A, \Delta$  where  $\Delta^* \subseteq X$ . In particular a 'proof' of  $A$  is a refutation of the sequent  $\neg A$ .

2 The axioms and rules of inference of the sequent calculus

A sequent  $\Delta$  is an *axiom* of the sequent calculus if either

- (i) the formula  $f$  belongs to  $\Delta$ , or
- (ii) for some atom  $A$ , both  $\neg A$  and  $A$  belong to  $\Delta$ , or
- (iii) a formula of the form  $\neg(t = t)$  belongs to  $\Delta$ .

If  $\Delta$  satisfies either (i) or (ii) above, it is called a *C-axiom*.

In defining the rules of inference we use the customary schematic notation. Thus to denote that  $\Gamma$  'follows' by some rule of inference from  $\Delta$  we write

$$\frac{\Delta}{\Gamma}$$

and to denote that  $\Gamma$  'follows' by some rule of inference from  $\Delta_1$  and  $\Delta_2$  we write

$$\frac{\Delta_1 \quad \Delta_2}{\Gamma}$$

The sequent written below the line is called the *conclusion* of that rule of inference, and the formula(e) written above the line the *premiss(es)*.

The *rules of inference* are as follows:

$$\begin{array}{l} \neg\neg\text{-rule} \quad \frac{A, \Gamma}{\neg\neg A, \Gamma} \\ \alpha_1\text{-rule} \quad \frac{\alpha_1, \Gamma}{\alpha, \Gamma} \\ \alpha_2\text{-rule} \quad \frac{\alpha_2, \Gamma}{\alpha, \Gamma} \\ \beta\text{-rule} \quad \frac{\beta_1, \Gamma \quad \beta_2, \Gamma}{\beta, \Gamma} \\ \gamma\text{-rule} \quad \frac{\gamma(t), \Gamma}{\gamma, \Gamma} \\ \delta\text{-rule} \quad \frac{\delta(\varepsilon\delta), \Gamma}{\delta, \Gamma} \\ \text{E1-rule} \quad \frac{\neg(s = t), \Gamma \quad \neg A_s^*, \Gamma}{* \Gamma} \end{array}$$

$$\text{cut-rule } \frac{A, \Gamma \quad \neg A, \Gamma}{\Gamma}$$

$$\text{structure rule } \frac{\Gamma}{\Delta}, \text{ where } \Gamma^* \subseteq \Delta^*$$

*Restriction.* In the E1-rule,  $A$  is an atom and  $x$  any variable not having a free occurrence in  $A$  within the scope of an  $\varepsilon$ -symbol.

Notice that the  $\beta$ -rule, E1-rule, and cut rule all have two premisses and the remaining rules have only one. The structure rule is a very useful one since it permits us to add new members to a sequent, to rearrange the order of the members of a sequent, and to remove repetitions. Thus the sequent  $A, B, C$  follows from the sequent  $C, C, B$  by the structure rule. In each of the above rules the specified formula in the conclusion is called the *major formula*, and the specified formula(e) in the premiss(es) the *minor formula(e)*. Thus in the application of the  $\beta$ -rule by which  $A \rightarrow B, C, D$  follows from  $\neg A, C, D$  and  $B, C, D$ , the formula  $A \rightarrow B$  is the major formula, and the formulae  $\neg A$  and  $B$  are the minor formulae. An application of the structure rule has no major or minor formulae, and an application of the cut rule has no major formula. In any application of the cut rule the minor formula which was denoted above by  $A$  is called the *cut formula*.

### EXERCISE

Let  $\Gamma, \Delta_1$ , and  $\Delta_2$  be any three sequents. Suppose that  $\Gamma$  follows by one of the above rules of inference from  $\Delta_1$  (or from  $\Delta_1$  and  $\Delta_2$ ). Prove that if  $\Delta_1^* \vdash_\varepsilon f$  (and  $\Delta_2^* \vdash_\varepsilon f$ ), then  $\Gamma^* \vdash_\varepsilon f$ . (cf. Theorem II.12.)

### 2.1 Refutations

A *refutation* of the sequent  $\Delta$  in the sequent calculus is any sequence of sequents such that the last member of the sequence is  $\Delta$  and each member of the sequence either is an axiom or follows by some rule of inference from a preceding member (or from preceding members) of the sequence. A refutation in which the cut rule is not used is called a *normal refutation*. If there exists a refutation of  $\Delta$ , we say  $\Delta$  is *refutable* and denote this fact by writing  $\text{Ref}(\Delta)$ . Furthermore, if there exists a normal refutation of  $\Delta$ , we write *norm-Ref*( $\Delta$ ).

In order to show that the sequent calculus fits our general definition of a formal system, we define the notion of a deduction of  $A$  from  $X$  as follows: For any formula  $A$  (of  $\mathcal{L}(\mathcal{V})$ ) and any set  $X$  of formulae (of  $\mathcal{L}(\mathcal{V})$ ) the *deductions* of  $A$  from  $X$  in the sequent calculus (for  $\mathcal{L}(\mathcal{V})$ ) are the refutations of sequents of the form  $\neg A, \Delta$ , where  $\Delta^* \subseteq X$ . Thus a *proof* of  $A$  is a refutation of the sequent  $\neg A$ .

By defining a refutation as a sequence of sequents we encounter a nota-

tional ambiguity, since the expression  $\langle \Delta_1, \dots, \Delta_n \rangle$  could mean the sequence whose members are the sequents  $\Delta_1$ , and  $\Delta_2$ , etc., or it could mean the sequence whose only member is the sequent  $\Delta_1, \dots, \Delta_n$ . To avoid such ambiguity we shall denote a sequence of sequents by inserting a semi-colon between each member of the sequence. Thus the expression

$$\langle \Gamma_1; \dots; \Gamma_m; \Delta, \Gamma \rangle$$

denotes a sequence of  $m + 1$  sequents where for each  $i = 1, \dots, m$ ,  $\Gamma_i$  is the  $i$ th member, and the sequent  $\Delta, \Gamma$  is the last member.

Having defined the notions of a refutation and a normal refutation, we can now give the precise statement of the major theorem of this chapter.

**THE CUT ELIMINATION THEOREM.** *For any  $\varepsilon$ -free sequent  $\Delta$ , if  $\text{Ref}(\Delta)$ , then  $\text{norm-Ref}(\Delta)$ .*

In the next section we shall see why the notion of a normal refutation is so important.

### 3 The subformula property of normal refutations

The notion of an *immediate logical subformula* is defined by the following rules.

1.  $A$  is an immediate logical subformula of  $\neg \neg A$ .
2. For  $i = 1, 2$ ,  $\alpha_i$  is an immediate logical subformula of  $\alpha$ , and  $\beta_i$  is an immediate logical subformula of  $\beta$ .
3. For any term  $t$ ,  $\gamma(t)$  is an immediate logical subformula of  $\gamma$ , and  $\delta(t)$  is an immediate logical subformula of  $\delta$ .
4. For any formulae  $A$  and  $B$ ,  $A$  is an immediate logical subformula of  $B$  only if it is so by virtue of one of the above rules.

A formula  $A$  is said to be a *logical subformula* of  $B$  if and only if there exists a finite sequence of formulae whose first member is  $B$  and last member is  $A$ , such that every member of the sequence, except the first, is an immediate logical subformula of the preceding member. (Since the sequence may have only one member, then any formula is a logical subformula of itself.)

Now consider the rules of inference of the sequent calculus. Except in the case of the cut rule and the E1-rule, any formula which belongs to the premiss of a rule of inference is a logical subformula of some formula which belongs to the conclusion. In the case of the E1-rule every formula which belongs to the premiss is either the negation of an atom or a logical subformula of some formula which belongs to the conclusion. These facts yield the following *subformula property* of normal refutations:

If  $\mathcal{R}$  is a normal refutation of  $\Delta$ , then every formula in  $\mathcal{R}$  (i.e., every formula which belongs to some sequent of  $\mathcal{R}$ ) is either (i) the negation of an atom, or (ii) a logical subformula of some formula belonging to  $\Delta$ .

Thus by proving the Cut Elimination Theorem, we can show that for every refutable  $\varepsilon$ -free  $\Delta$ , there exists a refutation whose formulae are 'no more complicated' than those which belong to  $\Delta$ .

### 3.1 The eliminability of the identity symbol

In order to demonstrate the usefulness of the subformula property we shall prove the following theorem concerning the eliminability of the identity-symbol from normal refutations.

A sequent is said to be *identity-free* if all its members are identity-free.

**THEOREM V.1.** *For any identity-free sequent  $\Delta$ , if  $\text{norm-Ref } \Delta$ , then there exists a normal refutation of  $\Delta$  in which every sequent is identity-free.*

*Proof.* To prove the theorem it is sufficient to prove that there exists a normal refutation  $\langle \Delta_1; \dots; \Delta_n \rangle$  of  $\Delta$ , where each  $\Delta_i$  either is a C-axiom (see page 115) or follows by some rule other than the EI-rule. For in such a refutation, the identity symbol plays no essential role, and hence every occurrence of a quasi-formula of the form  $s = t$  can be replaced by the formula  $f$ .

The proof hinges on the fact that if  $A$  is a subformula of some identity-free formula  $B$ , then its skeleton  $A^+$  (p.66) is identity-free. Let  $\langle \Gamma_1; \dots; \Gamma_n \rangle$  be any normal refutation of  $\Delta$ . By the subformula property of normal refutations, for any formula  $A$  in the  $\Gamma_i$ , if the skeleton of  $A$  is not identity free, then  $A$  must be the negation of an atom and hence of the form  $\neg(s = t)$ . Now for each  $i = 1, \dots, n$  let  $\Delta_i$  be obtained from  $\Gamma_i$  by replacing every member of  $\Gamma_i$  of the form  $\neg(t = t)$  by  $f$ , and by removing every member of  $\Gamma_i$  of the form  $\neg(s = t)$ , where  $s$  and  $t$  are different. Then the sequence  $\langle \Delta_1; \dots; \Delta_n \rangle$  is the required refutation of  $\Delta$ . For, if  $\Gamma_i$  is an axiom, then  $\Delta_i$  is a C-axiom; if  $\Gamma_i$  follows from  $\Gamma_j$  and  $\Gamma_l$  by the EI-rule, then  $\Delta_i$  follows by the structure rule from either  $\Delta_j$  or  $\Delta_l$ ; and if  $\Gamma_i$  follows by any other rule of inference, then  $\Delta_i$  follows by that same rule.

This theorem will be used later, in conjunction with the Cut Elimination Theorem, to give an alternative proof of the eliminability of the identity symbol from the predicate calculus.

### 4 The contradiction rule

In this section we shall prove the useful fact that if for some formula  $A$ , both  $A$  and  $\neg A$  occur in  $\Delta$ , then  $\text{norm-Ref}(\Delta)$ . The proof of this fact depends on the following notion of the index of a formula or term.

By the skeleton  $A^+$  of any formula or term  $A$  we understand that expression which is obtained from  $A$  by replacing every subterm of  $A$  by  $a_1$  (cf. p. 66). We now define the *index* of any formula or term to be the length of its skeleton, and we denote the index of  $A$  by  $\text{ind}(A)$ . Clearly the index function possesses the following properties:

- K1.  $\text{ind}(A) < \text{ind}(\neg A)$ ;
- K2. If  $A$  is an immediate logical subformula of  $B$ , then  $\text{ind}(A) < \text{ind}(B)$ ;
- K3. if  $A'$  is obtained from  $A$  by replacing every occurrence of some term  $t$  by  $s$ , then  $\text{ind}(A') = \text{ind}(A)$ ;
- K4. if  $\varepsilon yB$  is an  $\varepsilon$ -term of the form  $[p]^x$  and  $p$  is subordinate to  $QxA$ , then  $\text{ind}(\varepsilon yB) + 2 \leq \text{ind}(QxA)$ .

Thus the index function, like the rank function in Chapter III, is a useful measure of the complexity of any term or formula.

**THEOREM V.2** (The contradiction rule). *For any sequent  $\Delta$  and any formula  $A$ , if  $A$  and  $\neg A$  belong to  $\Delta$ , then  $\text{norm-Ref}(\Delta)$ .*

*Proof.* The proof is by induction on the index of the formula  $\neg A$ . At each stage of the induction it is sufficient to prove either  $\text{norm-Ref}(\neg A, A)$  or  $\text{norm-Ref}(A, \neg A)$ , since  $\Delta$  follows from each of the sequents  $A, \neg A$  and  $\neg A, A$  by the structure rule. By our unifying classification of formulae (page 14) one of the following four cases must hold.

*Case 1.  $A$  is an atom:* In this case  $\Delta$  is an axiom and therefore  $\text{norm-Ref}(\Delta)$ .

*Case 2.  $A$  is of the form  $\neg B$ :* Since  $\text{ind}(\neg B) < \text{ind}(\neg A)$ , we have  $\text{norm-Ref}(B, \neg B)$  by the induction hypothesis. Hence  $\text{norm-Ref}(\neg \neg B, \neg B)$  by the  $\neg \neg$ -rule.

*Case 3. One of the two formulae,  $A$  and  $\neg A$ , is a conjunctive formula  $\alpha$  and the other is a disjunctive formula  $\beta$ :* By the duality principle (page 16)  $\alpha_1$  and  $\beta_1$  are contradictory and  $\alpha_2$  and  $\beta_2$  are contradictory. Hence by the induction hypothesis we have

$$\text{norm-Ref}(\beta_1, \alpha_1, \alpha_2) \quad \text{and} \\ \text{norm-Ref}(\beta_2, \alpha_1, \alpha_2).$$

Hence  $\text{norm-Ref}(\beta, \alpha_1, \alpha_2)$  by the  $\beta$ -rule. The desired result,  $\text{norm-Ref}(\alpha, \beta)$ , now follows by applications of the structure rule, the  $\alpha_1$ -rule, and the  $\alpha_2$ -rule.

*Case 4. One of the two formulae,  $A$  and  $\neg A$ , is a universal formula  $\gamma$  and the other is an existential formula  $\delta$ :* By the duality principle  $\gamma(\varepsilon\delta)$  and  $\delta(\varepsilon\delta)$  are contradictory. Hence by the induction hypothesis we have  $\text{norm-Ref}(\gamma(\varepsilon\delta), \delta(\varepsilon\delta))$ . The desired result,  $\text{norm-Ref}(\delta, \gamma)$  now follows by applications of the  $\gamma$ -rule, structure rule, and  $\delta$ -rule.

**THEOREM V.3.** *For any sequent  $\Delta$ , if  $\Delta^*$  is truth functionally invalid, then  $\text{norm-Ref}(\Delta)$ .*



*Proof.* The proof is practically identical to the proof of Theorem II.13.

**COROLLARY (The Tautology Theorem).** *If  $A$  is a tautology, then  $\text{norm-Ref}(\neg A)$ .*

### EXERCISE

Prove that if  $A$  is any axiom of the  $\varepsilon$ -calculus other than an E2-axiom, then  $\text{norm-Ref}(\neg A)$ .

#### 4.1 The relationship between the sequent calculus and $\varepsilon$ -calculus

**THEOREM V.4.** *For any sequent  $\Delta$  and any formula  $A$ ,  $\text{Ref}(\neg A, \Delta)$  iff there exists a deduction of  $A$  from  $\Delta^*$  in the  $\varepsilon$ -calculus in which no E2-axioms are used.*

*Proof.* (i) Let  $\langle \Gamma_1; \dots; \Gamma_m \rangle$  be some refutation of  $\neg A, \Delta$ . It is easy to prove by induction that for each  $i = 1, \dots, m$  there is a deduction of  $\perp$  from  $\Gamma_i^*$  in the  $\varepsilon$ -calculus without the E2-axioms. (See the exercise at the end of § 2, page 116.) Hence  $\{A\} \cup \Delta^* \vdash_\varepsilon \perp$  (without E2), and therefore  $\Delta^* \vdash_\varepsilon A$  (without E2) by the  $\perp$ -rule.

(ii) Conversely, assume that  $\langle A_1, \dots, A_n \rangle$  is a deduction of  $A$  from  $\Delta^*$  in the  $\varepsilon$ -calculus (without E2). We shall prove by induction that for each  $i = 1, \dots, n$ ,  $\text{Ref}(\neg A_i, \Delta)$ . One of the following three cases must hold.

*Case 1.*  $A_i$  is an axiom (of the  $\varepsilon$ -calculus) other than an E2-axiom: Then  $\text{Ref}(\neg A_i)$  by the exercise at the end of the last section. Hence  $\text{Ref}(\neg A_i, \Delta)$  by the structure rule.

*Case 2.*  $A_i$  is a member of  $\Delta^*$ : Then  $\text{Ref}(\neg A_i, \Delta)$  by the contradiction rule.

*Case 3.*  $A_i$  follows from  $A_j$  and  $A_k$  by *modus ponens*: Then  $A_k$  is of the form  $A_j \rightarrow A_i$ . By the induction hypothesis we have

$$(1) \quad \text{Ref}(\neg A_j, \Delta) \quad \text{and}$$

$$(2) \quad \text{Ref}(\neg(A_j \rightarrow A_i), \Delta).$$

The structure rule applied to (1) gives

$$(3) \quad \text{Ref}(\neg A_j, \neg A_i, \Delta),$$

and by the contradiction rule, we have

$$(4) \quad \text{Ref}(A_i, \neg A_i, \Delta).$$

The  $\beta$ -rule applied to (3) and (4) gives

$$(5) \quad \text{Ref}(A_j \rightarrow A_i, \neg A_i, \Delta),$$

and the structure rule applied to (2) gives

$$(6) \quad \text{Ref}(\neg(A_j \rightarrow A_i), \neg A_i, \Delta).$$

Finally, the cut rule applied to (5) and (6) gives

$$(7) \quad \text{Ref}(\neg A_i, \Delta).$$

**THEOREM V.5.** *Let  $A$  be any proper formula and  $\Delta$  any sequent whose members are proper formulae. If  $\text{norm-Ref}(\neg A, \Delta)$ , then  $\Delta^* \vdash_\varepsilon A$ .*

*Proof.* Let  $\langle \Gamma_1; \dots; \Gamma_m \rangle$  be any normal refutation of  $\neg A, \Delta$ . Since every subformula of a proper formula is proper (and atoms are proper), then by the subformula property of normal refutations every member of  $\Gamma_i$  is proper, for each  $i = 1, \dots, m$ . Hence, as in the proof of Theorem V.4(i), one can prove that  $\Gamma_i^* \vdash_\varepsilon \perp$  for each  $i = 1, \dots, m$ .

#### 5 $k, F_k$ -refutations

We now turn to the main problem, that of proving the eliminability of the cut rule. In order to prove this result by an inductive argument we need the following definition.

For any  $k \geq 1$  and any finite collection  $F_k$  of formulae with index  $k$ , a  $k, F_k$ -refutation is a refutation  $\langle \Gamma_1; \dots; \Gamma_n \rangle$  such that a sequent  $\Gamma_i$  can follow by the cut rule from some  $\Gamma_j$  and  $\Gamma_l$  only if the cut formula  $A$  in that rule satisfies the following conditions: (i)  $\text{ind}(A) \leq k$  and (ii) if  $\text{ind}(A) = k$ , then  $A$  is a member of  $F_k$ . If there exists a  $k, F_k$ -refutation of  $\Delta$ , we write  $k, F_k\text{-Ref}(\Delta)$ . Obviously, if there exists a refutation of  $\Delta$ , then there exists a  $k \geq 1$  and a finite collection  $F_k$  of formula of index  $k$  such that  $k, F_k\text{-Ref}(\Delta)$ . Since every formula has index  $\geq 1$ , then a  $1, \emptyset$ -refutation is a normal refutation. Consequently, in order to prove the Cut Elimination Theorem it is sufficient to prove that if  $k, \{A\} \cup F_k\text{-Ref}(\Delta)$ , for some  $\varepsilon$ -free  $\Delta$ , then  $k, F_k\text{-Ref}(\Delta)$ .

Since our definition of a  $k, F_k$ -refutation imposes a limit on the index of any formula which is the minor formula in an application of the cut rule, we have the following modified subformula property for  $k, F_k$ -refutations:

If  $\mathcal{R}$  is a  $k, F_k$ -refutation of  $\Delta$ , then for any formula  $A$  which belongs to some sequent of  $\mathcal{R}$ , either (i)  $A$  is a subformula of some member of  $\Delta$ , or (ii)  $A$  is the negation of an atom, or (iii)  $\text{ind}(A) \leq k + 1$ .

#### 5.1 The invertibility of the logical rules

Our next theorem states that the  $\neg\neg$ -rule,  $\alpha$ -rules,  $\beta$ -rule,  $\gamma$ -rule, and  $\delta$ -rule are, in a certain sense, invertible. The proof of this theorem depends on the fact that a non-atomic formula  $A$  can be the major formula in one and only one type of rule. For example, if  $A$  is a conjunctive formula, it can be the major formula only in the  $\alpha$ -rules, and if  $A$  is disjunctive, only in the  $\beta$ -rule.

**THEOREM V.6.** *For any sequent  $\Delta$ , and any formula  $A$ , conjunctive formula  $\alpha$ , disjunctive formula  $\beta$ , universal formula  $\gamma$ , and existential formula  $\delta$ :*

- (i)  $k, F_k\text{-Ref}(\neg\neg A, \Delta)$  iff  $k, F_k\text{-Ref}(A, \Delta)$ ;
- (ii)  $k, F_k\text{-Ref}(\alpha, \Delta)$  iff  $k, F_k\text{-Ref}(\alpha_1, \alpha_2, \Delta)$ ;
- (iii)  $k, F_k\text{-Ref}(\beta, \Delta)$  iff  $k, F_k\text{-Ref}(\beta_1, \Delta)$  and  $k, F_k\text{-Ref}(\beta_2, \Delta)$ ;
- (iv)  $k, F_k\text{-Ref}(\gamma, \Delta)$  iff there exists terms  $t_1, \dots, t_n$  such that  

$$k, F_k\text{-Ref}(\gamma(t_1), \dots, \gamma(t_n), \Delta)$$
;
- (v)  $k, F_k\text{-Ref}(\delta, \Delta)$  iff  $k, F_k\text{-Ref}(\delta(\varepsilon\delta), \Delta)$ .

*Proof.* We shall prove only part (iv), since the other four parts can be proved in a similar fashion.

(iv) First assume  $k, F_k\text{-Ref}(\gamma(t_1), \dots, \gamma(t_n), \Delta)$ . Then by repeated applications of the  $\gamma$ -rule and the structure rule we obtain a  $k, F_k$ -refutation of  $\gamma, \Delta$ .

Conversely, let  $\langle \Gamma_1; \dots; \Gamma_m \rangle$  be a  $k, F_k$  refutation of  $\gamma, \Delta$ . For notational simplicity suppose that  $\gamma$  is the formula  $\forall xA$ . (The proof is identical for the case where  $\gamma$  is of the form  $\neg\exists xA$ .) Let  $t_1, \dots, t_n$  be all those terms  $t$  such that, for some  $j = 1, \dots, m$ ,  $\Gamma_j$  is of the form  $A(t), \Lambda$ . (If there are no such  $t$ , let  $t_1$  be  $a_1$ .) For any sequent  $\Lambda$ , let  $\Lambda^0$  denote the sequent obtained from  $\Lambda$  by removing every occurrence of  $\forall xA$ . For each  $i = 1, \dots, m$  let  $\Gamma_i'$  be the sequent  $\Gamma_i^0, A(t_1), \dots, A(t_n)$ . We now claim that the sequence of sequents

$$\langle \Gamma_1'; \dots; \Gamma_m'; A(t_1), \dots, A(t_n), \Delta \rangle$$

is a  $k, F_k$ -refutation of  $A(t_1), \dots, A(t_n), \Delta$ . Since  $\Gamma_m'$  is the sequent  $\Delta^0, A(t_1), \dots, A(t_n)$ , then the final sequent  $A(t_1), \dots, A(t_n), \Delta$  follows from  $\Gamma_m'$  by the structure rule. The justifications for the  $\Gamma_i'$  are as follows. If  $\Gamma_i$  is an axiom, then clearly  $\Gamma_i'$  is also an axiom since  $\forall xA$  is neither an atom nor the negation of an atom. Suppose  $\Gamma_i$  follows from  $\Gamma_j$  (and  $\Gamma_l$ ) by some rule of inference R. One of the following cases must hold.

*Case 1.* The formula  $\forall xA$  is neither a major nor minor formula in this application of the rule R: Then  $\Gamma_i'$  follows from  $\Gamma_j'$  (and  $\Gamma_l'$ ) by the same rule R. (This includes the case where R is the structure rule.)

*Case 2.* The formula  $\forall xA$  is the major formula in this application of the rule R: Then R must be the  $\gamma$ -rule,  $\Gamma_i$  is of the form  $\forall xA, \Lambda$ , and  $\Gamma_j$  of the form  $A(t), \Lambda$ . Hence  $t$  is one of the terms  $t_1, \dots, t_n$ , and therefore  $\Gamma_i'$  follows from  $\Gamma_j'$  by the structure rule.

*Case 3.* The formula  $\forall xA$  is a minor formula in this application of the rule R: Then  $\Gamma_i'$  follows from  $\Gamma_j'$  (or  $\Gamma_l'$ ) by the structure rule.

### EXERCISE

Show how the required refutations for the other four parts of Theorem V.6 can be constructed.

The technique employed in proving Theorem V.6 can be used to prove the following theorem.

**THEOREM V.7.** For any atom  $A$ , if  $k, F_k\text{-Ref}(\neg A, \Delta)$  and  $k, F_k\text{-Ref}(A, \Delta)$ , then  $k, F_k\text{-Ref}(\Delta)$ .

*Proof.* (Note: we cannot apply the cut rule to prove this theorem since the index of  $A$  may be greater than  $k$ .)

*Case 1.*  $A$  is the formula  $f$ : Let  $\langle \Gamma_1; \dots; \Gamma_m \rangle$  be any  $k, F_k$ -refutation of  $\neg f, \Delta$ . For each  $i = 1, \dots, m$ , let  $\Gamma_i^0$  be the sequent obtained from  $\Gamma_i$  by removing each occurrence of  $\neg f$ . Then  $\langle \Gamma_1^0; \dots; \Gamma_m^0; \Delta \rangle$  is a  $k, F_k$ -refutation of  $\Delta$ .

*Case 2.*  $A$  is not the formula  $f$ : Let  $\langle \Gamma_1; \dots; \Gamma_m \rangle$  be any  $k, F_k$ -refutation of  $A, \Delta$  and  $\langle \Lambda_1; \dots; \Lambda_n \rangle$  any  $k, F_k$ -refutation of  $\neg A, \Delta$ . For each  $i = 1, \dots, m$ , let  $\Gamma_i^0$  be the sequent obtained from  $\Gamma_i$  by removing each occurrence of  $A$  and let  $\Gamma_i'$  be the sequent  $\Gamma_i^0, \Delta$ . Then the sequence

$$\langle \Lambda_1; \dots; \Lambda_n; \Gamma_1'; \dots; \Gamma_m'; \Delta \rangle$$

is a  $k, F_k$ -refutation of  $\Delta$ . The proof of this assertion is identical to the proof of Theorem V.6, except for the following two changes: (1) Case 2 in the proof of Theorem V.6 does not apply since no rule of inference has an atom as its major formula, and (2) if  $\Gamma_i$  is an axiom by virtue of the fact that both  $A$  and  $\neg A$  occur in it, then  $\Gamma_i'$  follows from  $\Lambda_n$  (i.e.,  $\neg A, \Delta$ ) by the structure rule.

In order to prove the Cut Elimination Theorem we would like to improve on Theorem V.6(v) by showing that if  $k, F_k\text{-Ref}(\delta, \Delta)$ , then for any term  $t$ ,  $k, F_k\text{-Ref}(\delta(t), \Delta)$ . The following theorem shows that this is indeed the case provided that certain conditions are satisfied.

**THEOREM V.8.** Let  $\delta$  be any existential formula and  $\Delta$  any  $\varepsilon$ -free sequent. Suppose  $k, F_k\text{-Ref}(\delta, \Delta)$ , where  $\text{ind}(\delta) \geq k$  and the length of  $\delta$  is at least as great as the length of any member of  $F_k$ . Then for any term  $t$ ,

$$k, F_k\text{-Ref}(\delta(t), \Delta).$$

*Proof.* Since  $k, F_k\text{-Ref}(\delta, \Delta)$ , then by Theorem V.6(v)  $k, F_k\text{-Ref}(\delta(\varepsilon\delta), \Delta)$ . Recall that if  $\delta$  is  $\exists yB$ , then  $\delta(\varepsilon\delta)$  is  $B(\varepsilon yB)$ , and if  $\delta$  is  $\neg\forall yC$ , then  $\delta(\varepsilon\delta)$  is  $\neg C(\varepsilon y\neg C)$ . Thus the term denoted by  $\varepsilon\delta$  has the same index and the same length as  $\delta$ . Let  $\delta(\varepsilon\delta)$  be the formula  $B(\varepsilon yB)$  and let  $\langle \Gamma_1; \dots; \Gamma_m \rangle$  be the  $k, F_k$ -refutation of  $B(\varepsilon yB), \Delta$ . Notice that by our assumption on the length of  $\delta$  the term  $\varepsilon yB$  does not occur in any member of  $F_k$ . For each  $i = 1, \dots, m$ , let  $\Gamma_i^0$  be the sequent that we obtain from  $\Gamma_i$  by replacing each occurrence of  $\varepsilon yB$  (in the members of  $\Gamma_i$ ) by the term  $t$ . Let  $\Gamma_i'$  be the sequent  $\Gamma_i^0, B(t)$ . We shall prove that the sequence  $\langle \Gamma_1'; \dots; \Gamma_m'; B(t), \Delta \rangle$  is a  $k, F_k$ -refutation of  $B(t), \Delta$ , i.e., of  $\delta(t), \Delta$ .

First of all, since  $\Delta$  is  $\varepsilon$ -free, then  $\Gamma_m'$  is  $B(t), \Delta, B(t)$ , so that  $B(t), \Delta$  follows from  $\Gamma_m'$  by the structure rule. If  $\Gamma_i$  is an axiom, then  $\Gamma_i'$  is also an axiom,

and if  $\Gamma_i$  follows from  $\Gamma_j$  (and  $\Gamma_l$ ) by some rule of inference other than the  $\gamma$ -rule or  $\delta$ -rule, then clearly  $\Gamma_i'$  follows from  $\Gamma_j'$  (and  $\Gamma_l'$ ) by the same rule.

We now consider the case where  $\Gamma_i$  follows from  $\Gamma_j$  by the  $\gamma$ -rule or  $\delta$ -rule. First of all, suppose that  $\Gamma_j$  is of the form  $B(\varepsilon yB), \Lambda$ . Then  $\Gamma_j'$  is  $B(t), \Lambda^0, B(t)$ , and  $\Gamma_i'$  follows from  $\Gamma_j'$  by the structure rule. On the other hand, suppose  $\Gamma_j$  is of the form  $A(s), \Lambda$ , where  $A(s)$  is not  $B(\varepsilon yB)$ . Assume  $\Gamma_i$  is of the form  $\forall xA, \Lambda$ . (The other possibilities can be handled similarly.) As in the proof of Theorem III.6, we want to rule out the possibility that  $\varepsilon yB$  is of the form  $[p]_t^x$  where  $p$  is a quasi  $\varepsilon$ -term subordinate to  $\forall xA$ . Suppose that this is the case. Then the index of  $\forall xA$  is at least  $k + 2$  (since the index of  $\varepsilon yB$  is at least  $k$ ). Consequently, by the subformula property of  $k, F_k$ -refutations,  $\forall xA$  must be a subformula of some member of  $\Delta$ . But this is impossible since every member of  $\Delta$  is  $\varepsilon$ -free, and an improper formula, such as  $\forall xA$ , cannot be a subformula of an  $\varepsilon$ -free formula. Hence  $\Gamma_j'$  is of the form  $\forall xA^0, \Lambda^0, B(t)$ , and  $\Gamma_i'$  is of the form  $A^0(s^0), \Lambda^0, B(t)$ ; and  $\Gamma_i'$  follows from  $\Gamma_j'$  by the  $\gamma$ -rule. This completes the proof. (Why is this refutation still a  $k, F_k$ -refutation?)

**COROLLARY.** For any  $\varepsilon$ -free  $\Delta$ , any existential formula  $\delta$ , and any term  $t$ , if  $\text{Ref}(\delta, \Delta)$ , then  $\text{norm-Ref}(\delta(t), \Delta)$ .

### EXERCISE

Using the cut rule give a simple proof of the following: for any sequent  $\Delta$ , any existential formula  $\delta$ , and any term  $t$ , if  $\text{Ref}(\delta, \Delta)$ , then  $\text{Ref}(\delta(t), \Delta)$ . Using this result prove that Theorem V.8 still holds if we remove the condition that  $\text{ind}(\delta) \geq k$ .

### 6 The Cut Elimination Theorem

The next theorem is the last subsidiary result needed in the proof of the Cut Elimination Theorem.

**THEOREM V.9** (The cut property). For each  $i = 1, \dots, n$ , let  $A_i$  and  $B_i$  be contradictory formulae with index  $\leq k$ . If  $k, F_k\text{-Ref}(A_1, \dots, A_n, \Delta)$  and  $k, F_k\text{-Ref}(B_i, \Delta)$  for each  $i = 1, \dots, n$ , then  $k, F_k\text{-Ref}(\Delta)$ .

*Proof.* The proof is by induction on  $n$ .

*Case 1.*  $n = 1$ : Assume  $k, F_k\text{-Ref}(A, \Delta)$  and  $k, F_k\text{-Ref}(B, \Delta)$ , where  $A$  and  $B$  are contradictory and each has index  $\leq k$ . Then  $\Delta$  follows from  $A, \Delta$  and  $B, \Delta$  by the cut rule, where the cut formula has index  $< k$  (since its negation has index  $\leq k$ ), and hence  $k, F_k\text{-Ref}(\Delta)$ .

*Case 2.*  $n \geq 2$ : Assume  $k, F_k\text{-Ref}(A_1, \dots, A_n, \Delta)$  and  $k, F_k\text{-Ref}(B_i, \Delta)$  for each  $i = 1, \dots, n$ . By the structure rule, from  $k, F_k\text{-Ref}(B_1, \Delta)$  we infer

$k, F_k\text{-Ref}(B_1, A_2, \dots, A_n, \Delta)$ . Case 1 then gives  $k, F_k\text{-Ref}(A_2, \dots, A_n, \Delta)$ . The desired result now follows by the induction hypothesis.

**THEOREM V.10** (The Cut Elimination Theorem). For any  $\varepsilon$ -free sequent  $\Delta$ , if  $\text{Ref}(\Delta)$ , then  $\text{norm-Ref}(\Delta)$ .

*Proof.* By using an obvious inductive argument, we can reduce this theorem to the following lemma.

**LEMMA.** Suppose  $k, \{A\} \cup F_k\text{-Ref}(\Delta)$ , where  $\Delta$  is  $\varepsilon$ -free and  $A$  is any formula with index  $k$  whose length is at least as great as the length of any member of  $F_k$ . Then  $k, F_k\text{-Ref}(\Delta)$ .

*Proof.* Let  $\langle \Gamma_1; \dots; \Gamma_m \rangle$  be a  $k, \{A\} \cup F_k$ -refutation of  $\Delta$ . For each  $i = 1, \dots, m$ , let  $\Gamma_i'$  be  $\Gamma_i, A$ , and let  $\Gamma_i''$  be  $\Gamma_i, \neg A$ . Then the sequence

$$\langle \Gamma_1'; \dots; \Gamma_m'; A, \Delta \rangle$$

is a  $k, F_k$ -refutation of  $A, \Delta$ , and the sequence

$$\langle \Gamma_1''; \dots; \Gamma_m''; \neg A, \Delta \rangle$$

is a  $k, F_k$ -refutation of  $\neg A, \Delta$ . For, if  $\Gamma_i$  follows from  $\Gamma_j$  and  $\Gamma_l$  by an application of the cut rule where  $A$  and  $\neg A$  are the minor formulae, then  $\Gamma_i'$  and  $\Gamma_i''$  follow by the structure rule. Otherwise,  $\Gamma_i'$  and  $\Gamma_i''$  have the same justification that  $\Gamma_i$  has in the original refutation. Hence we have

$$(1) \quad k, F_k\text{-Ref}(A, \Delta), \quad \text{and}$$

$$(2) \quad k, F_k\text{-Ref}(\neg A, \Delta).$$

As in the proof of Theorem V.2, one of the following cases must hold.

*Case 1.*  $A$  is an atom: Then by Theorem V.7, (1) and (2) yield  $k, F_k\text{-Ref}(\Delta)$ .

*Case 2.*  $A$  is of the form  $\neg B$ : By Theorem V.6(i), (2) yields  $k, F_k\text{-Ref}(B, \Delta)$ .

Since  $\text{ind}(\neg B) = k$ , then by (1) and the cut property we have  $k, F_k\text{-Ref}(\Delta)$ .

*Case 3.* One of the two formulae,  $A$  and  $\neg A$ , is a conjunctive formula  $\alpha$  and the other is a disjunctive formula  $\beta$ : Thus we have

$$(3) \quad k, F_k\text{-Ref}(\alpha, \Delta), \quad \text{and}$$

$$(4) \quad k, F_k\text{-Ref}(\beta, \Delta).$$

By Theorem V.6, (3) and (4) yield

$$(5) \quad k, F_k\text{-Ref}(\alpha_1, \alpha_2, \Delta),$$

$$(6) \quad k, F_k\text{-Ref}(\beta_1, \Delta), \quad \text{and}$$

$$(7) \quad k, F_k\text{-Ref}(\beta_2, \Delta).$$

Since  $\alpha_1$  and  $\beta_1$  are contradictory and  $\alpha_2$  and  $\beta_2$  are contradictory, and the index of each of these formulae is less than  $k$ , then the cut property yields  $k, F_k\text{-Ref}(\Delta)$ .

Case 4. One of the two formulae,  $A$  and  $\neg A$ , is a universal formulae  $\gamma$  and the other is an existential formula  $\delta$ : Thus we have

$$(8) \quad k, F_k\text{-Ref}(\gamma, \Delta), \quad \text{and}$$

$$(9) \quad k, F_k\text{-Ref}(\delta, \Delta).$$

By (8) and Theorem V.6(iv) there exist terms  $t_1, \dots, t_n$  such that

$$(10) \quad k, F_k\text{-Ref}(\gamma(t_1), \dots, \gamma(t_n), \Delta).$$

Since  $\Delta$  is a  $\varepsilon$ -free and since  $\delta$  is either  $A$  or  $\neg A$  we can apply Theorem V.8 to (9) and get for each  $i = 1, \dots, n$

$$(11) \quad k, F_k\text{-Ref}(\delta(t_i), \Delta).$$

The desired result,  $k, F_k\text{-Ref}(\Delta)$  now follows from (10) and (11) by the cut property.

This completes the proof of the lemma and hence also of Theorem V.10.

The following counterexample shows that the Cut Elimination Theorem no longer holds if the condition that  $\Delta$  is  $\varepsilon$ -free is removed.

Let  $P$  be any 1-place predicate symbol. Since the sequent  $\exists xPx, \neg PexPx$  follows by the  $\delta$ -rule from the axiom  $PexPx, \neg PexPx$ , we have at once  $norm\text{-Ref}(\exists xPx, \neg PexPx)$ . Hence, by the exercise at the end of the last section, there exists a refutation of  $Pt, \neg PexPx$  for any term  $t$ . However, since  $Pt$  and  $PexPx$  are both atoms, it is easy to see that there exists no normal refutation of the sequent  $Pt, \neg PexPx$  when  $t$  is any term other than  $exPx$ . Consequently, if we let  $\Delta$  be the sequent  $Pa, \neg PexPx$  the Cut Elimination Theorem does not hold. (For possible ways of modifying the axioms and rules of inference of the sequent calculus so that the Cut Elimination Theorem holds for arbitrary  $\Delta$ , see Maehara [1955], [1957] and Curry [1963], page 342.)

## 6.1 Applications

In this section we show how the Cut Elimination Theorem can be used to provide new proofs of some of our earlier results. In each case the main function of this theorem is to prove the eliminability of certain symbols or certain types of formulae.

**THEOREM V.11** (The Second  $\varepsilon$ -Theorem—weaker form). *For any  $\varepsilon$ -free  $X$  and  $A$ , if  $X \vdash_{\varepsilon} A$  (without E2), then  $X \vdash_{PC} A$ .*

*Proof.* Since  $X \vdash_{\varepsilon} A$  (without E2), then there exist formulae  $B_1, \dots, B_n$  such that  $B_1, \dots, B_n \vdash_{\varepsilon} A$  (without E2). Let  $\Delta$  be the sequent  $B_1, \dots, B_n$ . Then  $Ref(\neg A, \Delta)$  by Theorem V.4. Hence  $norm\text{-Ref}(\neg A, \Delta)$  by the Cut

Elimination Theorem and therefore  $X \vdash_{\varepsilon} A$  by Theorem V.5. This implies  $X \vdash_{PC} A$  by Theorem III.3.

Notice that the Cut Elimination Theorem has been used in this proof to show that any improper formulae can be eliminated from the original deduction of  $A$  from  $X$ . (cf. Theorem V.5.)

**THEOREM V.12.** *For any  $\varepsilon$ -free  $X$  and  $A$ ,  $X \vdash_{PC} A$  iff there exists a sequent  $\Delta$ , such that  $\Delta^* \subseteq X$  and  $norm\text{-Ref}(\neg A, \Delta)$ .*

*Proof.* Assume  $X \vdash_{PC} A$ . Then  $X \vdash_{\varepsilon} A$  by Theorem III.2. Hence  $norm\text{-Ref}(\neg A, \Delta)$  as in the proof of Theorem V.11. Conversely, if  $norm\text{-Ref}(\neg A, \Delta)$ , where  $\Delta^* \subseteq X$ , then  $X \vdash_{PC} A$  as in the proof of Theorem V.11.

Using this essential link between deductions in the predicate calculus and normal refutations, we can now give new proofs of the First  $\varepsilon$ -Theorem and of the eliminability of the identity symbol (cf. Theorem III.10 and Theorem III.15).

**THEOREM V.13.** *Let  $X$  be any set of  $\varepsilon$ -free  $\forall$ -prenex formulae and  $B$  any  $\exists$ -prenex formula. If  $X \vdash_{PC} B$ , then  $A_1, \dots, A_m \vdash_{EC} B_1 \vee \dots \vee B_n$  where the  $A_i$  are certain substitution instances of the matrices of members of  $X$  and the  $B_j$  are certain substitution instances of the matrix of  $B$ .*

*Proof.* By Theorem V.12, there exists a sequent  $\Delta$  such that  $\Delta^* \subseteq X$  and  $norm\text{-Ref}(\neg B, \Delta)$ . By repeated applications of the invertibility of the  $\gamma$ -rule (Theorem V.6(iv)) we get  $norm\text{-Ref}(\neg B_1, \dots, \neg B_m, A_1, \dots, A_n)$ , where the  $B_j$  and  $A_i$  are of the required form. By applying the  $\alpha$ -rules and the structure rule, this yields

$$norm\text{-Ref}(\neg(B_1 \vee \dots \vee B_n), A_1, \dots, A_n).$$

Now since each of the  $A_i$  and  $B_j$  are quantifier-free, then by the subformula property of normal refutations every member of the refutation is elementary (since any stray  $\varepsilon$ -terms can be replaced by  $a$ ). Consequently, as in the proof of Theorem V.5 we get

$$A_1, \dots, A_m \vdash_{EC} B_1 \vee \dots \vee B_n.$$

**THEOREM V.14.** *For any identity-free ( $\varepsilon$ -free)  $X$  and  $A$ , if  $X \vdash_{PC} A$ , then there exists a deduction of  $A$  from  $X$  in the predicate calculus without identity.*

*Proof.* Since  $X \vdash_{PC} A$ , then by Theorem V.12,  $norm\text{-Ref}(\neg A, \Delta)$  for some  $\Delta^* \subseteq X$ . This implies by Theorem V.1 that there exists a normal refutation of  $\neg A, \Delta$  in which every sequent is identity-free. Returning now to the predicate calculus by means of Theorem V.12, we obtain an identity-free deduction of  $A$  from  $X$ .

## BIBLIOGRAPHY

ACKERMANN, W.

1924. 'Begründung des "tertium non datur" mittels der Hilbertschen Theorie der Widerspruchsfreiheit', *Mathematische Annalen*, **93**, 1–36.  
1937–8. 'Mengentheoretische Begründung der Logik', *Mathematische Annalen*, **115**, 1–22.  
1940. 'Zur Widerspruchsfreiheit der Zahlentheorie', *Mathematische Annalen*, **117**, 162–194.

ASSER, G.

1957. 'Theorie der logischen Auswahlfunktionen', *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, **3**, 30–68.

BERNAYS, P.

1927. 'Zusatz zu Hilberts Vortrag über "Die Grundlagen der Mathematik"', *Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität*, **6**, 89–92. (cf. von Heijenoort [1967], pp. 485–489.)  
1935. 'Hilberts Untersuchungen über die Grundlagen der Arithmetik', in Hilbert [1935], pp. 196–216.  
1958. *Axiomatic Set Theory*, Amsterdam. (With a historical introduction by A. A. Fraenkel.)

BETH, E. W.

1959. *The Foundations of Mathematics*, Amsterdam.

BOURBAKI, N.

1954. *Éléments de Mathématique, Livre I (Théorie des Ensembles), Chap. I et II*, Paris.

CARNAP, R.

1961. 'On the use of Hilbert's  $\varepsilon$ -operator in scientific theories', *Essays on the Foundations of Mathematics dedicated to A. A. Fraenkel on his Seventieth Anniversary*, Jerusalem, pp. 156–164.

CHURCH, A.

1956. *Introduction to Mathematical Logic, Volume I*, Princeton.

COHEN, P. J.

1963. *The independence of the axiom of choice* (mimeographed, Stanford University).  
1966. *Set Theory and the Continuum Hypothesis*, New York.

CURRY, H. B.

1963. *Foundations of Mathematical Logic*, New York.

- DENTON, J., and DREBEN, B.  
1969. 'The Herbrand theorem and the consistency of arithmetic'. Forthcoming.
- DREBEN, B., ANDREWS, P., and AANDERAA, S.  
1963. 'False lemmas in Herbrand', *Bulletin of the American Mathematical Society*, **69**, 699-706.
- FRAENKEL, A. A., and BAR-HILLEL, Y.  
1958. *Foundations of Set Theory*, Amsterdam.
- FRAYNE, T., MOREL, A. C., and SCOTT, D. S.  
1962. 'Reduced direct products', *Fundamenta Mathematicae*, **51**, 195-228.
- GENTZEN, G.  
1934-5. 'Untersuchungen über das logischen Schließen', *Mathematische Zeitschrift*, **39**, 176-210, 405-431.  
1936. 'Die Widerspruchsfreiheit der reinen Zahlentheorie', *Mathematische Annalen*, **112**, 493-565.
- GÖDEL, K.  
1931. 'Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I', *Monatshefte für Mathematik und Physik*, **38**, 173-198.  
1940. *The Consistency of the Axiom of Choice and of the Generalized Continuum Hypothesis with the Axioms of Set Theory*, Princeton.
- HAILPERIN, T.  
1957. 'A theory of restricted quantification, I and II', *Journal of Symbolic Logic*, **22**, 19-35 and 113-129.
- van HEIJENOORT, J.  
1967. (editor) *From Frege to Gödel, A Source Book in Mathematical Logic, 1879-1931*. Cambridge, Mass.
- HENKIN, L.  
1949. 'The completeness of the first-order functional calculus', *Journal of Symbolic Logic*, **14**, 159-166.
- HERBRAND, J.  
1930. 'Recherches sur la théorie de la démonstration', *Travaux de la Société des Sciences et des Lettres de Varsovie, Classe III sciences mathématiques et physiques*, no. 33, pp. 33-160.  
1931. Sur la non-contradiction de l'arithmétique, *Journal für die reine und angewandte Mathematik*, **166**, 1-8. (cf. van Heijenoort [1967], pp. 618-628.)
- HERMES, H.  
1965. *Eine Termlogik mit Auswahloperator*, Berlin-Heidelberg-New York.
- HILBERT, D.  
1923. 'Die logischen Grundlagen der Mathematik', *Mathematische Annalen*, **88**, 151-165.

1926. 'Über das Unendliche', *Mathematische Annalen*, **95**, 161-190. (cf. van Heijenoort [1967], pp. 367-392.)  
1928. 'Die Grundlagen der Mathematik', *Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität*, **6**, 65-85. (cf. van Heijenoort [1967], pp. 464-479.)  
1935. *Gesammelte Abhandlungen*, vol. 3, Berlin.
- HILBERT, D., and BERNAYS, P.  
1934. *Grundlagen der Mathematik*, vol. 1, Berlin. Reprinted Ann Arbor, Mich., 1944.  
1939. *Grundlagen der Mathematik*, vol. 2, Berlin. Reprinted Ann Arbor, Mich., 1944.
- KLEENE, S. C.  
1952. *Introduction to Metamathematics*, Amsterdam, Groningen, New York.
- KNEEBONE, G. T.  
1963. *Mathematical Logic and the Foundations of Mathematics*, London.
- KREISEL, G.  
1964. 'Hilbert's Programme', *Philosophy of Mathematics*, (edited by P. Benacerraf and H. Putnam), Oxford, pp. 157-180.
- LEISENRING, A. C.  
1968. 'An abstract property of formalized languages which contain Hilbert's  $\epsilon$ -symbol', *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, **14**, 81-92.
- LÉVY, A.  
1961. 'Comparing the axioms of local and universal choice', *Essays on the Foundations of Mathematics dedicated to A. A. Fraenkel on his Seventieth Anniversary*, Jerusalem, pp. 83-90.
- LYNDON, R. C.  
1966. *Notes on Logic*, Princeton.
- MAEHARA, S.  
1955. 'The predicate calculus with  $\epsilon$ -symbol', *Journal of the Mathematical Society of Japan*, **7**, 323-344.  
1957. 'Equality axioms on Hilbert's  $\epsilon$ -symbol', *Journal of the Faculty of Science, University of Tokyo*, Sect. 1; **7**, 419-435.
- MENDELSON, E.  
1964. *Introduction to Mathematical Logic*, Princeton.
- MONTAGUE, R., and HENKIN, L.  
1956. 'On the definition of formal deduction', *Journal of Symbolic Logic*, **21**, 129-136.
- von NEUMANN, J.  
1927. 'Zur Hilbertschen Beweistheorie', *Mathematische Zeitschrift*, **26**, 1-46.

QUINE, W. V.

1950. 'On natural deduction', *Journal of Symbolic Logic*, **15**, 93–102.

RASIOWA, H.

1956. 'On the  $\varepsilon$ -theorems', *Fundamenta Mathematicae*, **43**, 156–165.

ROBINSON, A.

1963. *Introduction to Model Theory and the Metamathematics of Algebra*, Amsterdam.

SCHOENFIELD, J. R.

1954. 'A relative consistency proof', *Journal of Symbolic Logic*, **19**, 21–28.

1967. *Mathematical Logic*, Reading, Mass.

SMULLYAN, R.

1965. 'Analytic natural deduction', *Journal of Symbolic Logic*, **30**, 123–139.

1966a. 'Trees and nest structures', *Ibid.*, **31**, 303–321.

1966b. 'Finite nest structures and propositional logic', *Ibid.*, **31**, 322–324.

TAIT, W. W.

1965. 'The substitution method', *Journal of Symbolic Logic*, **30**, 175–192.

TARSKI, A.

1930. 'Fundamentale Begriffe der Methodologie der deduktiven Wissenschaften, I', *Monatshefte für Mathematik und Physik*, **37**, 361–404.

WANG, H.

1955. 'On denumerable bases of formal systems', *Mathematical Interpretations of Formal Systems*, Amsterdam, pp. 57–84.

1963. 'The predicate calculus', *A Survey of Mathematical Logic*, Amsterdam, pp. 307–321.

## INDEX OF SYMBOLS

(The numbers refer to the pages on which the symbols are introduced.)

$\varepsilon$	1
$\neg, \vee, \wedge, \rightarrow, \leftrightarrow, \forall, \exists$	2, 6
$\mathcal{L}$	2, 10
$\vDash$	2, 22
$\emptyset$	3
$\mathcal{F}$	3, 38
$\vdash_{\mathcal{F}}$	3, 38
$\mathcal{F}$	4, 85
$f$	6
$\in, \subseteq, \cup, \cap, \setminus, \{x: \dots x \dots\}$	10
$\mathcal{L}(\mathcal{V})$	10
$V_r$	11
<i>Ind</i>	11
$V(A)$	12
$F_{\mathcal{L}}$	12
$T_{\mathcal{L}}$	12
$[A]_i^x$	12
$A(t)$	13
$[A]_{i_1 \dots i_n}^{x_1 \dots x_n}$	13
$\Sigma$	14
$\text{dom } \Sigma$	14
$\alpha, \beta, \gamma, \delta$	15
$\alpha_1, \alpha_2, \beta_1, \beta_2, \gamma(t), \delta(t), \delta(\varepsilon\delta)$	15
$\mathcal{B}$	16
$H_{\neg}, H_{\vee}, H_{\wedge}, H_{\rightarrow}$	16
$\overline{\mathcal{P}}$	17
$\mathfrak{M}, M, \Pi, \Phi$	18
$\theta_{\mu}^x$	19
$\mathfrak{M}\theta(A)$	19
$\theta_1 \sim \theta_2$	19
$\mathfrak{M} \text{ Sat } X, \text{ Sat } X, \text{ m-Sat } X$	22
$C_s$	22
$\text{Max}_C X$	23
$\eta$	34

$\varepsilon(\mathcal{V})$	39
$\vdash_{\varepsilon}, \vdash$	41
$X \vdash \forall [A]$	51
$E_1 \equiv E_2$	52
$A_H, A_S$	56
$\text{EC}(\mathcal{V}), \text{EC}'(\mathcal{V})$	58
$\text{PC}(\mathcal{V})$	59
$\text{PC}'(\mathcal{V})$	62
$A^+$	66
$\vdash_{\varepsilon^*}$	67
$rk(t)$	70
$\vdash^r, \vdash^{r, T_r}$	71
$', g, +, \cdot$	89
$N$	89
$\mathfrak{N}$	89
$\mathcal{F}_{\varepsilon}$	100
$i$	100
$\wedge$	100, 108
$\sigma$	108
$C(\mathcal{V})$	109
$\vdash_C$	109
$\Delta^*$	114
$\text{Ref}(\Delta), \text{norm-Ref}(\Delta)$	116
$\text{ind}(A)$	119
$k, F_k\text{-Ref}(\Delta)$	121
$1, \emptyset\text{-Ref}(\Delta)$	121



Formation rules G1–G9 for quasi-terms and quasi-formulae	11
Conditions C1–C4 for finitary closure	23
Conditions L1–L6 for logical closure	24
Axioms $(\varepsilon_0)$ , $(\varepsilon_2)$	33
Condition L7 for logical closure	34
Condition L8 for logical closure	36
Axioms P1–P10, Q1–Q4, E1–E3 for $\varepsilon(\mathcal{Y})$	39
Properties R1–R4 of rank	70
Axioms S1–S5, S4'–S5', S1''–S5''	86
Axioms N1–N7, I	89
Axioms N1'–N7', $\varepsilon_1$ , $\varepsilon_2$	92
Axiom $\varepsilon_3$	96
Axiom S8 (Bourbaki)	106
Axioms $A'_\sigma$ , $A''_\sigma$ (Bernays)	108
Formation rule G8 <sup>0</sup>	108
Axioms CS, $\sigma_1$ , $\sigma_2$	109
Axiom $A_\sigma$ (Bernays)	112
Properties K1–K4 of index	119

## INDEX OF RULES OF INFERENCE

*(The numbers refer to the pages on which the rules  
are introduced.)*

rule of modus ponens	40
MP-rule	42
$\rightarrow$ -elimination rule	42
$\rightarrow$ -introduction rule	42
sylogism rule	43
contrapositive rules	43
$\neg$ -rules	44
contradiction rule	46, 119
$\neg\neg$ -rule	46, 115
$\alpha$ -rule	46
$\beta$ -rule	46, 115
tautology rule	47
$\forall$ -elimination rule	48
$\forall$ -introduction rule	48
$\exists$ -elimination rule	48
$\exists$ -introduction rule	48
substitution rule	49
generalization rule	49
$\exists$ -rule	49, 60
$\forall$ -rule	49
distribution rule	50
equivalence rule	52, 69
rule of relabelling bound variables	54
$\alpha_1$ -rule	115
$\alpha_2$ -rule	115
$\gamma$ -rule	115
$\delta$ -rule	115
E1-rule	115
cut-rule	116
structure rule	116

## GENERAL INDEX

- Aanderaa, S., 83  
Ackermann, W., 4, 5, 35, 99, 103, 104, 105  
adequate, 18  
adjoining new symbols to vocabulary, 10  
admissible relabelling of a bound variable, 53  
Andrews, P., 83  
appears in, 11  
arithmetic, formal, 88 ff  
    restricted, 97  
    standard model of, 85  
Asser, G., 5, 33, 35  
assignment, 19  
assumption, 40  
atom, 11, 14  
axiom, 3, 39  
    equality, 40  
    non-logical, 4, 85  
    propositional, 40  
    quantificational, 40  
axiom of choice, 1, 5, 9, 24, 105 ff, 112  
axiom schema, 39  
axiom schema of induction, 89  
axiom schema of replacement, 80, 106  
axioms of arithmetic, 89  
axioms of  $\varepsilon$ -calculus, 39–40  
axioms of  $\varepsilon^*$ -calculus, 66  
axioms of elementary calculus, 58  
axioms of predicate calculus, 60  
axioms of sequent calculus, 115
- Bernays, P., 5, 99, 102, 103, 107, 108, 112  
Beth, E., 105  
bound occurrence, 11  
bound variable, 7, 11  
Bourbaki, N., 5, 35, 103, 104, 106
- calculus, elementary, 58 ff  
    predicate, 3, 59 ff  
cardinal number, 6, 104  
cardinality of a language, 16, 25  
cardinality of a model, 18, 22  
Carnap, R., 5, 105  
C-axiom, 115  
choice function, 1, 17, 33 ff, 108 ff  
Church's schema, 103  
class calculus, 109  
class language, 108  
class operator, 107  
closure, deductive, 38  
closure, logical, 24  
closure operation, finitary, 23  
Cohen, P., 1  
collectivizing, 103  
collision of bound variable, 7  
Compactness Theorem, 9, 29  
complete, 3, 38  
Completeness Theorem, 45  
comprehension operator, 103  
conclusion, 115  
conjunctive formula, 7, 15  
connectives, propositional, 2, 7  
consequence, semantic, 2, 22  
consistency of arithmetic, 96 ff  
consistency of  $\varepsilon$ -calculus, 47–48  
consistency of formal systems, 38  
consistency of formal theories, 4, 85  
consistency proof, 86  
Consistency Theorem, 87  
consistent, 4, 38, 47, 85  
constant, 8, 10  
constructive, 2  
contained in, 11  
contradiction rule, 46, 118 ff  
contradictory formulae, 15  
contrapositive rules, 46  
contrary, 15  
Curry, H., 126  
Cut Elimination Theorem, 117, 124 ff  
cut formula, 116  
cut property, 124  
cut-rule, 115
- deducibility, 3, 38  
deduction, 38, 39, 60, 66, 116

- Deduction Theorem, 8, 42, 60  
 deductive closure, 38  
 defined symbol, 6  
 degree, 81, 98  
 Denton, J., 83  
 derivation, 60  
 derived rule of inference, 42  
 disjunctive formula, 7, 15  
 distribution rule, 50  
 Dreben, B., 83  
 duality principle, 15–16
- $\varepsilon_1$ -axiom, 92  
 $\varepsilon_2$ -axiom, 92  
 $\varepsilon$ -calculus, 3, 39 ff  
 $\varepsilon^*$ -calculus, 66 ff  
 $\varepsilon$ -formula, 32  
 $\varepsilon$ -free, 11  
 $\varepsilon$ -symbol, 1  
 $\varepsilon$ -Theorem, First, 5, 64, 78–79  
 $\varepsilon$ -Theorem, Second, 1, 3, 5, 64, 79
- $\eta$ -rule, 34  
 $\eta$ -symbol, 34
- E1-axiom, 40, 50  
 E2-axiom, 40, 70, 75 ff  
 E3-axiom, 40, 45  
 E2-term, 71  
 E2 Elimination Theorem, 75  
 elementary calculus EC ( $V$ ), 58 ff  
 elementary calculus without identity, 58  
 elementary expression, 11  
 eliminability of  $\varepsilon$ -symbol, 50, 79  
 eliminability of E2-axioms, 75 ff  
 eliminability of identity symbol, 58, 83 f, 118, 127  
 eliminability of improper formulae, 72 ff, 127  
 eliminability of  $\iota$ -symbol, 80, 94  
 eliminability of quantifiers, 6, 48, 92, 95–96  
 eliminability of  $\wedge$  and  $\sigma$ , 111  
 equivalent theories, 85  
 existence condition, 101  
 existential formula, 15  
 expression, 10  
   well-formed, 10  
 extension, 10, 85  
   inessential, 85
- $f$ , 6, 10  
 finitary closure operation, 23  
 finitary consistency proof, 48, 85 ff, 90  
 finitary reasoning, 37  
 finitary semantic closure, 22  
 First  $\varepsilon$ -Theorem, 5, 64, 78  
 formal arithmetic, 88 ff  
 formal language, 2, 10 ff  
 formal system, 2, 38  
 formal, theory, 4, 85  
 formalist, 4, 85, 87–88, 99  
 formula, 2, 12  
   atomic, 11  
   conjunctive, 7, 15  
   cut, 116  
   disjunctive, 7, 15  
   existential, 15  
   improper, 66  
   major, 116  
   minor, 116  
   molecular, 17  
   prenex, 54  
   proper, 66  
   universal, 15  
 Fraenkel, A. A. and Bar-Hillel, Y., 67, 107  
 Frayne, T., 30  
 free for, 11, 20  
 free occurrence of variable, 11  
 free variable, 11  
 $f$ -rules, 44  
 function symbol, 2, 10
- generalization rule, 49  
 Gentzen, G., 99, 114  
*Gesamtersetzung*, 97  
 Gödel, K., 4, 90, 96, 107, 112
- van Heijenoort, J., 99  
 Henkin, L., 9, 25, 27, 60  
 Herbrand, J., 79 ff, 99  
 Herbrand function, 57  
 Herbrand resolution, 57  
 Herbrand's Theorem, 64, 80 ff  
 Hilbert, D., 4 ff, 31 ff, 88  
 Hilbert, D., and Bernays, P., 5, 69 ff, 87, 92, 96 ff, 101
- $\iota$ -symbol, 34, 100 ff  
 identity-free, 11, 118

- immediate logical subformula, 117  
 improper, 66  
 inconsistent, 4, 85  
 indeterminacy of  $\varepsilon$ -symbol, 5–6, 103 ff  
 index, 118–119  
 individual symbol, 7, 8, 10  
 inessential extension, 3, 85  
 instance of axiom schema, 39  
 instance, substitution, 51  
 interpretation, 2, 19  
 introduction rule for implication  
   symbol, 42  
 introduction rules for quantifiers, 48  
 invalid, truth functionally, 46  
 invertibility of logical rules, 121 ff
- $k$ ,  $F_k$ -refutation, 121 ff  
 Kreisel, G., 87
- Language, formal, 2, 10 ff  
   object, 2  
 least number operator, 36, 92  
 Lévy, A., 113  
 logical choice function, 1  
 logical closure, 24  
 logical constant, 2, 6  
 logical subformula, 117  
 logical symbol, 6  
 logically true, 3  
 Löwenheim-Skolem Theorem, 30  
 Lyndon, R., 60
- Maehara, S., 126  
 major formula, 116  
 mathematical induction, 37, 92  
 matrix, 54  
 maximal consistent set, 25  
 maximal under  $C$ , 23 ff  
 Mendelson, E., 63, 90, 100  
 metalanguage, 2, 38 ff  
 metamathematics, 4  
 metatheory, 2, 9, 37, 86  
 minimal value, 98  
 minor formula, 116  
 model, 2, 18 ff, 89  
 modus ponens, 40  
 molecular constituent, 17  
 molecule, 17  
 Montagu, R., and Henkin, L., 60  
 Morel, A. C., 30
- von Neumann, J., 99  
 non-constructive, 2, 45  
 normal refutation, 116 ff  
 $n$ -place function symbol, 10  
 $n$ -place predicate symbol, 10  
 null term, 6, 33 ff, 103  
 numeral, 90  
 numerical value, 91  
 numerically true, 91
- object language, 2  
 occurs in, 11  
 omission of parentheses, 11–12  
 order, 10
- parentheses, rules for omission of,  
 11–12  
 power set, 102  
 predecessor, 91  
 predecessor function, 90  
 predicate calculus PC, 3, 59 ff  
 predicate calculus without identity  
   PC', 62–63  
 predicate symbol, 2, 10  
 prefix, 54  
 premiss, 115  
 prenex formula, 54, 56  
 primitive symbol, 6  
 principle of the least number, 92  
 product, 91  
 proof, 38  
 proper formula, 66  
 proposition, 10  
 propositional connectives, 2
- quantifier, 2, 4, 6–7  
 quantifier-free proof, 95  
 quasi-formula, 11  
 quasi-term, 11  
 quasi  $\varepsilon$ -term, 11  
 Q-term, 71
- rank, 69 ff, 97  
 Rank Reduction Theorem, 72  
 refutable, 116  
 refutation, 114, 116  
   normal, 116  
 relabelling bound variables, 7, 53–54, 95

- replacement operator, 14
- representable, 27
- representative, 27
- resolution, Herbrand, 57
  - Skolem, 57
- $r$ ,  $T$ -deduction, 71
- rule of inference, 3, 39, 42
- rules of inference for  $\varepsilon$ -calculus, 40
- rules of inference for predicate calculus, 60
- rules of inference for sequent calculus, 115
  
- Satisfiability Theorem, 24 ff
- satisfiable, 22
- satisfies, 22
- schema, axiom, 39
- Schröter, K., 33
- scope, 11, 50, 97
- Scott, D. A., 30
- Second  $\varepsilon$ -axiom, 33
- Second  $\varepsilon$ -Theorem, 1, 5, 64, 79, 126
- semantic consequence, 2, 22
- semantics, 9
- separation symbol, 10
- sequent, 114
- sequent calculus, 114 ff
- set theory, 100 ff
- Shoenfield, J. R., 68, 113
- skeleton, 66, 118, 119
- Skolem function, 57, 102–103
- Skolem resolution, 57
- Skolem's Theorem, 64, 79
- Smullyan, R., 7, 15
- sound, 3, 38
- standard model of arithmetic, 89
- structure rule, 115
- subformula, logical, 117
- subformula property of  $k$ ,  $F_k$ -refutations, 121
- subformula property of normal refutations, 117–118
- subordination, 69 ff
- substitution instance, 51
- substitution rule, 49
- subterm, 66
- successor, 91
- successor function, 36, 89
- suitable interpretation of  $\varepsilon$ -symbol, 33 ff
- sum, 91
  
- $x$ -suppression, 14
- syllogism rule, 43
- symbol, 2, 10
  - defined, 6
  - function, 2, 10
  - individual, 7, 8, 10
  - predicate, 2, 10
  - primitive, 6
  - separation, 10
- syntactic, 7
- syntax, 9
- system, formal, 2, 38
  
- Tait, W. W., 36, 99
- Tarski, A., 23
- tautological consequence, 17–18, 46
- tautology, 17–18
- tautology rule, 47
- Tautology Theorem, 45 ff, 120
- term, 2, 12
  - null, 6
  - quasi, 11
- $\varepsilon$ -term, 1, 12
- theorem, 4, 38, 85
- transform, 110
- true in model, 2, 22
- truth assignment, 17
- truth function, 16
- truth functionally invalid, 46
- truth value, 16
  
- ultraproduct, 30
- uniqueness condition, 101
- universal closure, 51
- universal formula, 15
- universe, 18
  
- vacuous bondage, 7
- valid, 3, 22
- variable, 2, 7, 10
  - free, 7
  - bound, 7
- variant, 53–54
- vocabulary, 2, 10
  
- Wang, Hao, 99, 106, 107
  
- $x$ -suppression, 14
- Zorn's Lemma, 24