# The Logical Foundations of Mathematics

**WILLIAM S. HATCHER**
Departement de Mathematiques, Universite Laval, Quebec

Foundations & Philosophy of Science & Technology Series

Copyrighted material

# THE LOGICAL FOUNDATIONS OF MATHEMATICS

by

WILLIAM S. HATCHER

*Département de Mathématiques, Université Laval, Québec, Canada*

PERGAMON PRESS

OXFORD · NEW YORK · TORONTO · SYDNEY · PARIS · FRANKFURT

Copyrighted material

| | |
|---|---|
| U.K. | Pergamon Press Ltd., Headington Hill Hall, Oxford OX3 0BW, England |
| U.S.A. | Pergamon Press Inc., Maxwell House, Fairview Park, Elmsford, New York 10523, U.S.A. |
| CANADA | Pergamon Press Canada Ltd., Suite 104, 150 Consumers Road, Willowdale, Ontario M2J 1P9, Canada |
| AUSTRALIA | Pergamon Press (Aust.) Pty. Ltd., P.O. Box 544, Potts Point, N.S.W. 2011, Australia |
| FRANCE | Pergamon Press SARL, 24 rue des Ecoles, 75240 Paris, Cedex 05, France |
| FEDERAL REPUBLIC OF GERMANY | Pergamon Press GmbH, 6242 Kronberg-Taunus, Hammerweg 6, Federal Republic of Germany |

Copyright © 1982 William S. Hatcher

*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, electrostatic, magnetic tape, mechanical, photocopying, recording or otherwise, without permission in writing from the publishers*

First edition 1982

**British Library Cataloguing in Publication Data**

Hatcher, William S

The logical foundations of mathematics. – (Foundations and philosophy of science and technology series).

1. Mathematics – Philosophy
I. Title II. Series
510′.1 QA8.4 80–41253

ISBN 0–08–025800–X

*Printed in Hungary by Franklin Printing House*

Copyrighted material

## Contents





Copyrighted material

*Chapter 1*

# First-order Logic

IN ORDER to understand clearly the formal languages to be presented in later chapters of this book, it will be necessary for the reader to have some knowledge of first-order logic. This chapter serves to furnish the necessary tools. The reader who is already familiar with these questions can easily treat this chapter as a review, though some attention should be given to our particular form of the rules for the predicate calculus, which will be used in the remainder of this study.

## 1.1. The sentential calculus

By a *statement* or *sentence* of some language, we mean an expression of that language which is either true or false in the language. Other expressions of a language may be meaningful without being sentences in this sense. Commands, for example, are meaningful expressions of English, but they are not sentences in our sense. "Go help your brother", "Thou shalt not kill", and "Stop!" are commands in English. Though they are correctly structured English expressions, they do not qualify as sentences in our restrictive definition. It makes little sense to ask, "Is the command to 'stop' true or false?" The reader may consider that by "statement" or "sentence" we shall mean roughly what a grammarian might designate as an "English sentence in the indicative mood".

The *sentential* or *statement calculus* considers certain locutions by which sentences are combined to form more complicated sentences. The basic sentence connectives are: "not", symbolized by "$\sim$"; "and", symbolized by "$\wedge$"; "or", symbolized by "$\vee$"; "if ..., then ---", symbolized by "$\supset$"; and "... if and only if ---", symbolized by "$\equiv$". In a natural language such as English, these locutions undoubtedly vary in meaning according to certain contexts. We shall now proceed to fix their meaning by way of explicit conventions, and this will be our first step toward what is called "formalization". These conventions should not be construed as asserting that the above-mentioned locutions are always used in ordinary discourse in a manner consistent with our conventional meanings. They are rather to be regarded as an explicit statement of how we shall, in fact, agree to use these same locutions. This point has been frequently misunderstood by philosophers of "ordinary language".

In fixing our conventions, we shall be concerned only with the truth or falsity of sentences (as opposed to other aspects of sentences such as meaning or length). For this reason, the logic we obtain is often called *truth-functional*. We express our conventional meanings by way of diagrammatic tables which tell us the truth or falsity of a compound statement (formed by



Copyrighted material

means of one of our connectives) relative to the truth or falsity of its component parts. These tables are called *truth tables.*

For example, the effect of the operation of *negation* ("not") on a given sentence $X$ is given by the following table:

| $X$ | $\sim X$ |
|---|---|
| $T$ | $F$ |
| $F$ | $T$ |

The truth table tells us that if the sentence is true, then its negation is false and if the sentence is false, its negation is true. Notice that all possible cases of truth and falsity have been considered. Thus, we have completely described the (truth-functional) meaning of negation.

Associated with each of our connectives is an *operation,* which is just a function or mapping in the usual mathematical sense of the term. Negation is a mapping from sentences into sentences. Our other connectives are *binary* and thus mappings from pairs of sentences into sentences. As we have done with the truth table for "not", we shall introduce with each truth table the name we give to the operation associated with the particular connective.

The truth table for the *conjunction* (associated with "and") of two sentences is the following:

| $X$ | $Y$ | $X \wedge Y$ |
|---|---|---|
| $T$ | $T$ | $T$ |
| $T$ | $F$ | $F$ |
| $F$ | $T$ | $F$ |
| $F$ | $F$ | $F$ |

Since conjunction is a binary operation, the number of possible cases of truth and falsity is greater than that of the negation operation. Generally speaking, a sentence with $n$ component sentences will give rise to $2^n$ different possibilities of truth and falsity. We shall speak of "truth" and "falsity" as our two *truth values.* A given line of the truth table of a compound sentence $X$ assigns a unique truth value to each component sentence of $X$ and indicates the truth value of $X$ for that assignment of truth values to components. For example, the first line of the truth table for the conjunction tells us that the compound $X \wedge Y$ has the value $T$ for the assignment of values $\langle T, T \rangle$ (i.e. the assignment $T$ to the first conjunct and $T$ to the second).

We have not yet considered truth tables with more than two component sentences since we are still defining our basic connectives. However, we shall be able to iterate our connectives and thus build compound sentences of increasing complexity. We will therefore have truth tables for compounds with any finite number of component sentences. This will be clearer once we have completed the task of defining our basic connectives.

The truth table for the operation of *disjunction* ("or") is given as follows:

| $X$ | $Y$ | $X \vee Y$ |
|---|---|---|
| $T$ | $T$ | $T$ |
| $T$ | $F$ | $T$ |
| $F$ | $T$ | $T$ |
| $F$ | $F$ | $F$ |

Copyrighted material

Notice that the conjunction of $X$ and $Y$ is true when and only when the conjuncts are both true, whereas the disjunction is false when and only when the two disjuncts are false. The relation between these two connectives will become clearer in the light of later examples.

Because we interpret $X \vee Y$ to be true when both disjuncts are true, it is sometimes said that we have adopted the *inclusive* meaning of "or".

The *conditional* ("if . . ., then - - -") of two sentences is defined by the following table:

| $X$ | $Y$ | $X \supset Y$ |
|---|---|---|
| $T$ | $T$ | $T$ |
| $T$ | $F$ | $F$ |
| $F$ | $T$ | $T$ |
| $F$ | $F$ | $T$ |

In a conditional statement $X \supset Y$, we call $X$ the *antecedent* or *hypothesis* and $Y$ the *consequent* or *conclusion.* This table probably departs most radically from ordinary usage. Ordinarily a statement of the form "If $X$, then $Y$" is thought of as asserting $Y$ as true conditional upon the truth of $X$. For this reason, one does not even consider the cases in which $X$ is false. Logicians have extended the common usage by giving truth values in these two further cases. Failure to do so would leave the conditional undefined since we would not have exhausted all possible cases of truth and falsity. We have, therefore, decided to consider the conditional as false only when the antecedent is true and the consequent is false.

Let us consider some examples to motivate our particular choice. If a statement $Y$ is true, then it is true whether anything else is true or false. "If $X$, then $1 = 1$" is true for any statement $X$, true or false, since $1 = 1$. We thus give the conditional the value $T$ in the first and third lines of the table.

As for the fourth line, which involves the falsity of both antecedent and consequent, consider such statements as the following, which occur frequently in mathematical exposition: "For any $x$ whatever, if $x$ is a prime number greater than 2, then $x$ is odd." This statement is true. Now, let $x$ be 6. Both antecedent and consequent are false in the resulting conditional statement "If 6 is a prime number greater than 2, then 6 is odd". Yet we still wish to count the original statement as true (even though this statement is not, strictly speaking, a conditional one). In fact, we desire things to be arranged so that a statement such as the original one is true only when *every* conditional statement obtained by substituting particular values for $x$ is true. But this state of affairs will obtain only if we require conditionals to be true when both the antecedent and consequent are false; hence the fourth line of our truth table.

The reader may feel that, in our treatment, too much weight is being given to mathematical considerations and that a different treatment of the conditional would be possible and even preferable. Many philosophers have indeed refused to accept this treatment and have developed their own theory of the conditional connective. Suffice it to say that experience seems to have shown that our truth-functional treatment of the conditional is entirely adequate for mathematics and this usage is the principal application that we envisage for our logic.

The *biconditional* ("$\equiv$") of two sentences $X$ and $Y$ is true when $X$ and $Y$ agree in truth value:

| $X$ | $Y$ | $X \equiv Y$ |
|---|---|---|
| $T$ | $T$ | $T$ |
| $T$ | $F$ | $F$ |
| $F$ | $T$ | $F$ |
| $F$ | $F$ | $T$ |

Copyrighted material

Now that we have defined our truth-functional connectives, it is clear that any sentence built up from other sentences by means of these connectives will have a well-determined truth value for each assignment of values to its components. To make this more precise, let us introduce several definitions. Let us call a sentence *atomic* if it is not built up from other sentences by means of our sentential connectives. For example, "It is raining today" and "I am sick" are atomic, whereas "It is not raining today" or "If it is raining today, then I am sick" are not atomic. Notice that an atomic sentence $X$ may contain occurrences of some of our sentential connectives, but such occurrences of our connectives will not serve to build $X$ from other *sentences*. Our example of three paragraphs ago is a case in point.

Now clearly we can build a compound sentence by means of our logical connectives by starting with a given number of atomic sentences. Moreover, any such sentence will have a truth table, which will give the truth value of the compound sentence for each assignment of values to its atomic components. For example:

| $[[\sim$ | $X]$ | $\vee$ | $Y]$ | $\supset$ | $Z$ |
|---|---|---|---|---|---|
| $F$ | $T$ | $T$ | $T$ | $T$ | $T$ |
| $F$ | $T$ | $T$ | $T$ | $F$ | $F$ |
| $F$ | $T$ | $F$ | $F$ | $T$ | $T$ |
| $F$ | $T$ | $F$ | $F$ | $T$ | $F$ |
| $T$ | $F$ | $T$ | $T$ | $T$ | $T$ |
| $T$ | $F$ | $T$ | $T$ | $F$ | $F$ |
| $T$ | $F$ | $T$ | $F$ | $T$ | $T$ |
| $T$ | $F$ | $T$ | $F$ | $F$ | $F$ |

Notice here that we have used a more compact way of giving the truth table: we have put the values for the component parts under each atomic component and put the resulting value for each compound statement under the connective that forms the compound. This procedure is simpler than stating each component separately as we did in introducing the connectives.

Involved in the generalized construction of compound sentences from simpler ones is the question of grouping. Where $X$, $Y$, and $Z$ are sentences, the expression $X \wedge Y \supset Z$ is ambiguous as between $[X \wedge Y] \supset Z$ and $X \wedge [Y \supset Z]$. In the vernacular, punctuation serves to indicate the intended grouping. For instance, let $X$ stand for "I am sick", $Y$ for "I will stay home", and $Z$ for "The game is lost". The two translations of these groupings would then be (1) "If I am sick and if I stay home, then the game is lost", and (2) "I am sick, and if I stay home then the game is lost". These two sentences do not have the same meaning.

Often the grouping of compound sentences in the vernacular is not clear or it is left to contextual interpretation. For this reason, let us introduce a convention for the use of brackets; it will remove all possible ambiguity. This convention represents a further step toward formalization.

**Definition 1.** We say that an expression $W$ is *well formed according to the sentential calculus* if $W$ is an atomic sentence or if there are well-formed expressions $X$ and $Y$ such that $W$ is of the form $[\sim X]$, or $[X \wedge Y]$, or $[X \vee Y]$, or $[X \supset Y]$, or $[X \equiv Y]$. We further suppose that atomic sentences contain no brackets. No other expressions are well formed (according to the sentential calculus).

Copyrighted material

This definition is our first example of a *recursive* definition. This is a definition involving an inductive or iterative process (such as building up sentences from simpler ones).

As we have defined them, well-formed expressions are sentences. We further suppose that, from now on, all sentences with which we deal are well formed.

From the way we have defined well-formed expressions, any such expression will have the same number of left and right brackets. We use this fact to define precisely what we mean by the *principal connective* of a sentence. The principal connective is thought of as the last connective used in constructing the sentence from its component parts. More precisely:

**Definition 2.** Given a sentence $X$, let us count the brackets from left to right, counting $+1$ for all left brackets and $-1$ for all right brackets. The sentence connective we reach while on the count of "one" will be the *principal connective.*

*Example.* $[[X \supset Y] \supset Z]$. Counting brackets, we find that we cross the second "$\supset$" on the count of one. This "$\supset$" is the principal connective. Notice that the final count is always zero. We leave it as an exercise to the reader to prove that there is always one and only one principal connective for any nonatomic sentence. (*Hint:* The proof is by mathematical induction on the number of brackets occurring in the sentence.)

**Definition 3.** We say that a sentence is a *tautology* or is *tautologically true* if and only if the truth table of the sentence exhibits only $T$s under its principal connective.

In other words, a tautology is a sentence that has the truth value $T$ for every assignment of values to its atomic components. The reader can verify that $[X \supset X]$, $[[X \vee Y] \vee [\sim X]]$ are tautologies, where $X$ and $Y$ are any sentences.

From now on, we shall relax our convention on brackets by omitting them wherever there is no ambiguity possible. In particular, we shall often omit the outside set of brackets.

**Definition 4.** Given two sentences $X$ and $Y$, we say that $X$ *tautologically implies* $Y$ if the conditional $X \supset Y$ is a tautology.

THEOREM 1. *If $X$ and $Y$ are sentences and if the conditional $X \supset Y$ is a tautology and if, further, $X$ is a tautology, then $Y$ is a tautology.*

*Proof.* Suppose there is some assignment of values which makes $Y$ false. Then, for that assignment of values, the conditional $X \supset Y$ will be false since every assignment of values makes $X$ true. But this is impossible since the conditional $X \supset Y$ is a tautology. Thus, there is no assignment of values making $Y$ false, and $Y$ is tautology. (A complete proof of the above theorem can be given by mathematical induction.)

Theorem 1 embodies what we call a *rule of inference.* A rule of inference is an operation which allows us to pass from certain given sentences to other sentences. We say that the latter are *inferred* from the former. The traditional name of the rule of Theorem 1 is *modus ponens. Modus ponens* allows us to infer $Y$ from $X \supset Y$ and $X$.

We may sometimes refer to tautologically true statements simply as *logically true* statements, meaning "logically true according to the sentential calculus". The reason for the latter qualifi-

Copyrighted material

cation is that we shall subsequently study a class of logical truths that is broader than the tautologies. Tautologies can be thought of as those logical truths which are true strictly by virtue of their structure in terms of our five sentential connectives. Our broader class will include other types of logical operations that we have not yet considered.

**Definition 5.** Two statements will be called *tautologically equivalent* if their biconditional is tautologically true.

*Exercise.* Show that two statements which are built up from the same atomic components $X_1, X_2, \ldots, X_n$ are tautologically equivalent if and only if they have the same truth value for the same given assignment of values to the atomic components.

Notice that two statements may be tautologically equivalent even where they do not have the same atomic components. Thus, $X$ is equivalent to $X\wedge[Y\vee[\sim Y]]$ where $X$ and $Y$ are any sentences.

THEOREM 2. *Let $X$ be any sentence which is a tautology and whose atomic components are sentences $a_1, a_2, \ldots, a_n$. If any sentences whatever are substituted for the atomic sentences $a_i$, where the same sentence is substituted for each occurrence in $X$ of a given $a_i$, the resulting sentence $X'$ is also a tautology.*

*Proof.* A rigorous proof of this is done by induction, but the basic idea can be easily expressed. Since $X$ is a tautology, it is true for every assignment of values to its atomic components. Now consider $X'$. Its atomic components are the atomic components of the sentences $Y_i$ that have been substituted for the $a_i$. Now, for any assignment of values to the atomic components of $X'$, we obtain an assignment of values to the $Y_i$. But this assignment of values to the $Y_i$ is the same as some assignment of values to the $a_i$ of $X$ and thus yields the value $T$ as before.

Theorem 2 tells us that substitution in a tautology always yields a tautology. This again shows us the sense in which our analysis is independent of the meaning of the sentences making up a given compound sentence.

THEOREM 3. *If $A$ is tautologically equivalent to $B$, and if $A$ is replaced by $B$ in some sentence $X$ (at one or more places), then the resulting sentence $X'$ will be tautologically equivalent to $X$.*

*Proof.* The student will prove Theorem 3 as an exercise (*Hint:* Show that the conditional $[[A \equiv B] \supset [X \equiv X']]$ is tautologically true.)

*Exercise* 1. Establish the following equivalences where the variables represent sentences: (a) $[X\vee Y] \equiv [Y\vee X]$; (b) $[X\wedge Y] \equiv [Y\wedge X]$; (c) $[X\vee[Y\vee Z]] \equiv [[X\vee Y]\vee Z]$; (d) $[X\wedge[Y\wedge Z]] \equiv [[X\wedge Y]\wedge Z]$; (e) $[\sim[\sim X]] \equiv X$.

These equivalences establish the commutative and associative laws for disjunction and conjunction with respect to the equivalence relation of tautological equivalence. Furthermore, negation is involutory.

It is well known that we really have a Boolean algebra with respect to tautological equivalence, where negation represents complementation, disjunction represents supremum, and conjunc-

Copyrighted material

tion represents infimum. In other words, the substitutivity of equivalence (Theorem 3) shows that we have a congruence relation, and the quotient algebra is a Boolean algebra. This shows the precise sense in which two operations of conjunction and disjunction are dual to each other. They are dual in the Boolean algebraic sense. The equivalence class of sentences determined by the tautologies will be the maximal element of our Boolean algebra. The minimal or zero element of our Boolean algebra will be determined by the equivalence class of the negations of tautologies. These are the *refutable or tautologically false* sentences. Obviously the tautologically false sentences will be those whose truth table has all *F*s under the principal connective. They are dual to tautologies.

In view of the associative and commutative laws of conjunction and disjunction, we can speak unambiguously about the conjunction (or disjunction) of any finite number of sentences without regard to grouping or order. Where several different connectives are involved, grouping and order again become relevant.

*Exercise* 2. By an *argument* in English, we mean a finite collection of statements called *premisses* followed by a statement called the *conclusion*.† We say that an argument is *valid* if the conjunction of the premisses logically implies the conclusion. In each of the following arguments, test validity by using letters to represent atomic sentences, forming the conditional with the conjunction of the premisses as antecedent and the conclusion as consequent, and testing for tautology (thus, a tautological implication).

(a) If the team wins, then everyone is happy. If the team does not win, then the coach loses his job. If everyone is happy, then the coach gets a raise in pay. Everyone is happy if and only if the team makes money. Hence, either the coach gets a raise in pay, or the team does not win and the coach loses his job.

(*Hint:* When dealing with an implication involving many variables, it becomes prohibitive and unnecessary to give all possible cases. Observe that by the truth table for the conditional, a conditional statement is false only in one case, when antecedent is true and consequent is false. The best method, then, is a truth-value analysis, working from the outside in. Try to make the antecedent true and the consequent false. This gives necessary conditions on the variables. If these are satisfied, then we have a counterexample and implication does not hold. Otherwise, we establish the impossibility of the false case and implication holds.)

(b) If the sun shines today, then I am glad. If a high-pressure zone moves in, then the sun will shine. Either my friend will visit and the sun will shine, or a high-pressure zone moves in. Hence, I am glad.

(c) If I study, then I shall pass the test. If I pass the test, then I shall be surprised. If I fail the test, it is because the teacher is too difficult and grades unfairly. If I fail the test, then I do not blame the teacher. Thus, either I shall not be surprised, or I shall fail the test and blame the teacher.

*Exercise* 3. Verify that the following are tautologies where the variables represent sentences: (a) $[X \equiv Y] \equiv [[Y \supset Y] \wedge [Y \supset X]]$; (b) $[X \wedge Y] \equiv [\sim[[\sim X] \vee [\sim Y]]]$; (c) $[X \supset Y] \equiv [[\sim X] \vee Y]$.

† In using the double s in "premiss", we follow Church [3].



Copyrighted material

The equivalences of Exercise 3 show that we can express some of our basic connectives in terms of others. In particular, we can see that negation and disjunction suffice to define the others. A more precise understanding of what such definability of some connectives in terms of others really means involves the notion of a truth function which we shall now briefly sketch.

Given a set $E$ with exactly two elements, say $E = \{T, F\}$, the set of our two truth values, a *truth function* is defined as a function from $E^n$ into $E$, where $E^n$ is the set of all $n$-tuples of elements of $E$. Let some compound sentence $X$ be given, and suppose that $X$ has exactly $n$ atomic components. Then, for a given ordering of the atomic components of $X$ (there are $n!$ such orderings) there is a uniquely defined truth function $f$ from $E^n$ to $E$ associated with $X$. For a given $n$-tuple $\langle a_1, \ldots, a_n\rangle$ of truth values, $f(\langle a_1, \ldots, a_n\rangle)$ is determined as follows: We assign the truth value $a_1$ to the first atomic component in our ordering, $a_2$ to the second, and generally $a_i$ to the $i$th atomic component. When the assignment of truth values is complete, we have simply a line of the truth table of $X$ and $f(\langle a_1, \ldots, a_n\rangle)$ is defined to be the truth value $a_{n+1}$ of $X$ for the given assignment of values to atomic components.

Suppose now that we agree on some fixed ordering of atomic components for all sentences $X$, say the order of first occurrence from left to right. Then every sentence $X$ has a uniquely associated truth function $f$ as defined above. We call $f$ the *truth function expressed by X*. Now, we say that a set $K$ of sentential connectives is *adequate* if and only if it is possible to express every truth function by means of some compound sentence $X$ constructed from atomic sentences using only connectives from $K$.

It is easy to see that the set $K = \{\sim, \vee, \wedge\}$ is adequate, for let some truth function $f$ from $E^n$ to $E$ be given. Let $n$ atomic sentences $X_1, \ldots, X_n$ be chosen. We now construct a sentence in the following way: We consider only those $n$-tuples of truth values $\langle a_1, \ldots, a_n\rangle$ for which

$$f(\langle a_1, \ldots, a_n\rangle) = T.$$

Suppose first that there is at least one $n$-tuple whose image under $f$ is $T$. For each such $n$-tuple we form the following conjunction $X_1^* \wedge \ldots \wedge X_n^*$, where $X_i^*$ is $X$ if $a_i$ is $T$ and $X_i^*$ is $[\sim X_i]$ if $a_i$ is $F$. This compound obviously yields the value $T$ when the values $a_i$ are assigned to the $X_i$. We now take the disjunction $W_1 \vee \ldots \vee W_m$ of all the conjunctions $W_j$ of the $X_i^*$s. This is a compound sentence whose expressed truth function is obviously $f$. On the other hand, if $f\langle a_1, \ldots, a_n\rangle = F$ for all $n$-tuples of truth values, then any contradictory sentence with $n$ atomic components will express $f$. The sentence $X_1 \wedge [\sim X_1] \wedge W$, where $W$ is the conjunction of the rest of the $X_i$, will do fine.

If two sentences have the same atomic components in the same left-to-right order of first occurrence, then they will be tautologically equivalent if and only if they express the same truth function. The equivalences of Exercise 3 thus show that we can express all truth functions by means of $\sim$ and $\vee$, since $\wedge$ is definable by means of $\sim$ and $\vee$, and we have shown $K = \{\sim, \vee, \wedge\}$ to be adequate. In other words, the set $\{\sim, \vee\}$ is adequate.

*Exercise* 4. Show that each of $\{\sim, \supset\}$ and $\{\sim, \wedge\}$ is adequate. Show that $\{\sim\}$, $\{\wedge\}$, and $\{\supset\}$ are not adequate.

Although it may not appear possible at first, there are single binary connectives which are alone adequate to express all truth functions. If we define a new connective "$|$" by the table

Copyrighted material

| $X$ | $Y$ | $[X \mid Y]$ |
|---|---|---|
| $T$ | $T$ | $F$ |
| $T$ | $F$ | $T$ |
| $F$ | $T$ | $T$ |
| $F$ | $F$ | $T$ |

then $[\sim X]$ is definable by $[X \mid X]$, and $[X \vee Y]$ is definable as

$$[[X \mid X] \mid [Y \mid Y]].$$

Since $\{\sim, \vee\}$ is adequate, the adequacy of $\{\mid\}$ follows immediately.

The binary connective "$\downarrow$" defined by

| $X$ | $Y$ | $[X \downarrow Y]$ |
|---|---|---|
| $T$ | $T$ | $F$ |
| $T$ | $F$ | $F$ |
| $F$ | $T$ | $F$ |
| $F$ | $F$ | $T$ |

is also adequate as the reader can check by defining $\sim$ and $\vee$ (or $\sim$ and $\wedge$) purely in terms of $\downarrow$. It is easy to show that $\mid$ and $\downarrow$ are the only two adequate binary connectives and we leave this as an exercise to the reader. Church [3] contains a detailed discussion of the definability and expressibility of various connectives for the sentential calculus, as well as detailed references to early work on the question.

## 1.2. Formalization

In the preceding section, we have considered the English language and subjected it to a certain logical analysis having mathematical overtones. The reader may have noticed that, as we progressed, we moved further and further away from the necessity of giving examples involving specific sentences of English such as "I am sick". It was sufficient simply to consider certain forms involving only letters representing sentences together with our logical connectives. The reason for this should be clear if one reflects on our point of departure. Because our analysis was truth-functional, we were concerned only with sentences as objects capable of being considered true or false. Not much else, not even meaning, except on a very rudimentary level, was relevant to our analysis.

The whole process is typical of mathematics. One starts with a particular concrete situation, and then subjects it to an analysis which ignores some aspects while considering others important. In the empirical sciences, this process leads eventually to certain "laws" or general statements of relationship, which are afterwards capable of various degrees of verification. In mathematics, the process leads to the definition of abstract structures independent of any concrete situation. These abstract structures are then studied in their own right. Because they supposedly carry certain important features shared by many concrete cases, applications of these abstract structures are often found (sometimes in surprising ways), and one gains in economy by finding general results true, within a single given structure, for many different cases. Finally, any parti-



Copyrighted material

cular structure having all of the properties of a given abstract structure is called a "model" of the abstract structure.

Mathematical logic has given rise to the study of abstract structures called *formal systems* or *formal languages*. We define these as follows:

**Definition 1.** A *formal system F* is a quadruple $\langle A, S, P, R\rangle$ of sets which satisfies the following conditions: $A$ is a nonempty set called the *alphabet* of $F$ and whose elements are called *signs*, $S$ is a nonempty subset of the set $\mathscr{M}(A)$ of all finite sequences of elements of $A$, $P$ is a subset of $S$, and $R$ is a set of finitary relations over $S$ (a finitary relation over $S$ being some subset of $S^n$, where $n \geqslant 1$ is some positive integer). Elements of $\mathscr{M}(A)$ are called *expressions* of $F$, elements of $S$ are called *well-formed formulas*, abbreviated as *wffs*, or just *formulas* of $F$, elements of $P$ are called *axioms* of $F$, and elements of $R$ are called *primitive rules of inference* or just *rules of inference* of $F$. By the *degree* of a rule of inference $r$ in $R$ we mean the integer $n$ such that $r$ is a subset of $S^n$. Finally, the triple $\langle A, \mathscr{M}(A), S\rangle$ is called the *language* of $F$, and the pair $\langle P, R\rangle$ is called the *primitive deductive structure* of $F$ *relative to* the language of $F$.

The above definition is a bit more general than we really need. In most cases, the alphabet will be countable. Moreover, we shall often require that the alphabet as well as the sets $S$ and $P$ be "effective" or "decidable" in the sense that there is some procedure allowing us to decide whether a given object is or is not a sign of the system, or whether a given expression is or is not a wff, or whether a given wff is or is not an axiom. We may similarly require that a given rule of inference $R_n$ be effective in the sense that, for any $n$-tuple $\langle x_1, \ldots, x_n\rangle$ of elements of $S$, there is some effective procedure which allows us to determine whether or not the $n$-tuple is in the relation $R_n$. Also, the set $R$ of rules of inference may usually be assumed finite.

The intuitive idea underlying our informal notion of "effectiveness" in the preceding paragraph is that there exists a set of rules or operations which furnish a mechanical test allowing us to decide a question in a finite length of time. Elementary arithmetic calculation is a good example of an effective procedure. Our bracket-counting method for determining the principal connective of a sentence is another. In Chapter 6, we give a precise mathematical definition of the notion of effectiveness by means of the so-called recursive functions.

The primitive deductive structure of a formal system $F$ allows us to develop a fairly rich *derived* deductive structure which we now describe.

**Definition 2.** Given a formal system $F$ and a set $X$ of wffs of $F$, we say that a wff $y$ is *immediately inferred* from the set $X$ of wffs if there exists a primitive rule of inference $R_n$ of degree $n$ and a finite sequence $b_1, \ldots, b_{n-1}$ of elements of $X$ such that the relation $R_n(b_1, \ldots, b_{n-1}, y)$ holds (i.e. the $n$-tuple $\langle b_1, \ldots, b_{n-1}, y\rangle$ is an element of the relation $R_n$).

**Definition 3.** Given a set $X \subset S$ of wffs, we say that a wff $y \in S$ is *deducible from the hypotheses X* if there exists a finite sequence $b_1, \ldots, b_n$ of wffs such that $y$ is $b_n$ and such that every member of the sequence is either (i) an element of $X$, or (ii) an axiom of $F$, or (iii) immediately inferred from a set of prior members of the sequence by some primitive rule of inference. The finite sequence $b_1, \ldots, b_n$ itself is called a *formal proof* (or *formal deduction*) *from the hypotheses X*. If $X$ is empty (thus only axioms are used in the deduction together with rules of inference), then the sequence $b_1, \ldots, b_n$ is called simply a *formal deduction* or *formal proof in F*. In this case, $y$ is said to be a *provable* wff or a *theorem* of $F$. A formal proof $b_1, \ldots, b_n$ is a *proof of* $b_n$.

Copyrighted material

We denote the set of all wffs which can be deduced from a given set of hypotheses $X$ by $K(X)$, the set of consequences of $X$. The theorems are thus the set $K(P)$ (which is the same as the set $K(\Lambda)$).† The theorems are those wffs which can be obtained as the last line of a formal deduction. Notice that any axiom is a one-line proof of itself, and every formal proof must begin with an axiom.

A formal deduction in a formal system is an abstract analogue of our usual informal notion of a proof in which we "prove" an assertion by showing that it "follows from" other previously proved statements by successive application of logical laws or principles. These previously proved statements are ultimately based on axioms, and our logical principles or laws (such as the rule of *modus ponens*) allow us to pass from a given statement or statements to other statements.

We write $X \vdash y$ for $y \in K(X)$. When $X$ is empty, we write simply $\vdash y$ instead of $\Lambda \vdash y$. Since $K(\Lambda)$ is the set of theorems, $\vdash y$ is a short way of asserting that "$y$ is a theorem" or "$y$ is provable". Sometimes we talk about provability with respect to different systems $F$ and $F'$, and we indicate this by writing "$\vdash_F y$", "$\vdash_{F'} y$", and so on. Also, if $\{z\} \vdash y$, for some one-element set $\{z\}$, we prefer to write simply "$z \vdash y$". This latter is properly read "$y$ is deducible from the hypothesis $z$". Finally, we may sometimes wish to emphasize the fact that certain wffs occur as hypotheses and so we display them by writing $X, z_1, z_2, \ldots, z_n \vdash y$ to mean $X \cup \{z_1, z_2, \ldots, z_n\} \vdash y$.

Let us note in passing that a formal deduction from hypotheses $X$ in a system $F$ is the same thing as a formal deduction in the system which is the same as $F$ except that its set of axioms is $P_F \cup X$. It is useful to keep this simple fact in mind since it means that some metatheorems about formal deduction have greater generality than may appear at first glance.

Given a formal proof $b_1, \ldots, b_n$ from the hypotheses $X$ in a formal system $F$, each member of the sequence is *justified* in that it satisfies at least one of the conditions (i) to (iii) of Definition 3. It is quite possible that a given member of the proof could satisfy more than one of these conditions, even all three of them. By a *justification* for a member $b_i$ of the proof, we mean any one of the three conditions (i) to (iii) which is true of $b_i$. Later on, when we deal with the languages known as first-order theories, we shall countenance a slightly different notion of proof, one in which it is necessary that every line in the proof be accompanied by an explicit justification.

As has already been mentioned, we generally suppose that the set $S$ of wffs and the set $P$ of axioms are decidable sets, and that our rules of inference are effective. These requirements have the result that the notion of formal proof is effective; given any finite sequence of wffs, we can decide whether or not it is a formal proof (from no hypotheses). It will not follow, however, that the set $K(\Lambda)$ of theorems is decidable even though the notion of deduction from no hypotheses is. A system $F$ for which the theorems $K(\Lambda)$ are a decidable set is called a *decidable* system. Most interesting systems will not be decidable, though there are some exceptions.

In practically all systems which logicians consider, the set of wffs is decidable and the rules of inference are effective. However, certain branches of logic, such as model theory, do consider systems whose set $P$ of axioms is not decidable (and even some for which the set $S$ of wffs is not decidable). A system $F$ whose set of axioms is decidable is called *axiomatized*.

† "$\Lambda$" is our symbol for the null set.

Copyrighted material

A system $F$ for which there is a decidable set of axioms yielding the same theorems (keeping the rules and wffs the same) is said to be *axiomatizable*. There are many interesting axiomatizable systems.

*Exercise* 1. Prove our assertion that $K(P) = K(\Lambda)$.

*Exercise* 2. Show that the $K$ function has the following properties in any formal system $F$: $X \subset K(X)$; $K(K(X)) = K(X)$; if $X \subset Y$, then $K(X) \subset K(Y)$; $K(X) = \bigcup_{Y \in \mathcal{F}} K(Y)$ where $\mathcal{F}$ is the class of finite subsets of $X$.

*Exercise* 3. Let two wffs, $x$ and $y$, be *deductively equivalent* if each is deducible from the other. Show that any two theorems are deductively equivalent for any formal system $F$.

*Exercise* 4. Let some formal system $F$ be given and let $X = a_1, a_2, \ldots, a_n$ and $Y = b_1, b_2, \ldots, b_m$ each be sequences of wffs which constitute formal proofs in $F$. Prove that the juxtaposition $XY$ of these two deductions is again a deduction. Explain how this fact justifies the usual practice in mathematics of citing previously proved theorems, as well as axioms, in a proof.

Obviously, a provable wff of a formal system $F$ will have an infinite number of different proofs. This is because any given proof of a wff $b$ can be arbitrarily extended by adding new lines which are justified but which are not essential to obtaining $b$. For example, we can uselessly repeat hypotheses, add extraneous axioms and inferences obtained from them, or even repeat $b$ itself any number of times. It is useful to make all of this precise by defining clearly when a formula in a deduction *depends* on another formula. Intuitively, a given formula will depend on another if the other formula has been used in obtaining the given formula.

**Definition 4.** Let $b_1, \ldots, b_n$ be a formal deduction from hypotheses $X$ in some formal system $F$. Then we say that an occurrence $b_i$ of a formula *depends on* an occurrence $b_j$ of a formula, $j \leq i$, if $i = j$ or else if $j < i$ and $b_i$ is immediately inferred from a set of prior formulas at least one of whose occurrences depends on $b_j$. We say that a formula $y$ occurring in the deduction *depends on* a formula $z$ occurring in the deduction if at least one occurrence of $y$ depends on an occurrence of $z$.

Notice that every formula occurring in the deduction depends on itself, and no formula occurring in the deduction depends on any formula not occurring in the deduction.

Using these concepts, we now prove:

THEOREM 1. *If $b_1, b_2, \ldots, b_n$ is a formal deduction from hypotheses $X$, in a formal system $F$, then there exists a subsequence $b_{i_1}, \ldots, b_{i_k}$ which is a formal deduction in $F$ of $b_n$ from hypotheses $X$ and such that $b_{i_k}$ depends on every occurrence of every formula in the new deduction.*

*Proof.* We prove this by induction on the length $n$ of the original deduction. We assume the proposition is true for all deductions of length less than $n$ and consider the case of length $n$. If $b_n$ is an hypothesis or an axiom, then the sequence of length 1 consisting of the formula $b_n$

Copyrighted material

alone is a proof from the hypotheses $X$ of the formula $b_n$. Moreover, this is a subsequence of the appropriate kind, chosen by putting $i_1 = n$.

Now, suppose $b_n$ is immediately inferred from prior wffs $b_{j_1}, \ldots, b_{j_m}$ in order of occurrence in the deduction. Since each of these wffs is the last line of a formal deduction from hypotheses $X$ and of length less than $n$, we can apply the induction hypothesis to each of them and obtain subsequences of our original deduction which are proofs of each of them consisting only of formulas on which they each depend. We now choose the smallest subsequence $Y$ containing all of these (which will, in fact, just be their set-theoretic union). $Y$ will be a formal deduction from hypotheses $X$ because each line is justified (if it was justified as a part of one of the subsequences proving one of the $b_{j_s}$, then it remains justified as a part of the union of all of these subsequences). Moreover, $b_{j_m}$ is the last line of the formal deduction $Y$ in which all the $b_{j_s}$ appear. Now, applying the rule originally used to infer $b_n$, we obtain $b_n$ as the next line of a formal deduction $Yb_n$. $Yb_n$ is a subsequence of the original deduction as is clear. Finally, $b_n$ depends on every member of $Yb_n$ since it depends on each of $b_{j_1}, \ldots, b_{j_m}$, and every other occurrence of a formula in $Y$ is depended upon by at least one of the $b_{j_s}$. This completes the proof of the theorem.

Several consequences of Theorem 1 deserve to be stated explicitly as corollaries.

COROLLARY 1. *In any formal system F, every theorem has a proof involving only formulas on which it depends.*

*Proof.* Any theorem has a formal deduction and hence, by Theorem 1, a formal deduction involving only formulas on which it depends.

COROLLARY 2. *In any formal system F, if* $y \vdash x$ *and if* $x$ *does not depend on* $y$*, then* $\vdash x$.

*Proof.* Since $y \vdash x$, there is a proof of $x$ from the hypothesis $y$ involving only formulas on which $x$ depends. Since $x$ does not depend on $y$, $y$ does not appear in the new proof. This new deduction thus involves only axioms and rules of inference and is therefore a proof of $x$ from no hypotheses.

We shall sometimes speak of the *interpretation* of a formal system $F$. This notion will be made quite precise for a large class of formal systems, the so-called first-order systems, which will be treated later in this chapter. For the moment we will not attempt a precise definition, but we will have in mind some language whose sentences can be interpreted as wffs of $F$ and whose true sentences (or some of whose true sentences) can be interpreted as theorems of $F$. By way of example, we will now formulate a formal system whose intuitive interpretation will be precisely the statement calculus of Section 1.

## 1.3. The statement calculus as a formal system

We define a formal system **P** whose alphabet consists of the signs "[" ,"]", "*" (called *star*), "~", "∨", and the small italic letter "*x*". An expression of **P** is, of course, any finite sequence of occurrences of these signs. By a *statement letter* of **P**, we mean the letter "*x*"

Copyrighted material

followed by any nonzero finite number of occurrences of star. $x_n$ will stand for the letter "$x$" followed by $n$ occurrences of star. Thus, "$x_5$" stands for "$x^{*****}$", for example.

We define recursively the set $S$ of wffs of **P**: (i) Any statement letter is a wff. (ii) If $X$ and $Y$ are wffs, then expressions of the form $[\sim X]$ and $[X \vee Y]$ are wffs. (iii) $S$ has no other members except as given by (i) and (ii).

The reader will notice that the wffs, as we have defined them, are completely analogous to the informal wffs of Section 1.1, as we here simply replace the informal notion of "atomic sentence" by the formal notion of a statement letter. The introduction of the star into our formal system is simply a device to enable us to obtain a countably infinite number of distinct statement letters from a finite alphabet. There is no particular virtue in this and we shall countenance countably infinite alphabets many times in this book.

The reader will also notice that we have not included some of our basic connectives as elements of our alphabet. This is because we can economize by *defining* some connectives in terms of others. The exact meaning of this sort of definition is best illustrated by an example.

**Definition 1.** $[X \supset Y]$ for $[[\sim X] \vee Y]$.

In this definition the letters "$X$" and "$Y$" are not signs of our system but rather dummy letters representing arbitrary wffs. The sign "$\supset$" is not part of our system either, and it is not intended that our definition introduces the sign into the system, but rather that, in every case, the expression on the left is an abbreviation for the expression on the right.

More precisely, the meaning of such a definition is as follows: When any two wffs of **P** are substituted for the letters "$X$" and "$Y$" in the above two forms of Definition 1, then the expression resulting from substitution in the left-hand form is an abbreviation of the expression resulting from the same substitution in the right-hand form. The expression resulting from such substitution in the right-hand form will be a wff of **P** while the expression resulting from substitution in the left-hand form will not. Such forms, involving letters together with brackets and other special signs, are called *schemes*. Particular expressions resulting from substitution of formal expressions for the letters of a scheme are called *instances* of the scheme. Thus, in the foregoing example, an instance of the left-hand scheme arising from substitution of wffs of **P** for the letters "$X$" and "$Y$" is an abbreviation of the *corresponding* instance (resulting from the same substitution for the letters "$X$" and "$Y$") of the right-hand scheme.

The distinction between *use* and *mention* of signs is highly important when dealing with formal systems. The alphabet of a given formal system may contain signs that are used in ordinary English. In order to avoid confusion, we must carefully distinguish between the roles such signs play within the formal system and the informal use made of them in the vernacular. We speak of the formal system we are studying as the *object language*, and the language, such as English, we use in studying the formal system is called the *metalanguage*. In the system **P**, we have, for instance, the sign "$x$", which is part of our formal system but which also has a usage in ordinary English, our metalanguage. Our abbreviative definitions of certain expressions of a formal system are technical parts of the metalanguage and not part of the formal system itself. To avoid confusion, logicians make the following rule: In order to talk about an object (to *mention* it) we must *use* a name of the object. Where the objects of a discourse are nonlinguistic, there is little danger of confusion. We would not, for instance, use New York

Copyrighted material

in order to mention New York. But where the objects are themselves linguistic ones (i.e. signs) we must be more careful.

The introduction of such metalinguistic definitions is practically necessary for any sort of manipulation of a formal language. Because we deal so extensively with such abbreviations, we speak of the original, formal notation as the *primitive* notation, and the abbreviated notation as *defined*. Thus, in the case of the formalization of **P** presently at hand, the symbols "$\sim$" and "$\vee$" are primitive while "$\supset$" is defined.

We often form the name of a linguistic object (a sign) by putting it into quotation marks, but it is possible to use signs as names of themselves. For example, we can say that "$x$" is a sign of **P**. Here, ' "$x$" ' is being used as a name for the sign which is the italic twenty-fourth letter of the Latin alphabet. But we often write that "$x$ is a sign of **P**", thus using $x$ as a name of itself. This is called an *autonomous* (self-naming) use of a sign. We can also use other symbols explicitly designated as names for our signs if it suits our purpose.

Similarly, expressions of a formal system, formed by juxtaposing the signs of the alphabet (writing them one after the other) must also be distinguished from expressions of English. We can form the names of formal expressions by juxtaposing in the metalanguage the names for the signs making up the finite sequence in the formal system. Again, we shall often use formal expressions as names of themselves.

There should be no confusion in this autonomous use of signs and expressions of a formal system, since those expressions and signs which are part of the system will always be explicitly designated for every formal system with which we deal in this study. Autonomous use avoids the myriad quotation marks which result from rigorous adherence to the use of quotation marks for name-forming.

Names, as we have here understood them, are simply *constants* in the sense commonly understood in mathematics. A constant is a symbol (linguistic object) which names or designates a particular object. A *variable* is a symbol which is thought of not as designating a particular object but rather as designating ambiguously any one of a given collection of objects. The collection of objects thus associated with a given variable is called the *domain of values* of the variable. It is, of course, we who decide what the domain of values of a given variable is, either by explicit designation or through some convention or contextual understanding. A given variable may obviously have different domains in different contexts, but its domain must be fixed and unambiguous in any particular context.

When substitutions are made for variables, it is usually understood that constants, names of elements of the domain of values, are actually substituted for the variables, rather than the values themselves. For example, in the phrase "$x$ is prime", where the domain of $x$ is the natural numbers, we can substitute "7" for "$x$" and obtain the sentence "7 is prime". We did not substitute the number seven (which is an abstract entity) but the numeral "7" for $x$.

However, when we are talking about linguistic entities, it does become possible to directly substitute the thing about which we are talking (a value) for another symbol. Thus, in dealing with formal systems, we can make use of dummy letters like the "$X$" and "$Y$" in the schemes of Definition 1. These dummy letters are not variables in the sense we have just defined, but rather *stand for* arbitrary wffs, and we can substitute wffs directly for them when our wffs are linguistic objects as is the case for our system **P**. (There is nothing in our definition of a formal system which requires that the set $S$ of wffs must consist of linguistic objects.) Of course, if we want to consider such dummy letters as variables, we are free to do so, for we

Copyrighted material

can regard the wffs as names of themselves and thus as constants which can be substituted for variables. In this case, a particular instance of a scheme is technically a name of a wff (itself) rather than a wff.

A variable in the metalanguage whose domain of values consists of signs or expressions of an object language is called a *metavariable* or *syntactical variable.* For example, our use of the letters "$X$" and "$Y$" in our recursive definition of the wffs of **P** is as metavariables rather than as dummy letters. We will not usually bother to distinguish explicitly between the use of letters of the metalanguage as metavariables and as dummy letters, since in the final analysis, the main difference in the two uses lies more in the way we regard what is going on than in what is going on. What we are doing is talking about formal expressions by means of letters and schemes, and we are allowing substitution of formal expressions for the letters of a scheme. In every case the class of formal expressions will be clearly designated. Hence, the main distinction resides in whether we regard a particular instance of a scheme as the autonym of a wff or as a wff, or whether, in the case of an abbreviative expression, we regard an instance of a scheme as standing for the wff it abbreviates or as naming it.

In any case, juxtaposition of letters in the metalanguage with other special signs will always represent the operation of juxtaposing the formal expressions they represent. Also, as follows from our discussion of abbreviations, we prefer to regard an abbreviation of a formal expression, such as "$[x^* \supset x^{**}]$" for "$[[\sim x^*] \vee x^{**}]$", as a technical part of the metalanguage which stands for the formal expression it abbreviates, rather than naming it. All these semantic matters will not concern us much in this book and we enter into a brief discussion here only to dispose of the matter for the rest of our study.

In the spirit of this last remark, let us point out one further complication that may have already occurred to the semantically precocious reader: When we speak of a *sign* such as the star "*", we cannot really mean the particular blob of ink on the particular part of this page. Such an ink blob is rather a *token* of the sign. The sign itself is the equivalence class of ink blobs under the relation "sameness of shape". A token is thus a representative of the equivalence class (i.e. the sign) in the usual mathematical sense. Similarly, sequences of signs are represented by linear strings of tokens.

We now return to our consideration of **P** and state several more definitions preparatory to designating its axioms.

**Definition 2.** $[X \wedge Y]$ for $[\sim[[\sim X] \vee [\sim Y]]]$.

**Definition 3.** $[X \equiv Y]$ for $[[X \supset Y] \wedge [Y \supset X]]$.

The intuitive justification for these definitions is clear in the light of Exercise 2 following Theorem 3 of Section 1 of this chapter.

We now use schemes and dummy letters to describe the set $P$ of axioms of **P**. $P$ consists of all wffs that are instances of the following schemes:

$$[[X \vee X] \supset X]; \quad [X \supset [X \vee X]]; \quad [[X \vee Y] \supset [Y \vee X]];$$
$$[[X \supset Z] \supset [[Y \vee X] \supset [Y \vee Z]]]$$

where $X$, $Y$, and $Z$ stand for any wffs of **P**.

Notice that our axiom set is infinite.

Copyrighted material

The set $R$ of rules of inference of **P** consists of one relation of degree 3, which is as follows: $R_3(X, Y, Z)$ if and only if there are wffs $A$ and $B$ such that $X$ is $[A \supset B]$, $Y$ is $A$, and $Z$ is $B$. This is the formal analogue of *modus ponens*, and we will apply the name *modus ponens* to it.

This completes the description of **P**.

*Exercise.* Determine which, if any, of the following are instances of our axiom schemes:

$$[[[x_1 \supset x_2] \vee [x_1 \supset x_2]] \supset [x_1 \supset x_2]];$$

$$[x_1 \supset [x_1 \vee x_2]]; \qquad [[x_1 \supset x_1] \supset [[x_1 \vee x_1] \supset [x_1 \vee x_1]]].$$

We have already indicated the analogy between statement letters of our formal system and atomic sentences of English. It is quite clear that we can define the *principal connective* of a wff and apply the truth-table method to wffs in a purely mathematical way to determine which of our wffs are *tautologies*. Notice that any instance of one of our axiom schemes will be a tautology. Furthermore, we can prove that *modus ponens* preserves the property of being a tautology. (The proof of this latter fact is essentially the same as that given for statements of English in Theorem 1 of Section 1.) These two facts immediately give us the result that all theorems are tautologies since our theorems are, by definition, obtained from axioms by successive applications of *modus ponens*. The rigorous proof of this is by induction in the metalanguage.

To illustrate formal deduction in $P$, we prove that $\vdash [x_1 \supset x_1]$. Recall that $[X \supset Y]$ is $[[\sim X] \vee Y]$ by definition.

$$[[x_1 \vee x_1] \supset x_1] \supset [[[\sim x_1] \vee [x_1 \vee x_1]] \supset [[\sim x_1] \vee x_1]]$$
$$[[x_1 \vee x_1] \supset x_1]$$
$$[[[\sim x_1] \vee [x_1 \vee x_1]] \supset [[\sim x_1] \vee x_1]]$$
$$[[\sim x_1] \vee [x_1 \vee x_1]]$$
$$[[\sim x_1] \vee x_1]$$

The first line of this deduction is an axiom of the last type listed in our axiom schemes. The second line is an axiom of the first type. The third line is obtained from the first two by *modus ponens*. The fourth line is an axiom of the second type (recall our defining abbreviations) and the fifth line is obtained from the third and fourth by *modus ponens*.

If we intended to engage in considerable formal deduction in **P**, then we would use many techniques to shorten proofs and render them more readable. This will be done for formal proofs of first-order theories after they are introduced. The reason why we do not do so for **P** will be clarified shortly.

We have the notion of deduction within the object language. This is the purely formal, mathematical process of deduction given in Definition 3 of Section 1.2. But we also have a notion of deduction within the metalanguage, which is the usual informal mathematical notion of deduction. Since deduction within the object language is a precisely defined mathematical operation, we can study its properties just as we can study the properties of any mathematical system. We can prove theorems about the operations within the formal system. These theorems about the formal system are called *metatheorems*. They are carried through in our intuitive logic, which underlies our thinking about mathematical structure.

Copyrighted material

The fact that all the theorems of our system **P** are tautologies, the proof of which we sketched several paragraphs ago, is a metatheorem about the system **P**. It is a theorem in the metalanguage about the set $K(P)$ of theorems of **P**.

It is possible that a given object language, about which we are proving metatheorems, has an interpretation as a part of English and that many of the statements we make about the object language can, in this sense, be made "within" the formal language itself; that is, within that part of English which is the interpretation of the formal language. For example, it is possible to explain, within English, the grammar of English itself. In such a case, it could happen that some theorems and metatheorems coincide. In other words, some metatheorems about the object language might be among the statements of the metalanguage that are also part of the interpretation of the object language. This would all depend on how we defined the interpretation of our object language in the first place.

It will naturally occur to the reader that it would be possible to envisage formalizing mathematically the metalanguage itself and studying its own internal structure. This is certainly true, but such a formalization could be done only within a meta-metalanguage, for we must *use* some language to communicate. Of course, it is conceivable that object language and metalanguage may be the same. Such is the case found in our example of explaining the grammar of English within English. However, Tarski has shown, by reasoning too involved for inclusion here, that in most cases we can avoid certain contradictions, which arise from the circularity of speaking about a language within the language itself, only when the metalanguage si strictly stronger than the object language (see Tarski [1], pp. 152 ff.). We delay further discussion of these delicate questions to Chapter 6 where we shall have the tools necessary to engage in a more precise analysis.

In closing this discussion, let us return once again to a consideration of our language **P**. Our observation that all the theorems of **P** are tautologies leads naturally to the question of whether or not the converse is true: are all tautologies theorems of **P**? The answer is "yes", but we will not give the details of the proof, which can be found in any standard work on mathematical logic such as Church [3] or Mendelson [1].

This second metatheorem, that all tautologies are theorems of **P**, is known as the *completeness theorem* for **P**. Logicians have extensively studied many different, partial (incomplete) versions of the statement calculus. Church [3] contains a long discussion of these questions.

Notice that, since the theorems of **P** are precisely the tautologies, we could have designated the tautologies as axioms (we still would have an effective test for the set of axioms) and let the set $R$ of rules of inference be empty. This form is often given to **P** and we will use it extensively in the present study.

Now it is clear why we are not very eager to engage in protracted formal deductions in **P**. It is because the theorems of **P** are precisely the tautologies and it is easier to prove that a given wff is a tautology than to find a formal deduction for it. **P** is an example of a decidable formal system. Deduction is, in principle, unnecessary in any such decidable system (though the test for theorem determination may be much more complicated than the truth-table method).

Copyrighted material

## 1.4. First-order theories

Our goal in this section will be to describe a whole class of formal languages known as *first-order theories* or *first-order systems.* The importance of these languages is that they have a well-defined structure which is relatively simple while being adequate for a surprisingly large number of purposes.

Since we are describing not one particular language but a class of languages, we cannot specify the exact alphabet for each such language. Rather we shall describe a list of signs from which the particular alphabet of any given first-order theory will have to be chosen.

By an *individual variable*, we mean the small italic "$x$" with a positive integral numeral subscript. Examples are "$x_1$", "$x_2$", and so on.

By an *individual constant letter*, or simply *constant letter*, we mean the small italic "$a$" followed by any positive integral numeral subscript. Examples are "$a_1$", "$a_2$", and so on.

By a *function letter* we mean the small italic "$f$" together with positive integral numeral sub- and superscripts. Examples are "$f_1^1$", "$f_3^2$", etc. The reason for the double indexing will be made clear subsequently.

By a *predicate letter* we mean the capital italic letter "$A$" together with positive integral numeral sub- and superscripts. Examples are "$A_2^1$", "$A_7^4$", etc.

To these four *syntactical categories* of signs we add the following specific signs: "(" called *left parenthesis*; ")" called *right parenthesis*; "$\sim$" called *negation sign*; "$\vee$" called *disjunction sign*; and "," called *comma.* These signs plus the individual variables are called *logical signs.*

We now require that the alphabet of any given first-order theory include at least the following: all logical signs and at least one predicate letter. In addition, a particular first-order language may contain any number of constant letters, function letters, or additional predicate letters. A first-order theory is not defined until one has specified precisely which signs belong to its alphabet. (The possibility exists of a first-order theory having signs other than the above, but we defer this question until later.)

Some of our signs, such as the parentheses and comma, are signs of ordinary English and will thus be used in our metalanguage, as well as in our formal language. This will cause no confusion since the formal usage of these signs will be precisely defined.

It will often be convenient to speak of the variables and constant letters as being ordered in some well-defined manner. The ordering by increasing order of subscript is called *alphabetic order* and it will be the one most frequently used.

By a *term* of a first-order theory, we mean (i) an individual variable, (ii) a constant letter of the theory, or (iii) a function letter $f_m^n$ of the theory followed by one left parenthesis, then $n$ terms, separated by commas, and then a right parenthesis; i.e. an expression of the form

$$f_m^n(t_1, t_2, \ldots, t_n)$$

where the $t_i$ are all terms.† (iv) These are the only terms by this definition. (Actually, we

† Here the signs "$f_m^n$", "$t_1$", and the like are metavariables that we use in order to talk about signs and expressions of our formal system. We use letters such as "$A$", "$B$", "$x$", "$x_i$", etc., which are not signs of our system, as metavariables to refer to arbitrary wffs, arbitrary variables, and the like. In some cases the use of these letters in the metalanguage will be more like the dummy letters used in connection with our system **P**, though we will not worry about distinguishing between these two uses. As usual, juxtaposition of variables and special signs in the metalanguage represents juxtaposition in the object language.

Many books use different type styles for metavariables. We shall not, since we have unequivocally designated

Copyrighted material

shall later introduce other types of terms involving an extension of the present definition.)

Notice that the superscript number $n$ of the function letter in part (iii) of the above definition is the same as the number of terms $t_i$ involved in forming the new term. We say that the superscript of any function letter is the ***argument number*** of the function letter. Similarly, we call the superscript of a predicate letter its ***argument number***. The argument number of a function letter tells us how many terms are necessary to combine with the function letter to form a new term. The particular terms $t_i$ used in forming the new term are called the *arguments* of the new term.

The subscript numerals of the function letters simply distinguish between different function letters. Also, the subscript numerals of the individual variables and constant letters serve, in each case, to differentiate them. In the latter two cases, we need no argument number since the use we make of these two syntactical categories does not involve their having other things as arguments.

We define a wff of a first-order theory as follows: (i) Any predicate letter $A_m^n$ of the system followed by $n$ terms as arguments, i.e. an expression of the form $A_m^n(t_1, t_2, \ldots, t_n)$ where the $t_i$ are all terms, is a wff. (ii) If $X$ is a wff, then the expression $(y)X$ is a wff (where $y$ is any individual variable) and the expression $(\sim X)$ is a wff. (iii) If $X$ and $Y$ are any two wffs, then the expression $(X \vee Y)$ is a wff. (iv) These are the only wffs. Wffs of type (i), from which our other wffs are built up, are called *prime* formulas.

Having now defined clearly the wffs of a first-order language, we must give some attention to the question of interpretation. We shall eventually consider a precise, mathematically defined notion of interpretation. For the moment, however, we proceed on the intuitive level.

The intuitive interpretation of a first-order system is that the predicate letters express properties or relations, depending on the argument number. For example, the predicate letter $A_1^1$ might be thought of as expressing the property "to be red". Then, where $a_i$ is some constant letter thought of as naming some object, the wff $A_1^1(a_i)$ would mean, intuitively, "the object designated by $a_i$ is red". The negation and disjunction signs have the same meaning in this system as they do in the statement calculus. Thus, "the object designated by $a_i$ is not red" would be rendered as $(\sim A_1^1(a_i))$.

In any first-order system, we introduce the usual definitions for the signs "$\supset$", "$\wedge$", and "$\equiv$", following Definitions 1, 2, and 3 of Section 3, where the wffs of **P** are replaced, in each case, by the wffs of the first-order theory in question. Henceforth, whenever we speak of a first-order system, we suppose these definitions to have been made.

Using our connectives, we can now express new properties by using our original ones together with the names of other objects. "If the object $a_1$ is red, then the object $a_2$ is not red" can be expressed by the wff $A_1^1(a_1) \supset (\sim A_1^1(a_2))$. The number of possible combinations is clearly not finitely limited.

---

those signs which can be in the alphabet of a first-order system and these will never be used as metavariables.

Notice that the alphabet of any first-order system is infinite. Our individual variables *are* such signs as "$x_1$", "$x_2$", and the like, rather than being abbreviations for other expressions as was the case with our statement letters of **P**. Since we require that any first-order theory have all the individual variables among the signs of its alphabet, it follows that the theory has an infinite alphabet.

Technically, the sub- and superscripts on the variables, constant letters, function letters, and predicate letters are all numerals, names of integers. We now agree to use the words "subscript" and "superscript" to designate also the number that is named by a particular numeral subscript or superscript respectively. Thus, the number two is the subscript of the variable $x_2$, the number three is the superscript of the function letter $f_2^3$, and so on.

Copyrighted material

Notice that we here use parentheses for grouping in a way similar to our previous use of brackets. But our parentheses also have other uses such as applying a predicate letter to its arguments. To achieve an economy of notation, we have assimilated all these functions to parentheses and dispensed with brackets entirely.

A two-argument predicate letter represents a binary relation. For example, $A_1^2$ might express the relation "less than". Then, $A_1^2(a_1, a_2)$ would mean intuitively "$a_1$ is less than $a_2$". Relations of higher degree have more argument places. The subscript of the predicate letters simply distinguishes among different predicate letters.

From this brief discussion, we have seen one way of obtaining a statement from a predicate letter, namely by using individual constant letters as arguments. Another basic way is *quantification*, which we now explain.

The expression "$(x_i)$", where $x_i$ is some individual variable, is thought of as expressing the words "for all $x_i$". The variables $x_i$ may be thought of as playing the role of the pronouns of ordinary speech, just as the constant letters play the role of proper names. For example, if the predicate letter $A_1^1$ is thought of as expressing the property "to be red", the expression "$A_1^1(x_2)$" means "it is red" with ambiguous antecedent for "it". By affixing now our expression "$(x_2)$" we obtain "$(x_2)\, A_1^1(x_2)$", which means "Whatever thing you may choose, that thing (it) is red", or more succinctly, "Everything is red".

Notice that the expression "$A_1^1(x_2)$" does not represent a sentence, since the variable "$x_2$" is not thought of as being a proper name for some object as was the case with our constant letters. Thus, properly speaking, the phrase "it is red" or "$x$ is red" has no subject since the pronoun "it" (or "$x$") cannot be regarded as a subject if it has no antecedent at least understood from the given context. We use the term *open sentence* to refer to an expression which is obtained from a sentence by replacing one or more substantives (nouns) by pronouns with ambiguous antecedents. A formal expression such as "$A_1^1(x_2)$" involving one or more predicate letters with individual variables as arguments is thought of as representing in our formalism the intuitive, informal linguistic notion of an open sentence.

Our individual variables are thus thought of as variables ranging over some given domain $D$ called the *universe* or *universe of discourse*. The variables, unlike the constant letters, do not stand for a particular object, but they are thought of as naming ambiguously any arbitrary member of the given domain $D$. For our variables to have such an interpretation, we must specify the domain $D$, just as we must specify the meaning we assign to the predicate letters such as $A_1^1$ and the objects which are named by the constant letters.

There are thus two basic ways to obtain a sentence from an open sentence such as "$x$ is red", which is represented in our formalism, let us say, by $A_1^1(x_2)$. One way is the obvious device of substituting a name for the variable, thus replacing the variable by a constant. If we replace "$x$" by "the Washington Monument", we obtain the false sentence "The Washington Monument is red". Again, if we have some constant letter, say "$a_1$", which is thought of as naming the Washington Monument, then $A_1^1(a_1)$ will mean "The Washington Monument is red" in our formalism.

The second method of obtaining a sentence from an open sentence is by *quantification of variables*, which is the application of our prefix "for all", using some variable $x$. We say that we quantify in the *name* of the given variable. Thus "For all $x$, $x$ is red" is represented as previously given by "$(x_2)\, A_1^1(x_2)$". Hence, quantification is a new logical operation just like negation, disjunction, and the like. The latter operate on sentences to give new sentences, whereas quan-

Copyrighted material

tification operates on open sentences to give sentences or open sentences (this latter case may occur when there is more than one variable in the expression to which quantification is applied).

Substitution of constants for variables is likewise a logical operation used for obtaining sentences (or open sentences when some variables are not quantified or replaced by constants) from open sentences.

The logic of first-order theories, which will be embodied in logical axioms and rules that we have yet to describe, can be thought of as a generalization of the logic of the sentence connectives, the sentential calculus. The sentential calculus deals with valid or universally true ways of operating on sentences with sentential connectives. The *predicate calculus*, as we shall define it, can be thought of as the analysis of the valid ways of combining the sentential connectives plus the additional operations of quantification and substitution. In the predicate calculus, moreover, we are operating with a larger class of expressions, representing open sentences as well as sentences.

Returning to our examples again, we observe that we can iterate and compound our various operations just as with the sentential calculus. We can say "$(x_2)(\sim A_1^1(x_2))$", "For all $x_2$, $x_2$ does not have the property expressed by '$A_1^1$'", or "$(\sim (x_2) A_1^1(x_2))$", "Not everything has the property $A_1^1$". Using other predicate letters we can make more complicated statements such as "$(x_1)(x_2)(x_3)((A_1^2(x_1, x_2) \wedge A_2^2(x_2, x_3)) \supset A_1^2(x_1, x_3))$". "For all $x_1$, $x_2$, $x_3$ if $x_1$ bears the relation $A_1^2$ to $x_2$, and if $x_2$ bears the relation $A_2^2$ to $x_3$, then $x_1$ bears the relation $A_1^2$ to $x_3$."

Particularly interesting is the combination "$(\sim (x_i)(\sim A))$" where $x_i$ is any individual variable and $A$ is some wff. This says, "It is not true that, for all $x_i$, $A$ is not true." Otherwise said, "There exists some $x_i$ such that $A$ is true." By means of quantification and negation, we can express the notion of existence. This fact was first recognized by Frege [1].

**Definition 1.** Where $A$ is any wff of a first-order system and $x_i$ is any individual variable, we abbreviate $(\sim (x_i)(\sim A))$ by $(Ex_i)A$. We call "$(x_i)$" *universal* quantification and "$(Ex_i)$" *existential* quantification. We read the existential quantifier as, "There exists $x_i$ such that". "Quantification", without modification, will henceforth mean either existential or universal quantification. Quantification is said to be *in the name* of the variable $x_i$ appearing in the parentheses of the quantifier $(Qx_i)$, be it existential or universal.

We suppose this definition made for all first-order theories.

From now on, and for the rest of this book, we shall omit parentheses at will, but consistent with the following convention. The negation sign and the quantifier $(Qx)$, existential or universal, apply to the shortest wff that follows them. Next in line of increasing "strength" is the conjunction sign, then the disjunction sign, the conditional sign, and finally the biconditional. Let us take some examples: $(x_1) A_1^1(x_1) \vee A_2^1(x_1)$ is properly read as $((x_1) A_1^1(x_1) \vee A_2^1(x_1))$ and not $(x_1)(A_1^1(x_1) \vee A_2^1(x_1))$ since $(x_1)$ is weaker than $\vee$. $A_1^1(x_1) \wedge A_2^1(x_1) \vee A_3^1(x_1)$ is read $((A_1^1(x_1) \wedge A_2^1(x_1)) \vee A_3^1(x_1))$, since $\wedge$ is weaker than $\vee$. $A_1^1(x_1) \vee A_3^1(x_1) \supset A_2^1(x_1) \equiv A_4^1(x_1)$ is unambiguously read as $(((A_1^1(x_1) \vee A_3^1(x_1)) \supset A_2^1(x_1)) \equiv A_4^1(x_1))$. $\sim \sim \sim A_1^1(x_1) \vee A_2^1(x_1)$ is read as

$$((\sim(\sim(\sim A_1^1(x_1)))) \vee A_2^1(x_1)).$$

Any wff $B$ which has an occurrence within another wff $A$ is called a *subformula* of $A$. By the way we have defined the wffs of first-order theories, any occurrence of a universal or existential quantifier $(Qx)$ in a wff $A$ applies to some particular subformula $B$ of $A$. $B$ will be the

Copyrighted material

(unique) wff immediately following the given occurrence of $(Qx)$. The occurrence of $B$ immediately following the given occurrence of $(Qx)$ is called the *scope* of that occurrence of $(Qx)$.

For example, in the wff $(x_1)((x_1)A_1^1(x_1) \vee (\sim (x_1)A_1^1(x_1)))$ the scope of the first occurrence of $(x_1)$ is the entire formula which follows it. The scope of the second occurrence of $(x_1)$ is the first occurrence of $A_1^1(x_1)$ and the scope of the third occurrence of $(x_1)$ is the second occurrence of $A_1^1(x_1)$.

We now define an occurrence of a variable $x_i$ in a wff $A$ to be *bound* if it occurs in the scope of some occurrence of a quantifier in its name or if it occurs within the parentheses of a quantifier $(Qx_i)$. Otherwise, the occurrence is *free*. Clearly a given variable may have both bound and free occurrences in the same given formula. Nevertheless, if $x_i$ has at least one free occurrence in $A$, we say that $x_i$ is *free in* $A$ or that it is a *free variable of* $A$. Also, if $x_i$ has at least one bound occurrence in $A$, it is *bound in* $A$ and is a *bound variable of* $A$. Thus, bondage and freedom are opposites for occurrences but not for variables.

Now from our description of the intuitive interpretation of the formalism of a first-order theory, and from the definition of wffs, it is clear that a wff of a first-order theory may fail to represent a sentence. As we have seen, it will represent an open sentence if it contains variables which are not quantified; in other words, if it contains free variables. We define a wff to be *closed* if it contains no free variables. The set of closed wffs of a given first-order theory $F$ will be called the *sentences* or *propositions of* $F$. The closed wffs represent sentences of the vernacular under our intuitive interpretation. A closed wff of a first-order theory is thus the formal analogue of a sentence in the vernacular just as a wff which is not closed is a formal analogue of an open sentence.

We extend the notion of bondage and freedom to cover more general terms and subformulas: An occurrence of a term $t$ or of a subformula $B$ in a formula $A$ is *bound* if at least one of its free variables falls within the scope of a quantifier in its name. Otherwise, the occurrence is free. A term or formula is bound in another formula if it has at least one bound occurrence in that formula, and free if it has at least one free occurrence.

*Exercise*. State which occurrences of which terms and subformulas are free or bound in each of the following:

$$(x_2)(x_1)A_1^2(x_1, x_2) \supset \sim A_1^2(x_2, f_1^1(x_1));$$
$$(x_3)(A_1^2(x_1, x_2) \supset A_2^2(x_1, f_1^2(x_2, x_3)) \vee A_1^1(x_3);$$
$$(x_4)(A_1^1(x_1) \supset A_2^1(x_2)).$$

From the way we have defined our wffs, we can apply quantifiers indiscriminately to any wff and obtain a wff. What intuitive meaning do we give to such wffs as $(x_1)A_1^1(x_2)$ where the variable $x_2$ is still free since we have applied a quantifier in the name of another variable? Intuitively we regard the quantifier as vacuous in this case. The displayed wff above means the same as $A_1^1(x_2)$.

Another case of vacuous quantification is double quantification such as in the wff $(x_1)(x_1)A_1^1(x_1)$. Intuitively, this wff means the same as the one in which the initial quantifier is dropped.

We will so formulate our rules and axioms of logic that these intended equivalences in meaning turn out to be provable.

It remains to give the intuitive interpretation of the function letters $f_m^n$. Once our domain $D$



Copyrighted material

of the individual variables $x_i$ is chosen, the function letters are thought of as representing functions from $n$-tuples of elements of $D$ into $D$. By $D^n$ we mean the $n$-fold cartesian product of $D$ with itself, the set of $n$-tuples of elements of $D$. Where the argument number of a function letter is $n$, the interpretation of a function letter $f_m^n$ is a function from $D^n$ into $D$.

For example, suppose our domain $D$ is the natural numbers. Then we can think of $f_1^2$ as the operation of addition, which associates with any two elements $x_1$ and $x_2$ of $D$ (remember the variables range over $D$) the sum $f_1^2(x_1, x_2)$. For a less mathematical example, we can let $D$ be the set of all people living or dead and $f_1^1$ the function "father of". This would associate with every person $x_1$ his father $f_1^1(x_1)$.

We have given an intuitive explanation of the interpretation of the wffs and terms of a first-order theory. We now proceed to give a precise mathematical definition, which should be regarded primarily as a precise statement of our intuitive notion of an interpretation.

**Definition 2.** An *interpretation* $\langle D, g\rangle$ of a given first-order theory $F$ consists of a given, non-empty set $D$ together with a mapping $g$ from the set consisting of the function letters, individual constants, and predicate letters of $F$ into the set $D \cup \mathcal{P}(\mathcal{M}(D))$, which is the set $D$ together with all the subsets of the set $\mathcal{M}(D)$ of all finite sequences of elements of $D$. The mapping $g$ assigns to each predicate letter $A_m^n$ of $F$ some subset $g(A_m^n)$ of $D^n$; to each function letter $f_j^i$ of $F$ some subset $g(f_j^i)$ of $D^{i+1}$, $g(f_j^i)$ a functional relation, i.e. a mapping from $D^i$ into $D$; and to each constant letter $a_i$ some element $g(a_i)$ of the set $D$. We call $g(a_i)$ the object *named* by $a_i$.

Intuitively, $D$ is our universe of discourse over which the individual variables range. The mapping $g$ assigns relations (considered as sets of $n$-tuples) to predicate letters, operations (functions) to function letters, and elements of $D$ to constants. The set $D$ together with the selected elements of $D$, functions on $D^n$, and relations over $D$ assigned by the mapping $g$ to some constant letter, function letter, or predicate letter of $F$ (respectively) is sometimes called a *structure for F*.

Now, for any interpretation $\langle D, g\rangle$ of a formal system $F$, the sentences or closed wffs of the system should be either true or false in the interpretation since it was part of our original definition of sentences that they were either true or false. We now proceed to define truth and falsity, for a formal system relative to a given interpretation, in a purely mathematical way.

Suppose a first-order system $F$ given, and let some interpretation $\langle D, g\rangle$ be given. Let us define for each *infinite* sequence $s = s_1, s_2, \ldots, s_i, \ldots$ of elements of $D$ a function $g_s$ from the set of terms of $F$ into the domain $D$. If $t$ is an individual constant $a_i$, then $g_s(t)$ is the element $g(a_i) \in D$ that is named by $a_i$. If $t$ is the individual variable $x_i$, then $g_s(x_i) = s_i$ (i.e. the $i$th member of the sequence). Finally, if $t$ is of the form $f_j^n(t_1, \ldots, t_n)$, then $g_s(t) = g(f_j^n)(g_s(t_1), \ldots, g_s(t_n))$; that is, we apply the associated operation of $f_j^n$ to the objects in $D$ which correspond to the terms $t_i$ that make up the term $t$. We have used the notation $g_s$ to emphasize the dependence on the chosen sequence $s$.

Next we define what it means for a sequence $s$ to *satisfy* a given wff $A$. If $A$ is of the form $A_k^n(t_1, \ldots, t_n)$ and if the $n$-tuple

$$\langle g_s(t_1), g_s(t_2), \ldots, g_s(t_n)\rangle$$

is in the set $g(A_k^n)$ then we say that the sequence $s$ *satisfies* the wff $A$. Otherwise, the sequence does not satisfy that wff. If $A$ is of the form $(\sim B)$, then $s$ *satisfies* $A$ if and only if it does not

Copyrighted material

satisfy $B$. If $A$ is of the form $(B \vee C)$, then $s$ satisfies $A$ if and only if $s$ satisfies $B$ or $C$ or both. If $A$ is of the form $(x_i)B$, then $s$ satisfies $A$ if and only if every infinite sequence of elements of $D$ which differs from $s$ in at most the $i$th component satisfies $B$.

We now say that a wff $A$ is *true* under a given interpretation if and only if *every* sequence $s$ satisfies $A$. We say that $A$ is *false* if and only if *no* sequence satisfies $A$.

According to our definition of truth, it is possible for a wff of a formal system to be neither true nor false under a given interpretation. It may be satisfied by some sequences but not by all. However, it is not possible for a closed wff to be neither true nor false. For any closed wff $X$, either every sequence satisfies $X$ or no sequence satisfies $X$.

The rigorous proof of this last statement is by induction, and we give only a sketch. For this purpose, let us define the *closure* or the *universal closure* of a given wff. Let $t_1, t_2, \ldots, t_n$ be the free variables of a given wff $X$ in increasing order of subscript (i.e. if $x_r$ is $t_j$ and $x_m$ is $t_i$, then $j \leq i$ if and only if $r \leq m$). The universal closure of $X$ is the wff

$$(t_n)(t_{n-1}) \ldots (t_1)X,$$

obtained by prefixing universal quantifiers in the name of each free variable of $X$ in the indicated order. Actually, the question of order does not really matter, but we choose some order so that the closure will be uniquely defined for a given wff. We also define the *existential closure* of $X$ to be the wff obtained by prefixing existential quantifiers in the name of all free variables in the indicated order; that is, the wff

$$(Et_n)(Et_{n-1}) \ldots (Et_1)X.$$

If $X$ is a closed wff, then $X$ is its own universal closure and its own existential closure.

The first observation concerning our definition of truth and the concepts we have just defined is the following: Any wff $X$ whatever is true if and only if its universal closure is true. By definition a sequence $s$ satisfies $(x_i)X$ if and only if *every* sequence which differs from $s$ in at most the $i$th place satisfies $X$. But if every sequence satisfies $X$ to begin with (i.e. if $X$ is true), then it immediately follows that every sequence will satisfy $(x_i)X$. Also, if every sequence satisfies $(x_i)X$ then certainly every sequence satisfies $X$ (remember, any sequence $s$ differs from itself in at most the $i$th component). Iterating this argument to any number of applications of universal quantification to $X$, we obtain that the universal closure of $X$ is true if and only if $X$ is true.

Thus, one finds that the truth of any wff $X$ of any system $F$ under a given interpretation $\langle D, g \rangle$ is equivalent to the truth of some closed wff, namely the universal closure of $X$.

Another important observation concerning the satisfaction relation between sequences and formulas is the following: Let $A$ be any formula all of whose free variables are included in the list $x_{i_1}, \ldots, x_{i_n}$. Then any two sequences $s$ and $s'$ for which $s_{i_j} = s'_{i_j}$, $1 \leq j \leq n$, either satisfy or fail to satisfy $A$ together. In other words, the values of a sequence at indices corresponding to variables which are not free in $A$ do not affect the satisfaction of $A$ by $s$. Thus, even though sequences are infinite in length, we are never really concerned with more than a finite number of values at any given time since any formula has only a finite number of free variables.

Using these observations, we now want to sketch the inductive proof that any closed wff $X$ in any system $F$ is either true or false under any given interpretation $\langle D, g \rangle$ of $F$. Notice that this is equivalent to proving that whenever one sequence $s$ satisfies $X$, then every sequence does. The induction is on the number of sentence connectives and quantifiers of $X$.



Copyrighted material

If $X$ has none of these, then it must be a prime formula, and since it is closed it must contain only variable-free terms as arguments. $X$ is thus of the form $A_j^n(t_1, t_2, \ldots, t_n)$ where the $t_i$ contain no variables. In this case, $g_s(t_i)$ is the same for every sequence $s$. Hence, either all sequences will satisfy $X$ or none will, and the theorem holds in this case.

Assuming, inductively, that the theorem holds for all wffs with fewer than $n$ sentence connectives and quantifiers, we suppose $X$ has $n$ connectives and quantifiers. If $X$ is prime, the proof is the same as above. Otherwise, $X$ is of the form $(\sim B)$, $(B \vee C)$, or $(x_i)B$, and the theorem holds for $B$ and $C$ since they each have less than $n$ quantifiers and sentence connectives. Let us consider each case. If $X$ is of the form $(\sim B)$, then $B$ is closed since $(\sim B)$ has the same free variables as $B$. Thus, applying the theorem to $B$, either every sequence satisfies $B$ and thus does not satisfy $(\sim B)$ (by the definition of satisfaction), or else no sequence satisfies $B$ in which case every sequence satisfies $(\sim B)$, again by the definition of satisfaction. If $X$ is of the form $(B \vee C)$, then suppose there is no sequence satisfying either $B$ or $C$. Then none satisfies $(B \vee C)$ and it is false. Otherwise, there is at least one sequence satisfying one of $B$ or $C$, say $C$. Now, $C$ must be closed since $(B \vee C)$ is. Applying the induction hypothesis to $C$, we conclude that every sequence satisfies $C$. Thus, by the definition of satisfaction, every sequence satisfies $(B \vee C)$.

Finally, we consider the case where $X$ is of the form $(x_i)B$. Either $x_i$ is free in $B$ or not. If not, then $B$ is closed (since $(x_i)B$ is) and the result is immediate since the quantifier now changes nothing. Otherwise, $(x_i)B$ is the universal closure of $B$ and $x_i$ is the only free variable in $B$. Thus, for any sequence $s$, the value $s_i$ is the only relevant one for determining the satisfaction of $B$ by $s$. Suppose, now, that at least one sequence $s$ satisfies $(x_i)B$. Then every sequence $s'$ which differs from $s$ in at most the $i$th place satisfies $B$ (the definition of satisfaction). In particular, $s$ satisfies $B$. But every sequence which differs from $s$ in any place other than the $i$th one also satisfies $B$ by our second observation above (i.e. because $x_i$ is the only free variable of $B$). Thus, every sequence satisfies $B$ and $B$ is true. Hence, by our first observation, its universal closure $(x_i)B$ is true and is thus satisfied by every sequence. This completes the proof.

The fact that closed wffs are either true or false under any interpretation justifies applying the term "sentence" or "proposition" to them.

*Exercise*. Prove that a sequence $s$ satisfies a wff $(Ex_i)B$ for a given interpretation if and only if there is at least one sequence differing from $s$ in at most the $i$th place and satisfying $B$.

The possibility of a rigorous definition of truth was first conceived and executed by Tarski (see Tarski [1], p. 152). Now that we have such a notion at hand, we can define what we mean by a logically valid wff of a first-order theory. We say that a wff $X$ of a first-order theory $F$ is *logically valid* or *universally valid* if $X$ is true for every interpretation of $F$.

The notion of logical validity for a first-order theory is analogous to the notion of tautology for the system **P** of the sentential calculus. In fact, it is a generalization of the notion of tautology. We ask the reader to show this.

*Exercise* 1. Show that any wff of a first-order system $F$ which is tautological in form will be logically valid as a wff of $F$ according to the given definition.

*Exercise* 2. Find some examples of wffs of first-order systems which are valid, but which are not tautologies.

Copyrighted material

It is important to understand that our definition of truth is rigorous. However, the reader should also see that the definition only makes precise the notion he would normally have of truth. As such, this intuitive notion will often suffice in understanding a given discussion of logical questions. Thus, to take an example, if we consider the wff $(x_1)(Ex_2)A_1^2(x_1, x_2)$ in some formal theory and if we let the domain $D$ be the natural numbers and the relation $g(A_1^2)$ be the relation "less than" (the set of all ordered pairs of natural numbers $x$, $y$ such that $x$ is less than $y$), then the wff says that for every natural number $x_1$ there is some natural number $x_2$ greater than $x_1$. This statement is true in the natural numbers.

Moreover, it is clear that there is a qualitative difference between the definition of logical validity and the definition of tautology. For tautology, we have a purely mechanical test, the truth table, which allows us to decide whether a given wff is a tautology or not. But the definitions of interpretation, truth, and validity all depend on general notions of set theory. There is no simple way of deciding whether a given wff is valid or not. In some cases, as with tautologies, we can decide. Church has proved, however, that there is no general method of decision by which one can decide for all wffs of any system $F$ whether or not they are logically valid (see Church [1]).

The concepts of truth, interpretation, and validity as we have defined them are as legitimate as any abstractly defined mathematical concept. But for practical purposes it would be very difficult to have to return again and again to these abstract definitions in order to prove facts about validity. We thus conceive of the following plan: We specify certain axioms and rules of inference and require that they hold for all first-order theories. A given first-order theory may have other axioms, but the axioms we specify are *logical axioms* required to hold in any system. These are called the *axioms and rules of the predicate calculus*. The logical axioms will be a decidable set and the rules will be simple, formal, decidable rules like the rule of *modus ponens* of **P** (in fact, *modus ponens* will be one of our rules). Furthermore, it will turn out that all logically valid wffs and only logically valid wffs can be formally proved from the axioms and rules of the predicate calculus. Thus, we can replace the notion of validity by the notion of formal deducibility or provability in the predicate calculus.

The virtue of this proposed plan is that, while the notion of validity depends for its definition on general set theory, the notion of formal deducibility does not. From the way we have defined formal deducibility in a formal system, it is clear that the only mathematical notions involved are those which are essentially number-theoretic, having to do with the length of proofs, and so on. Metatheorems about formal deducibility will usually involve no more tools than elementary number theory and the principle of mathematical induction. But metatheorems about validity easily involve highly nonconstructive principles of general set theory such as the axiom of choice.

This last point is very important for the purposes of our study in this book. We will be treating, in future chapters, different formal languages in which mathematics and general set theory can be expressed. These languages will, for the most part, be first-order theories. If our only approach to a logical discussion of these languages was in terms of interpretations and validity, there would be little virtue in the formal axiomatic approach. Instead, we shall proceed by formal deduction within these languages, using our decidable axioms and rules. Thus, theorems within these formal languages will be those wffs for which we actually exhibit a purely formal, mechanical deduction; no concepts of general set theory shall be necessary to justify the notion of proof.

Of course nothing prevents us from studying these set-theoretical languages from the point

Copyrighted material

of view of general set theory itself. Chapter 6 of this book contains a detailed discussion of this question, and the interested reader can proceed directly to this chapter if he so wishes.

We now address ourselves to the question of formulating the formal axioms and rules of the predicate calculus. The reader should note that our definitions are purely formal and do not involve appeal to general set-theoretic concepts.

We have used letters such as "$A$", "$B$", "$x$", and the like, which are not part of the alphabet or expressions of first-order systems, as variables in the metalanguage to represent such constructs as arbitrary wffs and arbitrary variables. We now use such forms as "$A(x)$" to represent a wff which may contain the variable $x$ free. Similarly, "$A(x, y)$" represents a wff which may contain $x$ and $y$ free (of course $x$ and $y$ may be the same variable here). If $A(x)$ is a wff which may contain $x$ free, then $A(t)$ will represent the result of replacing the term $t$ for $x$ in all of the free occurrences (if any) of $x$ in $A(x)$.

Given a wff $A(x)$, we say that another term $t$ is *free for* $x$ in $A(x)$ if every new occurrence of $t$ in $A(t)$ is free. For example, in the wff $(x_2)\, A_1^2(x_1, x_2)$, the term $x_2$ is not free for $x_1$, since it will become bound if we substitute $x_2$ for $x_1$ in this wff. Similarly, the term $f_1^2(x_3, x_2)$ is not free for $x_1$, since one of its variables, namely $x_2$, will become bound if the term is substituted for $x_1$ in the wff. Any variable other than $x_2$ and any term not containing $x_2$ free is free for $x_1$. Obviously, any variable is free for itself in any formula.

Given a wff $A(x)$ and a free occurrence of the variable $x$ in $A(x)$, we say that a term $t$ is *free for the occurrence of* $x$ in question if no free occurrence of a variable in $t$ becomes bound by a quantifier of $A(x)$ when $t$ is substituted for the given occurrence of $x$. A term $t$ is free for $x$ in the sense of the preceding paragraph if it is free for every free occurrence of $x$ in $A(x)$. It is possible for a term $t$ to be free for some free occurrences of $x$ and not for others.

Notice that substitution is not generally a symmetrical operation. If we obtain $A(y)$ from $A(x)$ by substitution where the variable $y$ is free for $x$, it does not follow that we can obtain $A(y)$ from $A(x)$. For example, $A_1^2(x_1, x_1)$ is obtained from $A_1^2(x_1, x_2)$ by substitution of $x_1$ for $x_2$ but we cannot reverse the procedure. Whenever two wffs are obtainable each from the other by opposite substitutions, we say they are *similar*. More precisely, if $y$ is a variable free for $x$ in $A(x)$ and if $y$ has no free occurrences in $A(x)$, them $A(y)$ is similar to $A(x)$. Thus, $A_1^2(x_1, x_2)$ is similar to $A_1^2(x_1, x_3)$.

We now require that the axiom set of any first-order system $F$ satisfy the following: (1) All wffs of $F$ which are tautological in form are axioms. (2) If $A(x)$ is some wff and $x$ a variable, then every wff of the form $(x)\, A(x) \supset A(t)$ is an axiom where $t$ is any term free for $x$ (remember that $A(t)$ is obtained from $A(x)$ by replacing $t$ for $x$ at all of the latter's free occurrences, if any, in $A(x)$). (3) Every wff of the form

$$(x)\,(B \supset A(x)) \supset (B \supset (x)\, A(x))$$

where $x$ is any variable that is *not* free in $B$, and $A(x)$ is any wff, is an axiom. (4) These are the only logical axioms.

The following are the rules of inference of any first-order system. (1) *Modus ponens*; that is, from $A$ and $A \supset B$ we can infer $B$ where $A$ and $B$ are any wff. (2) For any wff $A$, we can infer the wff $(x)A$. This is the rule of universal generalization (abbreviated "UG").

It is presumed that a first-order system may have other *proper* axioms (meaning proper to the particular theory in question), but the rules of inference are the same. Any first-order system having only the logical axioms is called a *predicate calculus*. By *the* predicate calculus we mean

Copyrighted material

the theory of formal deduction using our logical axioms and rules. By a *theorem of the predicate calculus* we mean any theorem of any predicate calculus. By the *pure predicate calculus* we mean the first-order predicate calculus with no function letters, no constant letters, and all the predicate letters.

In any formal system, a formal deduction is a finite sequence of wffs such that every element of the sequence is either an axiom or follows from prior members of the list of axioms by a rule of inference. Thus, the theorems of any predicate calculus are the wffs that can be obtained by formal deduction from our logical axioms by means of our rules of inference.

From now on, to the end of this chapter, the terms "system" and "formal system" will be restricted to mean "first-order system" unless otherwise specified.

It may seem surprising that such a simple set of axioms suffices to yield, as theorems, precisely the universally valid wffs, but this is so. We now wish to examine this question more closely.

Given a first-order system, we now have two parallel notions. We have the notion of a provable wff or theorem, and the notion of a true wff under a given interpretation. For any such first-order theory $F$, those properties of $F$ which are defined by means of interpretations of $F$ are called *semantical* properties of $F$. Those properties which are defined in terms of the deductive structure of $F$ are called *syntactical* properties or proof-theoretic properties of $F$. What we are then interested in is the precise relationship between syntax and semantics.

It is easy to see that the theorems of any predicate calculus must be logically valid. First, we establish that any tautology is universally valid (cf. preceding exercise). Next, consider a wff of the form $(x_i)\, A(x_i) \supset A(t)$ where $t$ is free for $x_i$ in $A(x_i)$. This is an axiom under scheme (2) of our logical axioms. This will be universally valid only if it is true under every interpretation; that is, if every sequence in every interpretation satisfies it. This will be the case only if, for every interpretation, every sequence satisfying $(x_i)\, A(x_i)$ satisfies $A(t)$.

Thus, let some interpretation $\langle D, g\rangle$ be chosen, and consider any sequence $s$ satisfying $(x_i)\, A(x_i)$. By our definition of satisfaction, this means that every sequence differing from $s$ in at most the $i$th place satisfies $A(x_i)$. Thus, whatever the object $g_s(t)$ may be, the sequence $s$ must *a fortiori* satisfy $A(t)$ (a rigorous proof of this is by induction on the structure of $A(t)$). Hence, any formula of the form $(x_i)\ A(x_i) \supset A(t)$ is true in any interpretation and thus universally valid.

*Exercise.* Sketch a proof of the fact that wffs of the form (3) of our logical axioms are universally valid.

Next, we observe that our rules of inference preserve universal validity. This is obviously true for *modus ponens* in view of the truth table for the conditional. We have already remarked that any wff is true if and only if its universal closure is. Thus, the rule of generalization also preserves universal validity.

Since a theorem of a predicate calculus is obtained from the axioms by our rules of inference, it follows that the theorems must be universally valid (again, we skirt an induction on the length of the proof of the theorem).

This situation is clearly analogous to our system **P** in which the rules of inference (namely *modus ponens*) preserved the property of being tautological, and all of our axioms were tautologies. It then followed that all theorems of **P** were tautologies. The converse, that all tautologies of **P** were theorems of **P**, was stated but not proved. Likewise, we state, but do not prove:

Copyrighted material

THEOREM 1. *For any first-order system F, the universally valid formulas of F are precisely those theorems of F deducible from the logical axioms of F by our rules of inference. Every universally valid wff of F is thus a theorem of F.*

The proof of this nontrivial theorem was first given by Gödel [1]. This article is reprinted in van Heijenoort [1], p. 582.

The only real justification for our logical axioms and rules of inference is Theorem 1. The extreme importance of the theorem lies in the fact that we can describe the universally valid wff syntactically as well as semantically. That is, we have a decidable set of purely formal axioms and formal rules which generate the universally valid formulas of any given first-order system $F$.

The importance of the axioms being a decidable set has already been emphasized, and should not be overlooked. It is easy to designate a set of axioms for the universally valid formulas of a system $F$ in a nonformal, undecidable way; just let the valid formulas be axioms, for example, and have no rules of inference. This may seem analogous to designating the tautologies as axioms, but again we emphasize that the tautologies are a decidable set of wffs. We can determine whether a given wff is a tautology by the truth-table method just as surely as we can determine whether or not a given wff really is or is not of the form $(x_i)\, A(x_i) \supset A(t)$.

In view of Theorem 1, one might hope to prove that the valid formulas are a decidable set after all by proving that the set of theorems of the predicate calculus is decidable (for these two sets are the same). As we have already mentioned, however, Church has proved the theorems of the predicate calculus to be undecidable (see Church [1]). There is, in short, no mechanical test which will allow us generally to determine whether a given wff is universally valid (or, which is the same thing by Theorem 1, a theorem of the predicate calculus). For this reason, technique and skill in logical deduction are of some importance.

Before continuing our general discussion of first-order theories, let us exhibit a formal deduction in the predicate calculus. We show, for example, $\vdash A_1^1(x_1) \supset (Ex_1)\, A_1^1(x_1)$ in any theory $F$ having $A_1^1$ as a predicate letter:

$$(x_1)\,(\sim A_1^1(x_1)) \supset (\sim A_1^1(x_1))$$

$$((x_1)\,(\sim A_1^1(x_1)) \supset (\sim A_1^1(x_1))) \supset ((\sim(\sim A_1^1(x_1))) \supset (\sim(x_1)\,(\sim A_1^1(x_1))))$$

$$(\sim(\sim A_1^1(x_1))) \supset (\sim(x_1)\,(\sim A_1^1(x_1)))$$

$$A_1^1(x_1) \supset (\sim(\sim A_1^1(x_1)))$$

$$\left\{\begin{array}{l}(A_1^1(x_1) \supset (\sim(\sim A_1^1(x_1)))) \supset (((\sim(\sim A_1^1(x_1))) \supset (\sim(x_1)\,(\sim A_1^1(x_1)))) \\ \qquad \supset (A_1^1(x_1) \supset (\sim(x_1)\,(\sim A_1^1(x_1)))))\end{array}\right.$$

$$((\sim(\sim A_1^1(x_1))) \supset (\sim(x_1)\,(\sim A_1^1(x_1)))) \supset (A_1^1(x_1) \supset (\sim(x_1)\,(\sim A_1^1(x_1)))$$

$$A_1^1(x_1) \supset (\sim(x_1)\,(\sim A_1^1(x_1)))$$

The above sequence of seven wffs is a formal proof whose last line is the desired wff. We leave as an exercise to the reader the task of verifying that this sequence of wffs really meets the criteria of a formal proof using only our logical axioms and rules of inference.

Recall from Section 1.2 that a given line in a formal proof is *justified* by the particular criteria for members of a formal proof that it satisfies. Thus, the *justification* for a line of a proof is (i) that it is an instance of one of our axiom schemes or (ii) that it is inferred from some prior members by one of our rules of inference. Although it is not strictly necessary, it is helpful in

Copyrighted material

verifying formal proofs if one states the justification for each line. It is also useful to number the lines and to abbreviate wffs to avoid rewriting long wffs. We illustrate this by showing that, in any first-order theory $F$, $\vdash (x_i)(A \vee B(x_i)) \supset (A \vee (x_i) B(x_i))$, where the variable $x_i$ is not free in the wff $A$ and $B(x_i)$ is any wff of $F$. In stating our justifications, we abbreviate "tautology" by "Taut", "universal generalization" by "UG", and "*modus ponens*" by "MP".

1. $(x_i)(A \vee B(x_i)) \supset (A \vee B(x_i))$ Axiom 2, $t$ is $x_i$
2. $(A \vee B(x_i)) \supset (\sim A \supset B(x_i))$ Taut
3. † $[1] \supset ([2] \supset ((x_i)(A \vee B(x_i)) \supset (\sim A \supset B(x_i))))$ Taut
4. ‡ $R[3]$ 1, 3, MP
5. $(x_i)(A \vee B(x_i)) \supset (\sim A \supset B(x_i))$ 2, 4, MP
6. $(x_i)((x_i)(A \vee B(x_i)) \supset (\sim A \supset B(x_i)))$ 5, UG
7. $[6] \supset ((x_i)(A \vee B(x_i)) \supset (x_i)(\sim A \supset B(x_i)))$ Axiom 3, $x_i$ is not free in $(x_i)(A \vee B(x_i))$
8. $R[7]$ 6, 7, MP
9. $(x_i)(\sim A \supset B(x_i)) \supset (\sim A \supset (x_i) B(x_i))$ Axiom 3, $x_i$ not free in $\sim A$ since it is not free in $A$
10. $[8] \supset ([9] \supset ((x_i)(A \vee B(x_i)) \supset (\sim A \supset (x_i) B(x_i))))$ Taut
11. $R[10]$ 8, 10, MP
12. $(x_i)(A \vee B(x_i)) \supset (\sim A \supset (x_i) B(x_i))$ 9, 11, MP
13. $(\sim A \supset (x_i) B(x_i)) \supset (A \vee (x_i) B(x_i))$ Taut
14. $[12] \supset ([13] \supset ((x_i)(A \vee B(x_i)) \supset (A \vee (x_i) B(x_i))))$ Taut
15. $R[14]$ 12, 14, MP
16. $(x_i)(A \vee B(x_i)) \supset (A \vee (x_i) B(x_i))$ 13, 15, MP

This last theorem is really a metatheorem, for we have not specified the wffs $A$ and $B(x_i)$, but only required that they satisfy certain conditions. Of course, for any particular wffs satisfying the conditions, and for any system $F$, the proof would be line for line the proof we have given.

Our first theorem also gives rise to a metatheorem that

$$\vdash A(x_i) \supset (Ex_i) A(x_i)$$

in any formal system $F$ where $A(x_i)$ is any wff of the theory. The proof in this general case will be obtained by replacing the wff $A(x_i)$, whatever it may be, for the particular wff $A_1^1(x_1)$ in the preceding proof. In fact, we obtain an even stronger metatheorem as the following exercise shows.

*Exercise*. Let $F$ be any first-order system and $A(x)$ a wff of $F$. If the term $t$ of $F$ is free for $x$ in $A(x)$ and if $A(t)$ results from $A(x)$ by substituting the term $t$ for all free occurrences of the variable $x$ in $A(x)$, then $\vdash A(t) \supset (Ex) A(x)$.

† The numbers in brackets form the name of the wff occurring at the line of the proof having the bracketed number. One is to imagine that we have written out the wff that would occur if we replaced the number by the indicated line of the proof.

‡ This means "the wff occurring on the right side of the principal connective of line 3". This device, like the bracketed numbers, is to shorten the writing of complicated wffs.

Copyrighted material

Prove, using our logical axioms and rules, that $\vdash_F (x)\,A(x) \equiv (y)\,A(y)$ and $\vdash_F (Ex)\,A(x) \equiv (Ey)\,A(y)$ in any system $F$, where $A(x)$ and $A(y)$ are similar.

Prove, using our logical axioms and rules, that $\vdash_F (x)(y)A \equiv (y)(x)A$ in any system $F$.

**Definition 3.** Given two wffs $A$ and $B$ of a first-order theory $F$, we say that *A implies B in F* if and only if $\vdash_F A \supset B$. If $F$ is a predicate calculus, we say that *A logically implies B* or simply *A implies B*. By Theorem 1, $A$ implies $B$ means that the conditional $A \supset B$ is logically valid.

Logical implication is a generalization of tautological implication defined in Section 1.1, Definition 4. Since tautologies are theorems of any predicate calculus, a tautological implication is also a logical implication. Obviously, the converse is not true in general, and logical implication is a broader relation than tautological implication.

**Definition 4.** Given two wffs $A$ and $B$ of a first-order theory $F$, we say that *A is equivalent to B in F* if $\vdash_F A \equiv B$. If $F$ is a predicate calculus, we say that $A$ is *logically equivalent to B* or simply $A$ is *equivalent to B*. By Theorem 1, this means that $A \equiv B$ is logically valid.

This is again a parallel generalization of the relation of tautological equivalence given in Definition 5 of Section 1.1.

We now return to our general discussion of first-order theories.

### 1.5. Models of first-order theories

**Definition 1.** Let $F$ be some first-order system. By a *model* for $F$ we mean any interpretation of $F$ in which the proper axioms of $F$ are all true. (Remember that the proper axioms of $F$ are those axioms of $F$, if any, other than our logical axioms. The logical axioms of $F$ are automatically true in any model for $F$, since they are universally valid and thus true under every interpretation.)

It does not follow that every first-order system has a model. The question of whether or not a system $F$ does have a model is closely related to the question of consistency.

**Definition 2.** A first-order system $F$ is called *inconsistent* or *contradictory* if there is some wff $A$ of $F$ such that the wff $(A \wedge (\sim A))$ is a theorem of $F$. A wff of the form $(A \wedge (\sim A))$ is called a *contradiction*. A system is *consistent* if it is not inconsistent.

Since the set of theorems of a first-order system may not be decidable, a system may well be inconsistent without our knowing it. The following shows why inconsistent systems may cause trouble.

THEOREM 1. *A system F is inconsistent if and only if every wff of F is a theorem.*

*Proof.* If $F$ is inconsistent, let $(A \wedge (\sim A))$ be a provable contradiction. Let $X$ be any wff. Now, $\vdash (A \wedge (\sim A)) \supset X$, since this is a tautology. Hence, since $\vdash (A \wedge (\sim A))$, we obtain $\vdash X$ by *modus ponens*. But $X$ was any wff, and so every wff is provable.

Conversely, if every wff of $F$ is provable, then let $X$ be any wff. $(X \wedge (\sim X))$ is also a wff and therefore provable. Thus, a contradiction is provable in $F$.

Copyrighted material

In an inconsistent system, everything is provable.

Now let us think again how we defined the notion of a model. A model is a certain kind of interpretation, one which makes every proper axiom of the given system true. As we have already observed in connection with validity, our two rules of inference, *modus ponens* and generalization, both preserve truth. That is, they yield true wffs when applied to true wffs (where truth is defined relative to any given interpretation). Thus, all the theorems of any system will be true in any model of the system. This yields the following theorem:

THEOREM 2. *No inconsistent system F has a model.*

*Proof.* Let $F$ be inconsistent. Then some contradiction $(A\wedge(\sim A))$ is a theorem of $F$. Now, the wff $\sim(A\wedge(\sim A))$ is a tautology and thus universally valid. It is true in every interpretation. But, in any interpretation, $\sim(A\wedge(\sim A))$ is true if and only if $(A\wedge(\sim A))$ is false. Thus, $(A\wedge(\sim A))$ is false in every interpretation. If $F$ had a model $\langle D, g\rangle$, then every theorem of $F$, and thus $(A\wedge(\sim A))$, would be true in $\langle D, g\rangle$. But $(A\wedge(\sim A))$ is false in every interpretation and thus false in $\langle D, g\rangle$. Thus, $(A\wedge(\sim A))$ would be both true and false in $\langle D, g\rangle$. But no wff $X$ can be both true and false under an interpretation for it is impossible for every sequence to satisfy $X$ and no sequence to satisfy $X$ (remember that $D$ must be nonempty and so there are sequences). Thus, the assumption that $F$ has a model is contradictory and $F$ has no model.

COROLLARY. *If F has a model, it is consistent.*

*Proof.* This is the contrapositive of Theorem 2.

The wffs that are false in every interpretation are called *logically false* wffs. Obviously, the negation of every logically false wff is logically true and the negation of every logically true wff is logically false.

Now, for every closed wff $X$ of any first-order system, $X$ is either true or false (and not both) under any interpretation. Furthermore, $\sim X$ is false if and only if $X$ is true. It is natural for us to think of the theorems of a first-order system as being the set of truths under some interpretation. But this will be possible only if the system is *complete*; i.e. if, for every closed wff, either $\vdash X$ or $\vdash \sim X$. Of course, any inconsistent system is complete. We are interested, however, in consistent, complete systems. We now prove that any predicate calculus is a consistent but incomplete (i.e. not complete) theory.

THEOREM 3. *Any predicate calculus is consistent.*

*Proof.* Given any wff of a first-order system, we define its *associated statement form*, abbreviated asf. We obtain the asf of a given wff $X$ in a purely formal manner by (1) suppressing all terms and quantifiers of the wff together with accompanying commas and parentheses; (2) replacing each predicate letter by a statement letter of the system **P**, replacing everywhere the same predicate letter by the same statement letter, and using different statement letters for different predicate letters; (3) replacing the remaining parentheses by brackets, left brackets for left parentheses and right brackets for right parentheses. (We order the replacement of predicate letters by statement letters by ordering the predicate letters of $X$; first according

Copyrighted material

to argument number, and then according to subscript number within each class of those having the same argument number. We then begin by replacing the first predicate letter by $x^*$, the second by $x^{**}$, and so on, according to our other restrictions.) For example, the associated statement form of $(x_1)(A_1^2(f_1^1(x_1), x_2) \supset A_2^1(x_4))$ is $[x^{**} \supset x^*]$. What we obtain is a wff of **P** in every case.

Now, observe that the asf of any of our logical axioms is a tautology. For the tautologies this is obviously true. For axioms of the second type we just get a form $[X \supset X]$, and for axioms of the third type the form $[[X \supset Y] \supset [X \supset Y]]$. In each case we get a tautology.

Furthermore, it is clear that if the asf of $A$ is a tautology and if the asf of $(A \supset B)$ is a tautology, then the asf of $B$ must be a tautology. Thus, the rule of *modus ponens* preserves the tautological property of the asf.

Similarly, the rule of generalization preserves the tautological property of the asf. In fact, the asf of $A$ is the same as the asf of $(x)A$. It thus follows that all theorems of any predicate calculus will be such that their asf is a tautology. But the asf of any contradiction $(A \wedge (\sim A))$ is not a tautology and so no contradiction can ever be a theorem of a predicate calculus. Hence, every predicate calculus is consistent.

THEOREM 4. *No predicate calculus is a complete theory.*

*Proof.* Let $A_m^n$ be some predicate letter of a predicate calculus $F$ (there must be at least one). Consider the wff

$$(x_1)(x_2)\ldots(x_n)A_m^n(x_1, x_2, \ldots, x_n).$$

This wff, call it $X$, is closed. Its asf is simply $x^*$. This is not a tautology and so $X$ is not a theorem of the predicate calculus $F$. The asf of $(\sim X)$ is $[\sim x^*]$ and this is not a tautology either. Thus, neither $X$ nor $\sim X$ are theorems of the predicate calculus $F$. But $F$ was any predicate calculus and so no predicate calculus is a complete system.

The method of proof of Theorem 3 and Theorem 4 hinges on the fact that having a tautology for an asf is a necessary condition for a wff to be a theorem of a predicate calculus (and thus a valid wff). This condition is not sufficient, however, or we would have a decision method for the predicate calculus. That is, there are wffs whose asf is a tautology, yet are not valid wffs. Of course, necessity does give us a negative test which is of some value.

We recall that a formal system is said to be axiomatic (or axiomatized) if its set of axioms is a decidable set. Since the logical axioms of any first-order system form a decidable set, a first-order system $F$ is axiomatic if its set of proper axioms is decidable. A first-order system $F$ is axiomatizable if there is another first-order system $F'$ with the same wffs and the same theorems of $F$ and whose proper axioms form a decidable set. In short, a first-order system is axiomatizable if there exists an axiomatization of it which yields the same set of theorems.

*Exercise.* Give an example of a wff which is not valid but whose asf is a tautology.

Notice that any consistent, complete, axiomatizable system is decidable. We merely order all formal proofs in some convenient way and grind them out one by one. Any wff $X$ is provable if and only if its universal closure is provable. Given any wff $X$, we take its universal closure $\bar{X}$ and $\sim \bar{X}$. Since these are closed wffs, we must eventually turn up a proof either of

Copyrighted material

$\bar{X}$ or $\sim \bar{X}$. If $\vdash \bar{X}$, then $\vdash X$ by application of *modus ponens* and our axiom of type 2 for a first-order system. If $\vdash \sim \bar{X}$, then $\bar{X}$ cannot be a theorem and hence neither can $X$. Hence, we can decide whether a given wff is a theorem or not.

Most interesting first-order theories of any great degree of expressiveness are neither complete nor decidable. Again, Chapter 6 contains a more detailed discussion.

The Corollary to Theorem 2 stated that every first-order theory which has a model is consistent. Obviously, an interesting question is whether the converse holds here. Does every consistent first-order theory have a model? The answer is "yes" and the theorem is a very profound one indeed. It says, essentially, that we cannot talk consistently without talking about something. We state the following theorem:

THEOREM 5. *Every consistent first-order theory has a model.*

For proofs of this, consult Robinson [1] or Mendelson [1].

This theorem is called the *completeness theorem of first-order logic*. What is asserted to be complete is not any given first-order system, but rather our logical axioms and rules which together constitute the underlying logic of all first-order systems. Theorem 5 says that, given any first-order theory $F$ which has no model, then we can establish this fact using only our logical axioms and rules by formally deducing a contradiction in $F$. In other words, our logic is completely capable of detecting any theory which does not have a model. (Of course, we may not be clever enough to find the proof of contradiction, but that is another matter entirely.)

It is interesting that Theorem 1 of Section 1.4, which states that every valid formula of a first-order theory is a theorem of it, can be deduced from Theorem 5. To see this, we prove as lemmas several theorems which do not depend on Theorem 1 of Section 1.4.

Recalling the notion of dependence of formulas in proofs examined in Section 1.2, we now establish a fundamental result of proof theory called the *deduction theorem.*

THEOREM 6. *If $A \vdash_F B$, and if no application of the rule of generalization applied to a wff which depends on $A$, and in which the quantified variable was free in $A$, has occurred in the proof, then $\vdash_F A \supset B$.*

*Proof.* Again we apply induction to the length of the proof in question. If the proof is of length 1, then the proof consists of the one member $B$. If $B$ is $A$, then $(A \supset A)$ is a tautology and thus a theorem of $F$. This gives $\vdash_F A \supset B$. If $B$ is an axiom of $F$, we deduce $\vdash_F A \supset B$ from $B$ itself and the tautology $(B \supset (A \supset B))$.

We assume that the proposition holds for deductions of length less than $n$ and thus consider a deduction of length $n$. If $B$ is $A$ or an axiom, we have the same argument as just given. Suppose that $B$ is inferred from prior wffs $C$ and $(C \supset B)$ by *modus ponens.* $C$ and $(C \supset B)$ are the result of deductions from $A$ of length less than $n$, since they precede $B$, and so an application of the induction hypothesis yields $\vdash_F A \supset C$ and $\vdash_F A \supset (C \supset B)$. Now, $(A \supset C) \supset ((A \supset (C \supset B)) \supset (A \supset B))$ is a tautology as can be checked by the truth-table method. Applying *modus ponens* twice, we obtain $\vdash_F A \supset B$. Finally, suppose $B$ is obtained from a prior wff $C$ by the rule of generalization. Then $B$ is $(x)C$ for some variable $x$. By the conditions assumed in the hypotheses of our theorem, either $C$ does not depend on $A$ or $x$ is not free

Copyrighted material

in $A$. If $C$ does not depend on $A$, then $\vdash_F C$ by Corollary 2 of Theorem 1, Section 1.2. Since $B$ is obtained from $C$ by generalization, we have immediately $\vdash_F B$ and thus $\vdash_F A \supset B$, again using the tautology

$$(B \supset (A \supset B)).$$

If $C$ does depend on $A$, then $x$ is not free in $A$. Also, $C$ is the result of a deduction from $A$ of length less than $n$ and so the induction hypothesis yields $\vdash_F A \supset C$. Applying generalization, we obtain $\vdash_F (x)(A \supset C)$, where $x$ is not free in $A$. Now, $((x)(A \supset C) \supset (A \supset (x)C))$ is a logical axiom, since $x$ is not free in $A$, and so *modus ponens* yields $\vdash_F A \supset (x)C$; that is, $\vdash_F A \supset B$, since $B$ is $(x)C$. Our proposition holds for $n$, and the corollary is established.

COROLLARY. *If* $A \vdash_F B$ *and if* $A$ *is closed, then* $\vdash_F A \supset B$.

*Proof.* The hypotheses of Theorem 6 are immediately satisfied.

Theorem 6 is called the deduction theorem because it establishes a fundamental connection between our metamathematical relation of deducibility "$\ldots \vdash ---$", the formal symbol "$\supset$" and the metamathematical relation "$\vdash \ldots \supset ---$". Notice that the two relations "$\ldots \vdash ---$" and "$\vdash \ldots \supset ---$" are not the same, since Theorem 6 contains certain restrictive conditions. These restrictive conditions are necessary, for we have $A \vdash_F (x)A$ in any first-order system $F$. However, it is generally not true that $A \supset (x)A$ is provable if the variable $x$ is free in $A$.

By *modus ponens*, $\vdash_F A \supset B$ implies $A \vdash_F B$ in any first-order system $F$, and so the relation "$\vdash_F \ldots \supset ---$" is strictly stronger than the relation "$\ldots \vdash_F ---$" in most first-order systems $F$. (In contradictory systems, of course, everything is provable.)

We use the notation $A \models_F B$ to stand for "$A \vdash_F B$ and the hypotheses of the deduction theorem are satisfied". We thus have $A \models_F B$ if and only if $\vdash_F A \supset B$ in any first-order system $F$. In fact, $X$, $A \models_F B$ it and only it $X \models_F (A \supset B)$ by the proof of theorem 6.

THEOREM 7. *Let $F$ be any first-order system and $X$ a closed wff which is not a theorem of $F$. Then if the wff $(\sim X)$ is added to the axioms of $F$, the system $F'$ thus obtained is consistent.*

*Proof.* Suppose $F'$ is inconsistent. Then every wff, in particular $X$, is provable in $F'$. To say that $X$ is provable in $F'$ means precisely that $(\sim X) \vdash_F X$ holds. But $(\sim X)$ is closed and so, by the Corollary to Theorem 6 we obtain $\vdash_F ((\sim X) \supset X)$. Applying *modus ponens* to this and the tautology $((\sim X) \supset X) \supset X$, we obtain $\vdash_F X$ which contradicts our hypotheses. Hence, $F'$ must be consistent.

Notice that the system $F$ of Theorem 7 is consistent, since there is a wff, namely $X$, which is not a theorem of $F$.

We can now see easily how the completeness theorem implies that all valid wffs must be theorems of any system. Let $F$ be any first-order system and let $X$ be any valid wff of $F$ which is not provable. The universal closure $\bar{X}$ of $X$ is not provable either. But $\bar{X}$ is closed, and so by Theorem 7, we can add $(\sim \bar{X})$ as an axiom and obtain thereby a new system $F'$, which is consistent. Since $F'$ is consistent, it has a model by Theorem 5. A model of $F'$ must make all the axioms of $F'$, in particular $(\sim \bar{X})$, true. But if $(\sim \bar{X})$ is true, $\bar{X}$ must be false. Thus,

Copyrighted material

$X$ is not true, because any wff $X$ is true if and only if its universal closure $\bar{X}$ is true. But $X$ is valid and thus true under every interpretation. Hence, $X$ is both true and not true, a contradiction establishing that $X$ must have been provable in the first instance. Since $X$ is any valid formula, it follows that all valid formulas are theorems of $F$ and Theorem 1 of Section 1.4 is proved.

Theorem 1 of Section 1.4 is sometimes referred to as the "weak completeness" of our logic. The notion of completeness there is that our logical axioms and rules are sufficient to enable us to prove all logically valid formulas as theorems in any first-order theory.

Theorem 7 also has the following interesting corollary:

COROLLARY. *Let $F$ be a first-order theory and let $X$ be any wff of $F$ true in every model of $F$. Then $\vdash_F X$.*

*Proof.* Assume that $X$ is true in every model of $F$, but that it is not provable in $F$. $F$ is consistent, since not every wff is provable. Moreover, the universal closure $\bar{X}$ of $X$ is not provable in $F$, since $\vdash \bar{X} \supset X$ in any system. Thus, we can add $(\sim \bar{X})$ as an axiom and obtain a consistent system $F'$. But $F'$ has a model $\langle D, g\rangle$, since it is consistent and $(\sim \bar{X})$ must be true in this model. Moreover, every axiom of $F$ is an axiom of $F'$ and so $\langle D, g\rangle$ is also a model for $F$. But $X$, and thus $\bar{X}$, is true in every model of $F$ and hence in $\langle D, g\rangle$. Thus, $(\sim \bar{X})$ is false in $\langle D, g\rangle$, contradicting our first conclusion. Hence, our assumption of the unprovability of $X$ is false and $\vdash_F X$.

*Exercise.* Prove that, in any first-order system $F$, $A \vdash_F B$ if and only if $B$ is true in every model of $F$ in which $A$ is true.

Theorem 7 brings up the important notion of one system being an *extension* of another.

**Definition 3.** A first-order system $F'$ is an *extension* of a first-order system $F$ if the alphabet and theorems of $F$ are each subsets of the alphabet and theorems of $F'$. We write $F \subset F'$ to mean that $F'$ extends $F$. If $F \subset F'$ while both have the same alphabet, then we say $F'$ is a *simple* extension of $F$. If $F \subset F'$ and there are wffs or theorems of $F'$ that are not wffs or theorems of $F$, then the extension $F'$ is said to be *proper*.

The method of Theorem 7 can be used to prove the following useful theorem:

THEOREM 8. *Every consistent first-order theory has a consistent, complete, simple extension.*

*Proof.* Since the wffs of any first-order theory are denumerable, we begin with some fixed enumeration of all closed wffs of $F$. Let $B_1, B_2, \ldots, B_n \ldots$ be the enumeration in question. Now $B_1$ is either provable or not. If it is, then we proceed to $B_2$. If it is not, then we add $(\sim B_1)$ as an axiom and obtain by Theorem 7 a consistent simple extension of $F$ (which may or may not be proper). We then proceed to $B_2$. In either case, we let $F_1$ be the system that results after our consideration of $B_1$. It is a consistent simple extension (perhaps proper) of $F$. We do the same with $B_2$, and so on. We let $F_n$ be the system that results after considering $B_n$, and let $P_n$ be the set of axioms of $F_n$. $F_n$ is a consistent extension of $F_{n-1}$ for all $n \geqslant 2$, and $F_1$ is a consistent extension of $F$. Let $F_\infty$ be the system which has the same

Copyrighted material

symbols and wffs as $F$ and whose axioms are the set

$$P_\infty = \bigcup_{n=1}^{\infty} P_n.$$

Now, $F_n$ must be consistent, for otherwise we can deduce a contradiction in $F_\infty$. But a formal deduction is of finite length and thus involves only a finite number of axioms of $F_\infty$. These are all contained in $F_i$ for some $i$, and so a contradiction must be forthcoming in $F_i$. But all of the systems $F_n$ are consistent and so we have a contradiction establishing the consistency of $F_\infty$.

Finally, $F_\infty$ must be complete, for it is a simple extension of $F$ and thus has the same wffs and closed wffs as $F$. For every closed wff $B_n$ of $F_\infty$, either $\vdash_{F_n} B_n$ or $(\sim B_n)$ is added as an axiom to $F_n$. Thus, either $B_n$ or $(\sim B_n)$ is provable in $F_\infty$, which is an extension of all the $F_n$. This completes the proof.

**Definition 4.** Let $S$ be the set of wffs of any first-order theory $F$, and let $X$ be any subset of $S$. We say that the set $X$ *has a model* if the first-order theory $F^*$, having $S$ as its set of wffs and $X$ as its proper axioms, has a model. In other words, $X$ has a model if there is some interpretation (not necessarily a model) of $F$ in which all the wffs of $X$ are true.

**Definition 5.** A set $X$ of wffs of a first-order theory $F$ is said to be *inconsistent* if the theory $F^*$, having the same wffs as $F$ and $X$ as its set of proper axioms, is inconsistent. $X$ is inconsistent if and only if it has no model.

The fact, useful in the proof of Theorem 8, that proofs are of finite length means that any inconsistent set $X$ must be inconsistent on some finite subset. For if a contradiction is deducible from the hypotheses $X$, it must be deducible from some finite subset, since proofs are of finite length. Using Theorem 5, we thus obtain the *compactness theorem*:

THEOREM 9. *If a set $X$ of wffs is such that every finite subset of it has a model, then $X$ has a model.*

*Proof.* If every finite subset of $X$ has a model, then every finite subset of $X$ is consistent. The set $X$ is thus consistent, since it is not inconsistent on any finite subset. But every consistent set $X$ has a model, and our theorem is established.

The compactness theorem has many useful applications to algebra and analysis. The interested reader should consult A. Robinson [1] and [2].

Finally, we state (without proof) a modern form of the famous Löwenhein–Skolem theorem:

THEOREM 10. *If a system $F$ has a model, then it has a finite or denumerable model; that is, a model $\langle D, g\rangle$ in which the set $D$ is finite or denumerable. Furthermore, if $D'$ is any set with cardinality greater than or equal to $D$ and if $F$ has a model with domain $D$, then $F$ has a model with domain $D'$.*

COROLLARY 1. *Every consistent first-order theory has a denumerable model.*

Copyrighted material

*Proof.* If the system is consistent, it has a model and thus, by Theorem 10, a finite or denumerable model. If the model is in some finite domain $D$, then any denumerable domain $D'$ has a cardinality greater than or equal to $D$ and the system thus has a model with domain $D'$.

COROLLARY 2. *Every consistent, first-order theory has models of every infinite cardinality.*

*Proof.* If the theory is consistent, it has a denumerable model, and every infinite cardinality is greater than or equal to the cardinality of a denumerable domain.

The reader should be apprised of the fact that extending models of a given system to domains of a higher cardinality is essentially trivial. It amounts to showing that, once we have a given model, we can throw in any number of extra objects without disturbing the original model. The more significant part of Theorem 10 is thus the part asserting that any consistent system has a countable model. It is in this form that the Löwenhein–Skolem theorem is most often stated. For a proof of this part of Theorem 10, the reader should consult Church [3].

Theorem 10 is basically true because the set of wffs of any first-order theory is denumerable. Theorem 10 has some surprising consequences, for in later chapters we shall deal with first-order theories which are set theories, and in which we can prove the existence of uncountable sets. Yet, these set theories have denumerable models as all first-order theories do. Chapter 6 contains a detailed discussion of this point.

*Exercise* 1. Prove that, for any first-order theory $F$, and any set $X$ of hypotheses, $X \models_F y$ if and only if, for every model $\langle D, g \rangle$ of $F$, every sequence $s$ which satisfies every wff in $X$ also satisfies $y$.

*Exercise* 2. In any first-order theory $F$, a proper axiom $p_1$ of $F$ is said to be *independent* if it is not provable in the theory $F^*$ obtained from $F$ by deleting $p_1$ as an axiom. Prove: $p_1$ is independent in $F$ if and only if there exists a model of $F^*$ in which $p_1$ is false. Such a model is called an *independence model* for $p_1$.

**Definition 6.** Any set $X$ of wffs of a first-order language $L$ is said to be *independent* if each wff $y$ in $X$ is independent (in the sense of Exercise 2 above) in the theory having the same nonlogical symbols as $L$ and having $X$ for its set of proper axioms.

This notion will be useful to us in the future.

Independence is a useful property of an axiom set as a point of simplicity and economy, but it is not so crucial from a logical standpoint as are other properties such as consistency. In fact, there are often times when a nonindependent axiomatization is more elegant and more readily understandable.

## 1.6. Rules of logic; natural deduction

The form we have given to the axioms and rules of the predicate calculus is especially adequate for proving metatheorems about first-order theories. Some of the metatheorems we have stated depend on nothing more than simple principles, such as mathematical induction, for their proof. Others involve highly nonconstructive principles of set wfeory. As we have already mentioned, it would be inappropriate to use these highly nonconstructive metatheorems to



Copyrighted material

prove theorems in first-order systems which are themselves involved with expressing set-theoretical principles. The prime object of study in this book is precisely such "foundational" systems, and so we are interested, for our purposes, in developing the technique of formal reasoning within first-order logic. Of course we still may refer to our nonconstructive metatheorems in talking about a particular system, especially in trying to get some idea of what a model of it looks like. But such metamathematical discussion must be clearly distinguished from the purely formal and constructive approach of proving theorems within the system.

As it turns out, our form of the axioms and rules of the predicate calculus is not particularly useful for the technique of formal deduction. It is too far removed from intuitive reasoning for this. Witness, for example, the samples of formal deduction given in Section 1.4. We can see the beginnings of a more natural kind of deduction with the introduction of the deduction theorem, Theorem 6 of Section 1.5.

Let us reflect on how one intuitively proves a proposition of the form "If $X$, then $Y$" where $X$ and $Y$ are statements. Traditionally, one begins by assuming $X$ is true; that is, taking $X$ as an hypothesis and showing that the truth of $Y$ follows. One establishes $\vdash X \supset Y$ by establishing $X \vdash Y$. The deduction theorem tells us under what conditions this method is valid. Notice that our proof in Section 1.4 of $\vdash (x_i)(A \vee B(x_i)) \supset (A \vee (x_i)B(x_i))$, $x_i$ not free in $A$, did not proceed by this method, since we had not yet proved the deduction theorem.

Another natural method of intuitive logic not directly provided for by our axioms and rules of the predicate calculus is the handling of existential quantification. How might we establish a proposition of the form $(Ex)\,B(x) \supset A$? We might first assume $(Ex)\,B(x)$ as an hypothesis. Then we might say, "since there is some $x$ such that $B$ is true, call it $a$; that is, let it be designated by some arbitrary new dummy constant $a$". Assuming, then, that $B(a)$ holds (where $B(a)$ result from $B(x)$ by replacing $a$ for $x$ in all its free occurrences), we deduce $A$, where $A$ does not contain the dummy constant $a$. We then conclude that $\vdash (Ex)\,B(x) \supset A$.

It is possible to prove that just such a procedure as this one is permissible with the rules we have already presented. We need first to state a few definitions:

**Definition 1.** Let $F$ be any first-order system. By a *dummy constant letter* for $F$, we mean any constant letter which is not a constant letter of $F$.

A first-order theory $F$ may well have a countably infinite number of constant letters, thus using all of the constant letters $a_1$, $a_2$, etc. However, we can always suppose that $F$ has an unlimited supply of dummy constant letters. We can, for example, choose all the odd-numbered constant letters $a_1, a_3, \ldots$ to be in the theory, and thus leave the even-numbered ones available as dummy constant letters.

**Definition 2.** Given a first-order system $F$, let $F'$ be an extension of $F$ that is the same as $F$ except for containing some (any nonzero finite number of) constant letters not in $F$. By a *dummy well-formed formula* of $F$, abbreviated dwff, we shall mean a wff of any such $F'$ which is not a wff of $F$.

A dwff of $F$ is exactly like a wff except for containing at least one dummy constant letter as a term. From now on, the word "formula" will be used to mean either a wff or a dwff.

The purpose of introducing dummy constants and dwffs into our logic is to allow for the more direct and flexible handling of the existential quantifier as indicated in our brief discus-

Copyrighted material

sion above. The operation of removing the existential quantifier by substituting a new dummy constant in place of the existentially quantified variable will be called "rule $c$" or the "choice rule". Deductions using this rule will be called $c$-deductions.

**Definition 3.** By a *c-deduction from the hypotheses* $X$ we mean a finite list of wffs or dwffs of $F$ such that, for each member $Y$ of the list, one of the following conditions holds: (1) $Y$ is in $X$, (2) $Y$ is an axiom of $F$ or $Y$ is a dwff which is a logical axiom (of the extension $F'$) all of whose dummy constants have already appeared in the proof, (3) there is a prior member of the list of the form $(Ex)\,A(x)$ and $Y$ is of the form $A(b)$ where $b$ is a dummy constant which does not appear in any dwff of the list prior to $Y$ (in other words, $b$ is *new* in the proof), (4) $Y$ is inferred from prior members of the list by MP or by UG except that UG is never applied to a variable $x$ which is free in a formula of the form $(Ez)\,A(z)$ to which operation (3) has been previously applied and on which the given formula $c$-depends (see following definition). We write $X \vdash_c A$ to stand for "$A$ is the last line of a $c$-deduction from the hypotheses $X$".

**Definition 4.** An occurrence of a dwff or wff *c-depends* on an occurrence of a wff or dwff if it depends on that occurrence of the formula in the sense of Definition 4, Section 1.2 in any deduction involving possible applications of rule $c$. Formulas *c-depend* on other formulas if at least one occurrence of one $c$-depends on an occurrence of the other.

From now, but only to the end of Theorem 1 below, we restrict the notion of dependence to apply only to uses of the rules MP and UG. This is only to emphasize the explicit uses made of rule $c$ and to examine precisely the relationship between deductions which involve rule $c$ and those which do not. Also, the deduction theorem (Theorem 6 of Section 1.5) has so far been proved to hold only where the notion of dependence involves MP and UG.[†]

In this terminology, the restrictions on the application of UG in Definition 3 insure, in particular, that UG is never applied to a variable $x$ free in any formula $B(b)$ which has been immediately inferred by rule $c$ from a prior formula, and on which the given formula depends.

THEOREM 1. *In any first-order system $F$, if, for some set of hypotheses $X$, $X \vdash_c A$ and none of the $X$ nor $A$ are dwffs, then $X \vdash A$ where* UG *is applied to some formula c-dependent on and variable free in some hypothesis in $X$ only if there was such an application of* UG *in the original c-deduction.*

*Proof.* We make use of the following lemma which we prove using only our logical rules and axioms and the deduction theorem.

LEMMA. $\vdash (z)\,(B(z) \supset A) \supset ((Ez)\,B(z) \supset A)$ *where $z$ is not free in $A$, in any system $F$.*

*Proof.* 1. $(z)\,(B(z) \supset A))$ Hyp
2. $B(z) \supset A$ 1, Log $Ax$ 2, MP
3. $\sim A \supset \sim B(z)$ 2, Taut, MP
4. $(z)\,(\sim A \supset \sim B(z))$ 3, UG ([3] depends on [1] but $z$ is not free in [1])
5. $\sim A \supset (z)\,(\sim B(z))$ 4, Log $Ax$ 3 ($z$ not free in $\sim A$), MP
6. $(Ez)\,B(z) \supset A$ 5, Taut, MP

We have now established [1] $\vDash$ [6] and so $\vdash_F$ [1] $\supset$ [6] by the deduction theorem.

[†] A deduction theorem involving rule $c$ is forthcoming in Theorem 9 of this section.

Copyrighted material

Returning now to the proof of the theorem, we let $(Ey_1)\, B_1(y_1), \ldots, (Ey_k)\, B_k(y_k)$ be the list of the wffs or dwffs to which the choice rule has been applied in the proof in order of application of the rule. Let $b_1, \ldots, b_k$ be the new dummy constants thereby introduced. Obviously, $X, B_1(b_1), \ldots, B_k(b_k) \vdash A$ since the choice rule will have served only to give us the formulas $B_i(b_i)$. In fact, since we have not applied UG to any variable free in any $B_i(b_i)$ and to a formula which depends on any $B_i(b_i)$, the conditions of the deduction theorem are met and we have $X, B_1(b_1), \ldots, B_{k-1}(b_{k-1}) \vdash B_k(b_k) \supset A$ by the deduction theorem.

Let us now replace the dummy constant $b_k$ everywhere it appears in the proof by an entirely new variable $z$ (this is possible since the number of variables in the proof is necessarily finite). Once we have completed this formal replacement operation, we will still have a valid proof since neither the form of axioms nor the validity of the application of the rules MP and UG will have been changed. We thus have $X, B_1(b_1), \ldots, B_{k-1}(b_{k-1}) \vdash (B_k(z) \supset A)$ where $z$ is not free in $A$. (Since $A$ was, by hypothesis, a wff, no dummy constants appeared in $A$, which thus remains unchanged by our replacement operation. The wffs in $X$ and the formulas $B_i(b_i)$, $1 \leqslant i \leqslant k-1$ are also unchanged since $b_k$ does not appear in them.) Applying UG we obtain $X, B_1(b_1), \ldots, B_{k-1}(b_{k-1}) \vdash (z)\,(B_k(z) \supset A.)$ Notice that $z$ was entirely new to the original proof, and so $z$ does not occur in any of the $B_i(b_i)$, $1 \leqslant i \leqslant k-1$, nor in any of the hypotheses $X$ actually appearing in the deduction.

Appealing now to the lemma, we have $X, B_1(b_1), \ldots, B_{k-1}(b_{k-1}) \vdash ((Ez)\, B_k(z) \supset A)$. But, we also have $X, B_1(b_1), \ldots, B_{k-1}(b_{k-1}) \vdash (Ey_k)\, B(y_k)$ since this latter formula was the last one to which the choice rule was applied in the original deduction. But, $B(y_k)$ is similar to $B_k(z)$. Hence, $\vdash (Ey_k)\, B_k(y_k) \equiv (Ez)\, B_k(z)$, (see the exercise on page 32), and applying *modus ponens* twice, we obtain $X, B_1(b_1), \ldots, B_{k-1}(b_{k-1}) \vdash A$.

By successively eliminating, in the same way, the other hypotheses $B_i(b_i)$, we arrive at the desired conclusion. Moreover in the final deduction of $A$ from the hypotheses $X$, there will be an application of UG to some formula $c$-dependent on and variable free in some hypothesis in $X$ only if there was such an application of UG in the original $c$-deduction (the application of UG to the variable $z$ does not violate this since $z$ was a completely new variable and thus one which did not appear in any of the hypotheses $X$ actually used in the deduction).

The reason we need constant letters outside a system $F$ for dummy constant letters is that the proper axioms of $F$ may well assume special properties about the constant letters of $F$. If this is so, and certainly there is no reason to have constant letters in a system unless some assumptions are made about them, then the constant letters of $F$ are not really ambiguous names at all, but names of specific objects. To use such constant letters in removing the existential quantifier would be similar to reasoning that, because we have proved that there are irrational numbers, then some particular constant, such as zero, is irrational. It would be an instructive (and not difficult) exercise for the reader to see exactly where the proof of Theorem 1 breaks down if we admit constants other than dummy constants in applications of rule $c$.

What Theorem 1 tells us is that we can use rule $c$ as freely as we want in deductions, as long as we observe the restrictions on UG. If we proceed in this manner, every proof using rule $c$ can be transformed into a proof without it, thus a proof using our original logical axioms and rules.

Copyrighted material

*Exercise.* Use rule $c$ to prove in the predicate calculus that $\vdash (B \supset (Ex)A) \equiv (Ex)(B \supset A)$ where $x$ is not free in $B$.

We now want to go even further in the direction of a more natural deduction by presenting a proof system for first-order systems which uses only rules and no logical axioms at all (except for tautologies). We will subsequently show that the deductive power of our new proof procedure is the same as with the logical axioms and rules.

**Definition 5.** Given a first-order system $F$, by a *proof in F* we now mean a finite sequence $B_1, \ldots, B_k$ of wffs or dwffs of $F$ together with a *justification* for each *line* of the proof. By a line of the proof we mean an ordered pair $\langle n, B_n \rangle$ where $B_n$ is the $n$th member of the sequence. A justification is a statement in English which accompanies a given line of the proof. The line $\langle n, B_n \rangle$ of a proof is called the *nth line* of the proof, and $B_n$ is called the *formula of* the $n$th line. For each line $\langle n, B_n \rangle$ of a proof in $F$, one and only one of the following must hold: (1) $B_n$ is a wff which is a proper axiom of $F$ and the justification for the $n$th line consists of designating which of the proper axioms of $F$ $B_n$ is; (2) $B_n$ is a tautological formula and it is a wff oı else a dwff whose dummy constant letters each appear in some previous line of the proof. The justification for the $n$th line is that $B_n$ is a tautology (we write "Taut"); (3) $B_n$ is a wff or else a dwff whose dummy constant letters each appear in some previous line of the proof, and the justification for the $n$th line is that $\langle n, B_n \rangle$ is an hypothesis (we write "$H$"); (4) $B_n$ is immediately inferred from the formulas of explicitly designated prior lines in the proof by one of the rules of inference given below, and the justification for the $n$th line consists in designating the prior lines and the rule of inference in question. We say also that the line $\langle n, B_n \rangle$ is *immediately inferred* from the explicitly designated prior lines in question.†

Notice that any finite sequence of wffs is capable of being considered a proof in a trivial way. Just let the justification for each line be that it is an hypothesis. Of course it will not be true that any sequence of wffs can be a proof if we insist on a particular type of justification. We will use this fact to generate exactly the same theorems with our new rules as with our old rules and axioms.

We now turn to the statement of our rules of inference in order to complete Definition 5. In the following, the metavariables represent formulas (wff or dwff) unless a specific restriction is indicated. A scheme of the form

$$\frac{X, Y, \ldots}{Z}$$

means that we can immediately infer the formula $Z$ from the formulas $X$, $Y$, etc.

$$\text{MP: } \frac{(A \supset B), A}{B}; \qquad eE\text{: } \frac{(Ex)\,A(x)}{A(b)}$$

where $b$ is some dummy constant letter not appearing in $(Ex)\,A(x)$ nor in any wff or dwff

† We will sometimes abuse our language by identifying a line of a proof with the formula of that line. In particular, we will sometimes refer to a formula $B_n$ as an hypothesis when it is really the line $\langle n, B_n \rangle$ which is the hypothesis. No confusion will result if we keep in mind that $B_n$ may be the formula of two different lines of a proof and these lines may have different justifications.

Copyrighted material

previously in the proof, and $A(b)$ represents the result of substituting $b$ for $x$ at all free occurrences of the variable $x$ in $A(x)$.

$$e\forall: \frac{(x)\,A(x)}{A(t)}$$

where $t$ is any term of $F$ free for $x$ or else a dummy constant letter previously introduced into the proof by an application of $eE$, and $A(t)$ represents the result of substituting $t$ for the variable $x$ in all the free occurrences of $x$ in $A(x)$.

$$iE: \frac{A(t)}{(Ex)\,A(x)}$$

where $t$ is any term of $F$ which is free for $x$ in $A(x)$ or any dummy constant letter, and $A(t)$ is the result of substituting $t$ for the variable $x$ in all the free occurrences of $x$ in $A(x)$.

$$i\forall: \frac{A(x)}{(x)\,A(x)}$$

where the variable $x$ is not free in any hypothesis on which $A(x)$ depends, and where $x$ is not free in any wff or dwff $(Ey)\,B(y)$ to which rule $eE$ has been previously applied in the proof unless $A(x)$ is a wff and depends only on hypotheses that are wffs. (In this latter case neither $A(x)$ nor any hypotheses on which it depends can contain dummy constant letters.)

$$eH: \frac{A,\ B}{(A \supset B)}$$

where $A$ is any hypothesis on which $B$ depends, occurring before $B$ in the proof.

In each of the above rules in which "previously" is used, it is understood to mean "previous to the line $\langle n, B_n\rangle$ which is being inferred from other (necessarily prior) lines".

**Definition 6.** Let $\langle n, B_n\rangle$ occur as an hypothesis in a given proof in a first-order system $F$. We say that the line $\langle i, B_i\rangle$ of the proof, $n \leqslant i$, *depends on the hypothesis* $\langle n, B_n\rangle$ if and only if (1) $n = i$, or (2) $\langle i, B_i\rangle$ is immediately inferred from prior lines of the proof at least one of which depends on $\langle n, B_n\rangle$, except that $\langle i, B_i\rangle$ does not depend on $\langle n, B_n\rangle$ if $B_i$ is of the form $(B_n \supset B_k)$, $n < k$, and the justification for $\langle i, B_i\rangle$ is that it is inferred from $\langle n, B_n\rangle$ and $\langle k, B_k\rangle$ by $eH$.

Notice that Definition 6 defines dependence in such a way that a line of a proof may depend only on a line which is an hypothesis.

**Definition 7.** A *theorem* of a first-order system $F$ is a wff of $F$ which can be obtained as the formula of the last line $\langle n, B_n\rangle$ of a proof in $F$ such that $\langle n, B_n\rangle$ depends on no hypotheses whatever.

Notice that the definition for a theorem of a first-order system excludes dwffs as theorems. Even tautologies that are dwffs of $F$ are not theorems of $F$, though wff tautologies certainly are, as always. Still, our rules permit us to introduce a dwff tautology $X$ into a proof without introducing further dependence on hypotheses (provided that the dummy constants of $X$ have been previously introduced into the proof).

In citing the rules of inference as justification for a given line of a proof, we give the numbers of the prior lines of the proof from which the given line follows and the name of the rule in

Copyrighted material

question. Notice also that a line $X$ of a proof does not necessarily depend on an hypothesis $B$ that occurs before $X$ in the proof. It depends on $B$ only if $B$ has somehow contributed to obtaining $X$, as is clear from Definition 6 of this section. Moreover, any hypothesis $B$ depends on itself. At each line $X$ of a deduction, we indicate all the hypotheses on which $X$ depends by displaying in parentheses the number of the line of each such hypothesis to the left of the given line $X$.

The names we have given the rules are meant to suggest the formal operation involved. "$i$" stands for "introduction" and "$e$" for "elimination". "$E$" stands for "existential quantifier" and "$\forall$" for "universal quantifier".

We call our new set of rules *natural deduction rules*. For the remainder of this work, most instances of formal deduction will take place in our natural deduction rule system. However, we may still appeal to the original rules and axioms in proving metatheorems about systems. We want now to see that our natural deduction rules are both *adequate*, meaning that everything provable by our old rules and axioms is provable with the natural deduction rules, and *sound*, meaning that everything provable by our natural deduction rules is provable with our original set of rules and axioms. We begin by a descriptive examination of our natural deduction system, comparing it to our original system.

The first of our natural deduction rules is just *modus ponens*. The second is the elimination of the existential quantifier in favor of a new dummy constant letter. Theorem 1 has already provided the justification for this procedure on the basis of our original rules and axioms. Notice that it is only by rule $eE$ that a given dummy constant letter can be first introduced into a proof. The rule $e\forall$ is obvious and is justified by *modus ponens* and our previous logical axioms. The rule $iE$ is also one we have previously seen to be valid from an example of formal deduction in Section 1.4. That is, we easily prove $\vdash A(t) \supset (Ex)\,A(x)$ according to our old rules and axioms where $A(x)$ and $A(t)$ are related as in the statement of rule $iE$. The rule $eH$ of hypothesis elimination is just the deduction theorem. The rule $i\forall$ of universal quantifier introduction is somewhat complicated by the various restrictions imposed on the natural deduction rules. The first restriction, that the variable in whose name universal quantification is introduced must not be free in any hypothesis on which the formula in question depends, is necessary to insure that our rule $eH$ (the deduction theorem) is valid. The other restrictions have to do with the rule $eE$ rather than $eH$. Let us take a closer look at the $eE$ restrictions in $i\forall$.

First, we notice that we have a certain flexibility in the $eE$ restriction, because we have a disjunction of two possible restrictions. One or the other of these two must be satisfied, but it is not necessary that both be satisfied. The variable $x$ in whose name universal quantification is applied must not be free in any prior wff $(Ey)\,B(y)$ to which $eE$ has been previously applied in the proof, or else the formula $A(x)$ to which quantification is applied must be a wff and depend only on wff hypotheses. If we did not have such restrictions for $eE$ in $i\forall$, we could reason falsely in the following manner:

(1) 1. $(x_1)(Ex_2)\,A_1^2(x_1, x_2)$ $H$

(1) 2. $(Ex_2)\,A_1^2(x_1, x_2)$ 1, $e\forall$

(1) 3. $A_1^2(x_1, a_1)$ 2, $eE$

(1) 4. $(x_1)\,A_1^2(x_1, a_1)$ 3, $i\forall$ (falsely!)

(1) 5. $(Ex_2)(x_1)\,A_1^2(x_1, x_2)$ 4, $iE$

6. $(x_1)(Ex_2)\,A_1^2(x_1, x_2) \supset (Ex_2)(x_1)\,A_1^2(x_1, x_2)$ 1, 5, $eH$

Copyrighted material

The only false step here is in line 4 where we introduce the universal quantifier. The variable $x_1$ is free in the wff of line 2 to which the rule *eE* was applied, and so one of our restrictions is violated. It would be permissible to violate this restriction if it were not also true that line 3, to which the rule $i\forall$ is applied in this case, contains a dummy constant letter and is thus a dwff. Thus, neither of our alternate conditions is satisfied. Below is an example of a proof in which $i\forall$ is correctly applied.

(1) 1. $(x_1)(Ex_2)(x_3)\,A_1^3(x_1, x_2, x_3)$ $H$
(1) 2. $(Ex_2)(x_3)\,A_1^3(x_1, x_2, x_3)$ 1, $e\forall$
(1) 3. $(x_3)\,A_1^3(x_1, a_1, x_3)$ 2, $eE$
(1) 4. $A_1^3(x_1, a_1, x_3)$ 3, $e\forall$
(1) 5. $(Ex_2)\,A_1^3(x_1, x_2, x_3)$ 4, $iE$
(1) 6. $(x_3)(Ex_2)\,A_1^3(x_1, x_2, x_3)$ 5, $i\forall$
(1) 7. $(x_1)(x_3)(Ex_2)\,A_1^3(x_1, x_2, x_3)$ 6, $i\forall$
8. $(x_1)(Ex_2)(x_3)\,A_1^3(x_1, x_2, x_3) \supset (x_1)(x_3)(Ex_2)\,A_1^3(x_1, x_2, x_3)$ 1, 7, $eH$

Here the rules are correctly applied. The application of $i\forall$ in line 6 satisfies both of our restrictions concerning *eE* (it is only necessary that one of the two be satisfied), and the variable $x_3$ is not free in line 1. In the application of $i\forall$ in line 7, the variable $x_1$ is free in a previous line (line 2) and to which the rule *eE* is applied (in line 3). However, when we apply $i\forall$ in line 7, all constants introduced by *eE* have been eliminated and the hypotheses (namely line 1) on which line 6 depends are also all wffs (i.e. they contain no dummy constants). Thus, our second restriction concerning *eE* is satisfied and $i\forall$ can be applied in the name of $x_1$ (again, upon required checking, we see that $x_1$ is not free in line 1).

Notice that our conclusion in the first (incorrect) proof is not logically valid. Think of $A_1^2$ as being the "less than" relation on real numbers. Then for every real number, there is a greater real number, but it is not true that there is a real number greater than every real number. On the other hand, the conclusion of our correct deduction depends on no hypotheses and was deduced without the aid of any proper axioms. It is thus a theorem of the predicate calculus and is universally valid. We will, of course, need to justify that our new rules really do give the same results as our old ones.

It would be possible to formulate our rules in such a manner as to forego the use of dummy constants in connection with the rule *eE* and use free variables instead. However, the rules then become much less visual and practical, because our various other restrictions, particularly those in rule $i\forall$, require that we remain aware of which variables have been introduced by an application of *eE* and which have not. But with dummy constant letters, which are visually different from free variables, checking our rules is much easier. The dummy constants serve as "markers" when we come to apply $i\forall$. If the formula to which we wish to apply $i\forall$ contains dummy constants, then we must check that the variable in whose name we wish to generalize does not occur free in any previous formula to which *eE* has been applied. If the formula in question contains no dummy constants, we have only to check that it depends on no hypotheses which do contain dummy constants or that the variable in question is not free in any formula $(Ey)\,B(y)$ to which *eE* has been previously applied. Of course, we always have to check that the variable in whose name we generalize is not free in any hypothesis on which the wff in question depends. The reader will find that checking these things becomes

Copyrighted material

rather natural after practice though descriptive statements of the procedure appear verbose.

Another point with respect to our restrictions for *eE* in $i\forall$ is that our system is really more flexible than is absolutely necessary. We can obtain an adequate set of rules if we replace $i\forall$ by the weaker rule

$$i\forall^*: \frac{A(x)}{(x)\,A(x)}$$

where the variable $x$ is not free is any hypothesis on which $A(x)$ depends, and where $A(x)$ is a wff and depends only on hypotheses which are wffs. The weaker rule $i\forall^*$ is the same as $i\forall$ except that we have suppressed one of our alternatives for *eE*.

The fact that our set of natural deduction rules with the weaker rule $i\forall^*$ really is adequate will be presently justified. Our reason for making the observation here is that it will help us to see that our natural deduction rules allow us to introduce previously proved theorems at any point in a deduction without introducing further dependence on hypotheses. Under our old rules and axioms, such citing of previously proved theorems as justification for a line of a formal deduction has depended on the fact that the juxtaposition of two formal deductions is a formal deduction (see Exercise 4, Chapter 1, p. 12). That is, according to our old rules, if $X$ is some sequence of wffs which constitutes a proof in any given system $F$, and if $Y$ is another such sequence, then $XY$ is also a formal proof in $F$. This justifies the introduction of a theorem at any point in a proof, since the formal proof of a theorem can be interposed at any line of a proof. Now under our natural deduction rules, a theorem depends on no hypotheses, and so if the proof of a theorem $A$ can be legitimately interposed at any point of a deduction, it introduces no further *dependence* on hypotheses (the hypotheses themselves may be introduced, but dependence on them will have been eliminated by the time $A$ is proved). However, for our natural deduction rules we have some complications not present in our old system of rules and axioms, and this fact requires that we examine carefully under what conditions it is legitimate to juxtapose two deductions. Let us take an example.

Suppose that we have a proof $X$ in which the dummy constant $b$ is introduced by an application of *eE*. Suppose now that we have a proof $Y$ of the wff $A$ and that $Y$ also involves introduction of the dummy constant $b$. The juxtaposition $XY$ is not a proof, for now the introduction of $b$ in the proof of $A$ violates one of our rules, namely *eE*; this occurrence of $b$ in the sequence $Y$ is no longer new to the proof $XY$, since it has been introduced in $X$, which precedes every formula of $Y$ in the sequence $XY$.

Of course, it is immediately clear that this is inconsequential. If $Y$ is a proof of the wff $A$, then there exists a proof $Y'$ of $A$ where the dummy constants introduced by applications of *eE* can be judiciously chosen to be new to any given proof $X$. Thus, the juxtaposition $XY'$ will not violate our rule *eE*, and the citing of the theorem of $A$ can be justified. We know that there always will be a proof of $A$ which avoids the difficulty of the new constants in any given case.

Clearly there are no difficulties with the rules $e\forall$ and *iE* that would prevent us from juxtaposing two proofs involving any finite number of applications of these rules. However, observe that with $i\forall$ we have difficulties similar to those of *eE*. Suppose, for example, that we have a proof $Y$ of the wff $A$ as a theorem and there is an application in $Y$ of $i\forall$ applied to a variable $z$ and a dwff $B$. Since $B$ is a dwff, this application of $i\forall$ is legitimate in the proof $Y$ only if $z$ is not free in any wff or dwff to which *eE* has been previously applied in $Y$. Suppose, however,

Copyrighted material

that $z$ is free in some wff or dwff $(Ew)C(w)$ to which $eE$ has been applied in the proof $X$. Then if we juxtapose $X$ and $Y$ to form the sequence $XY$, our restriction on the application of $i\forall$ to $z$ and $B$ in $Y$ is now violated though it was not before; for now the wff $(Ew)C(w)$ of $X$ contains $z$ free and occurs before $B$ in the new sequence $XY$, and there has been an application of $eE$ to $(Ew)C(w)$ prior to $B$ in the sequence $XY$. In the present case, it is not so obvious that there is another proof $Y'$ of $A$ which avoids this tedious difficulty. Of course, one feels that there ought to be some way of avoiding it, since the previous use of $eE$ in the sequence $X$ is obviously not related to the later one in $Y$.

Let us now observe that we have no difficulties of the above kind if the rule $i\forall$ is replaced by the weaker rule $i\forall^*$, for then the universal quantifier is applied only to wffs that depend on wff hypotheses. Clearly any such application of $i\forall^*$ in any proof sequence $Y$ is independent of any proof sequence $X$ with which $Y$ may be eventually juxtaposed.

We have already remarked that the introduction of hypotheses presents no problems for the citing of prior theorems in a deduction, since theorems depend on no hypotheses. We have thus justified the citing of prior theorems in the system of rules that are the same as our natural deduction rules, except that $i\forall$ is replaced by $i\forall^*$.

We now prove the adequacy of our natural deduction rules where we use only our weaker rule $i\forall^*$. This also justifies the citing of previous theorems in our natural deduction system.

THEOREM 2. *In any first-order system F,* $\vdash (x)A(x) \supset A(t)$, *where* $t$ *is any term free for* $x$ *in the wff* $A(x)$, *and* $A(t)$ *results from* $A(x)$ *by substituting* $t$ *for* $x$ *in all of the latter's free occurrences in* $A(x)$.

*Proof.* (1) 1. $(x)A(x)$ $H$
(1) 2. $A(t)$ 1, $e\forall$
3. $(x)A(x) \supset A(t)$ 1, 2, $eH$

THEOREM 3. *In any first-order system whatever,*

$$\vdash (x)(A \supset B(x)) \supset (A \supset (x)B(x))$$

*where* $x$ *is not free in* $A$, *and* $A$ *and* $B(x)$ *are wffs.*

*Proof.* (1) 1. $(x)(A \supset B(x))$ $H$
(1) 2. $A \supset B(x)$ 1, $e\forall$
(3) 3. $A$ $H$
(1, 3) 4. $B(x)$ 2, 3, MP
(1, 3) 5. $(x)B(x)$ 4, $i\forall$ ($x$ not free in 1 or 3)
(1) 6. $A \supset (x)B(x)$ 3, 5, $eH$
7. $(x)(A \supset B(x)) \supset (A \supset (x)B(x))$ 1, 6, $eH$

We have proved Theorem 2 and Theorem 3 for any system $F$ and so wffs of the indicated form are theorems of any system. These are, of course, logical axioms of our previous formulation of rules for the predicate calculus. In proving Theorems 2 and 3, we have used only our weak rule $i\forall^*$, since we have not even used universal generalization. Besides the two types

Copyrighted material

of logical axioms just proved, we had tautologies, which we also have in our natural deduction rules. Also, we have, in both cases, a rule of *modus ponens* (the rule MP) and generalization (the rule $i\forall^*$).[†] Our natural deduction rules thus yield all of our previous logical rules and axioms, and so every theorem provable according to our prior rules is provable according to our natural deduction rules.

Of course there is one further complication that must now be considered. In our natural deduction system, the use of the existential quantifier is defined explicitly, and we have no right to consider it as definable in terms of negation and the universal quantifier unless we can prove this fact from our rules. That is to say, we now consider that "$E$" figures in our alphabet for first-order systems, and our definition of wffs must be extended to include expressions obtained from wffs $A$ by formally applying $(Ex)$ to get $(Ex)A$ where $x$ is any variable. Occurrences of variables in the scope of $(Ex)$ are bound, etc. What we now need to prove is that $\vdash (Ex)\,A(x) \equiv (\sim (x)\,(\sim A(x)))$ for any wff $A(x)$ in any system $F$ whatever. Also, we need to prove a general theorem of the substitutivity of logical equivalence in order to show that we really can always replace the existential quantifier by its equivalent in terms of negation and the universal quantifier. Moreover, the reader should observe that all this will be proved where every application of universal generalization satisfies our weak rule $i\forall^*$, as is the case for Theorem 2 and Theorem 3 of this section. Once this program is complete, the adequacy of our natural deduction rules is clearly established, since every proof according to our old rules and axioms is shown to be directly translatable into a proof using our natural deduction rules; in fact a proof in which only the weaker rule $i\forall^*$ is used.

We first prove the following theorem:

THEOREM 4. *In any first-order system F,* $\vdash (\sim (x)\,(\sim A(x))) \supset (Ex)\,A(x)$ *where* $A(x)$ *is any wff.*

*Proof.*
(1) 1. $A(x)$ $H$
(1) 2. $(Ex)\,A(x)$ 1, $iE$
3. $A(x) \supset (Ex)\,A(x)$ 1, 2, $eH$
4. $(A(x) \supset (Ex)\,A(x)) \supset (\sim (Ex)\,A(x) \supset \sim A(x))$ Taut
5. $\sim (Ex)\,A(x) \supset \sim A(x)$ 3, 4, MP
(6) 6. $\sim (Ex)\,A(x)$ $H$
(6) 7. $\sim A(x)$ 5, 6, MP
(6) 8. $(x)\,(\sim A(x))$ 7, $i\forall$ ($x$ not free in [6])
9. $\sim (Ex)\,A(x) \supset (x)\,(\sim A(x))$ 6, 8, $eH$
10. $(\sim (Ex)\,A(x) \supset (x)\,(\sim A(x))) \supset (\sim (x)\,(\sim A(x))$
$\supset (\sim\sim (Ex)\,A(x)))$ Taut

[†] Notice that all our restrictions for $i\forall^*$ will always be satisfied whenever we apply $i\forall^*$ to a wff $X$ which is a theorem. Since a theorem depends on no hypotheses, the $eH$ restriction on $i\forall^*$ will be satisfied. Since a theorem is not a dwff, and since all hypotheses on which it depends (there are none) are wffs, the $eE$ restriction for $i\forall^*$ is also satisfied. In any proof of a theorem using our old rules and axioms, UG will be applied only to theorems and axioms, and it thus follows that $i\forall^*$ will be as strong as UG for the purpose of translating proofs of theorems from our old axioms and rules into the natural deduction rules using only $i\forall^*$.

Copyrighted material

11. $\sim (x)(\sim A(x)) \supset (\sim\sim (Ex)\, A(x))$ 9, 10, MP
12. $\sim\sim (Ex)\, A(x) \supset (Ex)\, A(x)$ Taut
13. $[11] \supset ([12] \supset \text{Concl})$ Taut
14. $[12] \supset$ Concl 11, 13, MP
15. Concl 12, 14, MP

Here we see the use of bracketed numbers again to avoid rewriting lines of the proof which have already occurred. "Concl" means "conclusion"; that is, the statement to be proved. As we progress in our techniques of formal deduction, we shall begin to omit certain steps and give as a justification for the lines that appear the collective justification for the omitted and presented steps.

In the previous proof, for example, we might have jumped from line 9 directly to line 11 by giving as a collective justification for line 11 (which would then be line 10): 9, Taut, MP. We call such formal proofs, in which some lines are omitted, *quasiformal.* If he wishes, the reader may take it as a standing exercise in this book to supply the missing lines to quasiformal proofs.

Of course, informal proofs given in most mathematical literature are not even quasiformal. They are informal arguments which tend to convince the reader that a formal proof does exist and which permit the knowledgeable reader to supply the missing steps. In our treatment of deduction in this book, there will be a decreasing component of formalism. In Chapter 3, we give fairly complete formal or quasiformal proofs. This is also the case in the beginning of Chapter 5. Then we gradually relax and revert to the more usual informal "discussion" type of proof familiar in mathematical literature. This approach should enable the reader to appreciate more fully the notion of a proof. He should be more adept at translating from formal to informal and back again.

We now complete our treatment of the existential quantifier by proving the following theorem:

THEOREM 5. *In any first-order system F,* $\vdash (\sim (x)(\sim A(x))) \equiv (Ex)\, A(x)$, *where* $A(x)$ *is any wff.*

*Proof.* (1) 1. $(Ex)\, A(x)$ $H$
(1) 2. $A(b)$ 1, $eE$, $b$ a new dummy constant
(3) 3. $(x)(\sim A(x))$ $H$
(3) 4. $\sim A(b)$ 3, $e\forall$, $b$ is free for $x$
5. $(x)(\sim A(x)) \supset \sim A(b)$ 3, 4, $eH$
6. $A(b) \supset \sim (x)(\sim A(x))$ 5, Taut, MP
(1) 7. $\sim (x)(\sim A(x))$ 2, 6, MP
8. $(Ex)\, A(x) \supset \sim (x)(\sim A(x))$ 1, 7, $eH$
9. $(\sim (x)(\sim A(x))) \equiv (Ex)\, A(x)$ 8, Th. 4, *Df.* $\equiv$, Taut, MP

Here we have given a quasiformal proof, omitting a few steps, of the converse of Theorem 4. In citing theorems in proofs we use the abbreviation "Th.", as well as the abbreviation "*Df.*" for citing definitions. Both Theorem 4 and Theorem 5 use only $i\forall^*$. In particular, this justifies our citing of Theorem 4 in the proof of Theorem 5.

Copyrighted material

Theorem 5 establishes the equivalence of $(Ex)\,A(x)$ with

$$(\sim (x)\,(\sim A(x))).$$

It thus shows that we can recover our definition of $(Ex)$ in terms of negation and universal quantification from our natural deduction rules. Of course, when we defined the existential quantifier in terms of negation and the universal quantifier, it meant that we could always replace $(Ex)$ by $\sim (x) \sim$. In order to establish the same thing here, we need the substitutivity of logical equivalence.

In Theorem 3 of Section 1.1, we proved the substitutivity of tautological equivalence. What we must now prove is that this principle of substitutivity holds for the predicate calculus. When we have done this, the full definability of $(Ex)$ as $\sim (x) \sim$ will have been established. We need some preliminary lemmas.

THEOREM 6. *In any first-order theory,* $\vdash (x)\,(A \equiv B) \supset ((x)A \equiv (x)B)$, *where* $A$ *and* $B$ *are any wffs.*

*Proof.*

| | |
|---|---|
| (1) | 1. $(x)\,(A \equiv B)$   $H$ |
| (1) | 2. $A \equiv B$   1, $e\forall$ |
| (3) | 3. $(x)A$   $H$ |
| (3) | 4. $A$   3, $e\forall$ |
| (1, 3) | 5. $B$   2, 4, Taut, MP |
| (1, 3) | 6. $(x)B$   5, $i\forall$ |
| (1) | 7. $(x)A \supset (x)B$   3, 6, $eH$ |
| (8) | 8. $(x)B$   $H$ |
| (8) | 9. $B$   8, $e\forall$ |
| (1, 8) | 10. $A$   2, 9, Taut, MP |
| (1, 8) | 11. $(x)A$   10, $i\forall$ |
| (1) | 12. $(x)B \supset (x)A$   8, 11, $eH$ |
| (1) | 13. $[7] \wedge [12]$   7, 12, Taut, MP |
| (1) | 14. $(x)\,A \equiv (x)B$   13, $Df \equiv$ |
| | 15. Concl 1, 14 $eH$ |

From now on, we shall not use the vernacular to mention specific conditions relating to the rule $i\forall$ as we did in line 8 of the proof of Theorem 4. It is up to the reader to see that each application of our rules is justified. Thus, in lines 6 and 12 of the proof of Theorem 6 of this section, the variable $x$ is not free in the hypotheses on which the wff involved in the application of $i\forall$ depends, but we do not state this explicitly. It is to be considered part of the notation "$i\forall$" that the application of the rule must satisfy all restrictive conditions, and we mentioned these in previous proofs only for emphasis. Again, we note that only $i\forall^*$ is used in proving Theorem 6.

THEOREM 7. *In any first-order system,*

$$\vdash (x)\,(A(x) \equiv B(x)) \supset ((Ex)\,A(x) \equiv (Ex)\,B(x))$$

*where* $A(x)$ *and* $B(x)$ *are any wffs.*

Copyrighted material

*Proof.*

| | |
|---|---|
| (1) | 1. $(x)(A(x) \equiv B(x))$ $H$ |
| (2) | 2. $(Ex)A(x)$ $H$ |
| (2) | 3. $A(b)$ 2, $eE$ |
| (1) | 4. $A(b) \equiv B(b)$ 1, $e\forall$ |
| (1, 2) | 5. $B(b)$ 3, 4, Taut, MP |
| (1, 2) | 6. $(Ex)B(x)$ 5, $iE$ |
| (1) | 7. $(Ex)A(x) \supset (Ex)B(x)$ 2, 6, $eH$ |
| (8) | 8. $(Ex)B(x)$ $H$ |
| (8) | 9. $B(c)$ 8, $eE$ |
| (1) | 10. $A(c) \equiv B(c)$ 1, $e\forall$ |
| (1, 8) | 11. $A(c)$ 9, 10, Taut, MP |
| (1, 8) | 12. $(Ex)A(x)$ 11, $iE$ |
| (1) | 13. $(Ex)B(x) \supset (Ex)A(x)$ 8, 12, $eH$ |
| (1) | 14. [7]$\wedge$[13] 7, 13, Taut, MP |
| (1) | 15. $(Ex)A(x) \equiv (Ex)B(x)$ 14, *Df.* $\equiv$ |
| | 16. Concl 1, 15, $eH$ |

Again, the dummy constants $b$ and $c$ in the proof of Theorem 7 are required to satisfy the conditions of being new constant letters and the other relevant restrictions. We suppose these requirements summed up in the notation "$eE$".

We are now in a position to prove the following theorem:

THEOREM 8. *In any first-order system, if* $\vdash A \equiv B$, *where* $A$ *and* $B$ *are any wffs, then* $\vdash X \equiv X'$ *where* $X'$ *is obtained from the wff* $X$ *by replacing* $B$ *for* $A$ *at zero, one, or more occurrences of* $A$ *in* $X$.

*Proof.* The proof is by induction on the number of sentence connectives and quantifiers in $X$. If $X$ has no sentence connectives or quantifiers, then $X$ is a prime formula. There are no well-formed (proper) parts to a prime formula, and so $X$ is $A$. Thus, either $X'$ is $B$ or $X'$ is $A$. In either case, the desired result follows.

We now suppose the theorem true for all wffs $X$ with fewer than $n$ quantifiers and sentence connectives. We must prove the assertion where $X$ has $n$ quantifiers and sentence connectives. If $X$ is a prime formula or if $X$ is $A$, then the argument is the same as in the foregoing case. Thus, we suppose that $A$ is a proper part of $X$ and that $X$ is not a prime formula. In this case, $X$ is of the form (1) $(x)C$, or (2) $(Ex)C$, or (3) $(\sim C)$, or (4) $(C \vee D)$ where $C$ and $D$ are wffs which necessarily have fewer than $n$ quantifiers or connectives. In order to complete the demonstration, we consider each of these cases.

If $X$ is of the form $(x)C$, then $A$ is a part (not necessarily proper) of $C$, since $A$ is a proper part of $X$. Let $C'$ be the result of replacing $A$ by $B$ in $C$ (in zero, one, or more occurrences). Then $X'$ is $(x)C'$. Now suppose that $\vdash A \equiv B$. Then, by induction hypothesis $\vdash C \equiv C'$, since $C$ has fewer than $n$ occurrences of quantifiers and sentence connectives. Now, since $C \equiv C'$ is a theorem, it depends on no hypotheses. Thus, we can apply $i\forall$ (in fact $i\forall^*$) to it and obtain $\vdash (x)(C \equiv C')$. Now, applying MP to this and Theorem 6, we obtain $\vdash (x)C \equiv (x)C'$; that is, $\vdash X \equiv X'$.

Copyrighted material

The proof for the case that $X$ is of the form $(Ex)C$ is exactly the same except that here we use Theorem 7.

If $X$ is of the form $(\sim C)$, then $X'$ will be the wff $(\sim C')$ and, by inductive hypothesis, $\vdash C \equiv C'$ if $\vdash A \equiv B$. But

$$\vdash (C \equiv C') \supset ((\sim C) \equiv (\sim C')),$$

since this last wff is a tautology. Applying MP we obtain the desired results.

In the last case, we have the tautology

$$\vdash (C \equiv C') \supset ((D \equiv D') \supset (C \vee D \equiv C' \vee D'))$$

which, with an argument analogous to our previous example (both $C$ and $D$ must have fewer than $n$ connectives and quantifiers), yields the desired result.

Thus, our assertion is established for the case of $n$ quantifiers and connectives and the theorem follows by mathematical induction.

*Exercise.* Let $X$ be a wff of some first-order theory and let $A$ be a wff which is a subformula of $X$. We say that $A$ *has simple occurrence* in $X$ if no free variable of $A$ is bound in $X$. Show that in any first-order theory $\vdash (A \equiv B) \supset (X \equiv X')$, where $X'$ is obtained from $X$ by replacing $B$ for $A$ at zero, one, or more occurrences of $A$ in $X$, $A$ has simple occurrence in $X$, and $B$ has simple occurrence in $X'$.

Again, we remark that all of the above theorems have been established by using only the weak rule $i\forall^*$.

Since Theorem 3 establishes $\vdash (Ex)\, A(x) \equiv (\sim (x)(A \sim (x)))$ for all wffs $A(x)$ in any system $F$, it follows from Theorem 8 that we can always replace "$(Ex)$" by "$\sim (x) \sim$" just as when $(Ex)$ was defined notation. The situation is analogous to our method of defining some of the sentential connectives in terms of others. We could just as easily have all five of our sentential connectives as basic signs of our alphabet, since our definitions of "$\supset$", "$\wedge$", and "$\equiv$" in terms of "$\vee$" and "$\sim$" are all tautological equivalences, and thus equivalences of the predicate calculus.

The adequacy of our rules is now fully established. We will no longer be concerned with whether or not a deduction satisfies our weaker rule $i\forall^*$. Although we know that the exclusive use of $i\forall^*$ will yield an adequate system of rules, we prefer the flexibility of our full rule $i\forall$.

*Exercise.* Use the exercise on page 32 and Theorems 6 and 7 above to prove that the universal closures of two similar wffs are logically equivalent, as well as the respective existential closures.

Notice that Theorem 8, in conjunction with the exercise on similarity on page 32, tells us the following: If $A(x)$ and $A(y)$ are similar, then $(Qx)\, A(x)$ can be replaced by $(Qy)\, A(y)$ in any formula $X$ to obtain an equivalent formula $X'$, where $(Qx)$ and $(Qy)$ represent either universal or existential quantification. We call this replacement rule the "change of name of a bound variable". It tells us that two formulas which differ only by the name of quantified variables are equivalent.

We now turn to the problem of the soundness of our natural deduction rules. What we need to show is that anything provable by our natural deduction system is also provable by our ori-

Copyrighted material

ginal rules and axioms. We will establish this by showing how to translate any natural deduction proof into a proof involving our original axioms and rules.

It might seem at first glance that Theorem 1 is already a justification of the soundness of natural deduction since it shows how to translate a $c$-deduction into a deduction without use of rule $c$, and the use of rule $c$ as a primitive rule is clearly the main innovation involved with our natural deduction rule system. However, what must be justified is the way rule $c$ is used in conjunction with the introduction and elimination of hypotheses involved in our natural deduction system. In short, we need a deduction theorem for deductions involving the use of rule $c$.

**Definition 8.** Let $B_1, \ldots, B_n$ and $A$ be wffs or dwffs of a given first-order theory $F$. We write $B_1, \ldots, B_n \vDash_c A$ to mean that $A$ is the last line of a $c$-deduction from the hypotheses $B_i$ in which no application of UG to a variable free in one of the $B_i$, and to a formula which $c$-depends on that $B_i$, has occurred.

It follows immediately from Definition 8 and Theorem 1 of this section that, in any system $F$, if $B_1, \ldots, B_n \vDash_c A$ where the $B_i$ and $A$ are all wffs, then $B_1, \ldots, B_n \vDash A$.

We are now in a position to state and prove our deduction theorem involving rule $c$:

THEOREM 9. *If, for some first-order theory F,* $B_1, \ldots, B_n, B \vDash_c X$ *and* $A_1(c_1), \ldots, A_k(c_k)$ *are the dwffs in order of first occurrence in the proof that result from an application of rule c, then* $\vec{B}_i \vDash_c (B \supset X)$ *and all of the (results of) applications of rule c which occur in this new c-deduction are among* $(B \supset A_1(c_1)), \ldots, (B \supset A_k(c_k))$.† *Moreover, the formula* $(B \supset X)$ *c-depends on any of the formulas* $(B \supset A_j(c_j))$ *or* $B_i$ *in the new c-deduction only if X c-depended on the formula* $A_j(c_j)$ *or* $B_i$ *in the original deduction.*

*Proof.* The proof is by induction on the length of the original deduction. Clearly the theorem holds for deductions of length 1. We thus suppose it holds for deductions of length less than $m$ and consider a deduction of length $m$.

Again, the result is immediate if $X$ is an hypothesis or an axiom.

If $X$ results by application of MP to formulas $C \supset X$ and $C$, then $\vec{B}_i \vDash_c B \supset (C \supset X)$ and $\vec{B}_i \vDash_c (B \supset C)$ by induction hypothesis. Thus, $\vec{B}_i \vDash_c B \supset X$ by tautology and *modus ponens.*

If $X$ is inferred from $C(y)$ by UG applied to $y$, then, again by induction hypothesis, $\vec{B}_i \vDash_c (B \supset C(y))$. If $y$ is not free in $B$, then $y$ is not free in any of the formulas $(B \supset A_j(c_j))$ on which $(B \supset C(y))$ $c$-depends since $X$ cannot $c$-depend on any $A_j(c_j)$ in which $y$ is free. Thus, $\vec{B}_i \vDash_c (y)(B \supset C(y))$ and thus $\vec{B}_i \vDash_c B \supset (y)C(y)$ by a logical axiom and *modus ponens.*

If $y$ is free in $B$, then $C(y)$ does not $c$-depend on $B$. Thus, $\vec{B}_i \vDash_c C(y)$ with the same applications of rule $c$ and where $y$ is not free in any $A_j(c_j)$ on which $C(y)$ depends. Let $z$ be an entirely new variable. Then we also have $\vec{B}_i \vDash_c C(z)$ with a deduction of equal or lesser length (and in particular, of length less than $m$). The $B_i$ and the $A_j(c_j)$ are unchanged since $y$ was not free in any $B_i$ or $A_j(c_j)$ on which $C(y)$ $c$-depended‡. Thus, we have $\vec{B}_i \vDash_c C(z)$ with a deduction of length less than $m$ and, trivially, $\vec{B}_i, B \vDash_c C(z)$ with a deduction of length less than $m$, whence $\vec{B}_i \vDash_c (B \supset C(z))$ by induction hypothesis. $z$ is not free in $B$. Moreover (and here is the only point of this contortion), the only possible applications of rule $c$ are of the indicated kind. Thus,

† $\vec{B}_i$ means $B_1, \ldots, B_n$.

‡ Let us recall that, by Theorem 1 of Section 1.2, we can always suppose that the first deduction only involves formulas on which $C(y)$ $c$-depended.

Copyrighted material

image not available

*image not available*

Of course, these are all examples and illustrations of the notion of a variable-binding term operator. Let us now give a formal definition.

**Definition 3.** By a *variable-binding term operator*, abbreviated "*vbto*", of a first-order theory $F$, we mean a symbol $v$ which is explicitly added to the alphabet of $F$ and which combines with a wff $A$ and a variable $x$ to yield a term $t$ (the particular formal manner of combination is unimportant, but must be specified clearly when defining the wffs of $F$). The variable $x$ is said to be *bound* in the term $t$, and it is considered bound in any wff $X$ that contains the term $t$. The rules concerning substitution, bound and free variables, freedom for, and so on apply to wffs containing $t$, and to $t$, respectively.

Though we have thought it better to avoid fixing the grammar of *vbto*s once and for all, we will generally write "$vxA(x)$" to represent the term formed by applying the unspecified *vbto* $v$ to the variable $x$ and wff $A(x)$.

We need now to extend our definition of an interpretation $\langle D, g\rangle$ of a formal system in order to incorporate *vbto*s. We need to define, for a given infinite sequence $s$ of elements of $D$, what the object $g_s(t)$ is, where $g_s$ is our function from terms of $F$ to objects of $D$ defined relative to the sequence $s$, and $t$ is a term defined by means of a *vbto* $v$. To do this we need to define the interpretation of a *vbto* under the mapping $g$.

**Definition 4.** Let a first-order theory $F$, a *vbto* $v$ of $F$, and an interpretation $\langle D, g\rangle$ of $F$ be given. The mapping $g$ assigns to $v$ a function $g_v$ from $\mathcal{P}(D)$ to $D$. $g_v$ thus assigns an object in $D$ to each subset of $D$.†

We now extend the definition of our functions $g_s$ in order to define $g_s(t)$ where $t$ is a term defined by a *vbto* $v$, and $s$ is any infinite sequence of elements of $D$ as before. This extension is accomplished by the following definition:

**Definition 5.** Let $A$ be some wff of a first-order theory $F$ and let $t$ be the term formed by a *vbto* $v$ from $A$, where $v$ binds the variable $x_i$. Let an infinite sequence $s$ of elements of $D$ be given. Then, $g_s(t) = g_v(Y)$ where $Y$ is the set of all elements $d \in D$ such that $s'$ satisfies $A$ where $s'_i = d$ and $s'_j = s_j$ for all $j \neq i$. We suppose this condition added to the recursive definition of $g_s$ previously given.

Using the extended definition of $g_s$, we now know what it means for a sequence $s$ to satisfy a formula involving *vbto*s. We thus know what it means to say that such formulas are true. We have therefore totally determined the semantics of *vbto*s.

Let us illustrate this semantics with an example. Let $v$ be the abstraction operator of set theory and consider the term $t = \{x_1 \mid x_1 = x_2\}$ in an appropriate language. For any sequence $s$ having $s_2 = d_2$, $g_s(t) = g_v(Y)$ where $Y$ is the set of all $d$ in $D$ such that the couple $\langle d, d_2\rangle$ is in $g(=)$. For a normal model, $Y$ will be the set $\{d_2\}$. Otherwise, it will be the set of all those elements of $D$ which bear the relation $g(=)$ to $d_2$. If, now, we consider the term $r = \{x_1 \mid x_1 = 2\}$, then $Y$ will be the same for every sequence, namely the set of all elements of $D$ which bear the relationship $g(=)$ to 2. For a normal model, this will be $\{2\}$. Finally, $g_v(Y)$ is, in each case, some element $d'$ in $D$.

† In order to accommodate this definition, the codomain of the mapping $g$ must now be extended to include the set $\mathcal{P}(D) \times D$.

Copyrighted material

image
not
available

This completes the list of proper axioms of **S**. We have an infinite set of axioms, since **S.**12 is an axiom scheme.

The axioms **S.**6, **S.**7, and **S.**12 are formal analogues of the Peano postulates 3, 4, and 5. The first two Peano postulates are already part of our language by the inclusion of the constant letter 0 and the singulary function letter ′. (It is common practice to omit initial universal quantifiers in stating axioms. Since the universal closure of any theorem is a theorem by $i\forall$, the closed form immediately follows. Such differences in presentation are immaterial.)

The first three axioms are the reflexive, symmetric, and transitive properties of equality. **S.**4 and **S.**5 express the substitutivity of equality with respect to our basic function letters. The general substitutivity of equality can be proved as a metatheorem by using induction (in the metalanguage) on the number of function letters in the terms considered. Thus, **S** is a system in which equality is definable.

*Exercise.* Prove the last assertion. That is, prove that

$$(x)(y)(x = y \supset (X \equiv Y))$$

is a theorem of **S** where $Y$ is a wff obtained from the wff $X$ by replacing $y$ for $x$ in zero, one, or more occurrences of $x$ in $X$, and where $y$ is free for $x$ in those occurrences of $x$ that it replaces. This, together with **S.**1, yields the result that equality is definable in **S**.

We could, of course, have chosen to state this metatheorem as an axiom scheme and thus make **S** a theory with equality. The axioms **S.**2 to **S.**5 could then have been omitted, since they are special cases of this metatheorem. Such differences in presentation are immaterial as is the question of whether or not to include initial universal quantifiers in stating axioms.

Axioms **S.**8 and **S.**9 are known as the *recursive definitions* of addition. Of course we are already given that addition is defined, since we have a binary function letter for it. Therefore, in the system **S**, these equations serve to determine certain necessary properties of addition. Similarly, **S.**10 and **S.**11 give necessary properties of multiplication. By means of these definitions, together with the axiom of induction **S.**12, the usual properties of addition and multiplication of natural numbers can be deduced.

This brings us to a discussion of **S.**12. Under the *standard model* of **S**, the one in which the domain is the set $N$ of natural numbers, 0 names zero, the successor function represents the addition of the unit 1, addition represents addition, multiplication represents multiplication, and equality stands for identity. Any wff $A(x)$ with exactly one free variable will express some set of natural numbers, its truth set $X$. **S.**12 thus says that if this set $X$ contains 0 and the successor of any natural number it contains, then everything (and thus, under the standard model, every natural number) is in the set. Consequently, **S.**12 would seem to be the formal analogue of the last Peano postulate. This is, however, not quite true. There are obviously only denumerably many wffs of **S** and hence only denumerably many wffs with one free variable. Thus, there are only denumerably many different truth sets that the wff $A(x)$ of **S.**12 can express. But the set $\mathcal{P}(N)$ of all subsets of $N$ is nondenumerable. Hence, there are nondenumerably many subsets of $N$ that are excluded by **S.**12. There is simply no way to "talk about" them in **S**. **S.**12 thus represents a weak form of Peano postulate (5). This seemingly innocent fact has some surprising consequences, which are examined in Chapter 6.

For an example deduction in **S**, let us prove the following theorem:

Copyrighted material

image
not
available