JOSEPH R. SHOENFIELD
Department of Mathematics
Duke University

MATHEMATICAL LOGIC

ADDISON-WESLEY PUBLISHING COMPANY

READING, MASSACHUSETTS · MENLO PARK, CALIFORNIA · LONDON · DON MILLS, ONTARIO

CHAPTER 1

THE NATURE OF MATHEMATICAL LOGIC

1.1 AXIOM SYSTEMS

Logic is the study of reasoning; and mathematical logic is the study of the type of reasoning done by mathematicians. To discover the proper approach to mathematical logic, we must therefore examine the methods of the mathematician.

The conspicuous feature of mathematics, as opposed to other sciences, is the use of proofs instead of observations. A physicist may prove physical laws from other physical laws; but he usually regards agreement with observation as the ultimate test of a physical law. A mathematician may, on occasions, use observation; for example, he may measure the angles of many triangles and conclude that the sum of the angles is always 180°. However, he will accept this as a law of mathematics only when it has been proved.

Nevertheless, it is clearly impossible to prove all mathematical laws. The first laws which one accepts cannot be proved, since there are no earlier laws from which they can be proved. Hence we have certain first laws, called *axioms*, which we accept without proof; the remaining laws, called *theorems*, are proved from the axioms.

For what reason do we accept the axioms? We might try to use observation here; but this is not very practical and is hardly in the spirit of mathematics. We therefore attempt to select as axioms certain laws which we feel are evident from the nature of the concepts involved.

We thus have a reduction of a large number of laws to a small number of axioms. A rather similar reduction takes place with mathematical concepts. We find that we can define certain concepts in terms of other concepts. But again, the first concepts which we use cannot be defined, since there are no earlier concepts in terms of which they can be defined. We therefore have certain concepts, called *basic concepts*, which are left undefined; the remaining concepts, called *derived concepts*, are defined in terms of these. We have a criterion for basic concepts similar to that for axioms: they should be so simple and clear that we can understand them without a precise definition.

In any statement, we can replace the derived concepts by the basic concepts in terms of which they are defined. In particular, we may do this for axioms. Hence we may suppose that all the concepts which appear in the axioms are basic concepts.

We may now describe what a mathematician does as follows. He presents us with certain basic concepts and certain axioms about these concepts. He then explains these concepts to us until we understand them sufficiently well to see that the axioms are true. He then proceeds to define derived concepts and to prove theorems about both basic and derived concepts. The entire edifice which he constructs, consisting of basic concepts, derived concepts, axioms, and theorems, is called an *axiom system*. It may be an axiom system for all of mathematics, or for a part of mathematics, such as plane geometry or the theory of real numbers.

We have so far supposed that we have definite concepts in mind. Even so, it may be possible to discover other concepts which make the axioms true. In this case, all the theorems proved will also be true for these new concepts. This has led mathematicians to frame axiom systems in which the axioms are true for a large number of concepts. A typical example is the set of axioms for a group. We call such axioms systems *modern* axioms systems, as opposed to the *classical* axiom systems discussed above. Of course, the difference is not really in the axiom system, but in the intentions of the framer of the system.

Guided by this discussion, we shall begin the study of mathematical logic by studying axiom systems. This will eventually lead us to a variety of problems, some of them only faintly related to axiom systems.

1.2 FORMAL SYSTEMS

An axiom (or theorem) may be viewed in two ways. It may be viewed as a sentence, i.e., as the object which appears on paper when we write down the axiom, or as the meaning of a sentence, i.e., the fact which is expressed by the axiom. At first sight, the latter appears much more important. The obvious purpose of the sentence is to convey the meaning of the sentence in a clear and precise manner. This is a useful purpose, but it does not seem to have much to do with the foundations of mathematics.

Nevertheless, there are two good reasons for studying axioms and theorems as sentences. The first is that if we choose the language for expressing the axioms suitably, then the structure of the sentence will reflect to some extent the meaning of the axiom. Thus we can study the concepts of the axiom system by studying the structure of the sentences expressing the axioms. This is particularly valuable for modern axiom systems, since for them our initial understanding of the basic concepts may be very weak.

The second reason is that the concepts of mathematics are usually very abstract and therefore difficult to understand. A sentence, on the other hand, is a concrete object; so by studying axioms as sentences, we approach the abstract through the concrete.

One point is apparent: there is no value in studying concrete (rather than abstract) objects unless we approach them in a concrete or constructive manner. For example, when we wish to prove that a concrete object with a certain property exists, we should actually construct such an object, not merely show that the nonexistence of such an object would lead to a contradiction.

Proofs which deal with concrete objects in a constructive manner are said to be *finitary*. Another description, suggested by Kreisel, is this: a proof is finitary if we can *visualize* the proof. Of course, neither description is very precise; but we can apply them in many cases to decide whether or not a particular proof is finitary.

Once the fundamental difference between concrete and abstract objects is appreciated, a variety of questions are suggested which can only be answered by a study of finitary proofs. For example Hilbert, who first instituted this study, felt that only finitary mathematics was immediately justified by our intuition. Abstract mathematics is introduced in order to obtain finitary results in an easier or more elegant manner. He therefore suggested as a program to show that all (or a considerable part) of the abstract mathematics commonly accepted can be viewed in this way. The question of how far such a program can be carried out is of obvious interest, even to those who do not find Hilbert's view of abstract mathematics congenial.

The study of axioms and theorems as sentences is called the *syntactical* study of axiom systems; the study of the meaning of these sentences is called the *semantical* study of axiom systems. For the above reasons, we shall often keep the syntactical and the semantical parts of our investigations separate. When it is possible and reasonably convenient, we shall carry out our syntactical investigations in a finitary manner. We shall always consider axioms and theorems to be sentences, and hence syntactical objects; when we wish to study them semantically, we speak of the meaning of the axiom or theorem.

To guide us in our syntactical study, we introduce the notion of a *formal system*. Roughly, a formal system is the syntactical part of an axiom system. We shall give a precise definition.

The first part of a formal system is its *language*. As previously stated, this should be chosen so that, as far as possible, the structure of the sentences reflects their meaning. For this reason among others, we generally use artificial languages for our formal systems.

To specify a language, we must first of all specify its *symbols*. In the case of English, the symbols would be the letters, the digits, and the punctuation marks. Most of our artificial languages will have infinitely many symbols.

Any finite sequence of symbols of a language is called an *expression* of that language. It is understood that a symbol may appear several times in an expression; each such appearance is called an *occurrence* of that symbol in the expression. The number of occurrences of symbols in an expression is called the *length* of that expression. (Thus the English expression *boot* has length 4.) We allow the empty sequence of symbols as an expression; it is the only expression of length 0.

It is possible for one expression to appear within another expression. Each such appearance is called an *occurrence* of the first expression in the second expression. Thus the English expression *on* has two occurrences in the English expression *confront*. However, we do not count it as an occurrence when the symbols of the first expression appear in the second expression in a different order or separated by other symbols. Thus *on* has no occurrences in *not* or in *corn*.

Most expressions in English are meaningless. Among the meaningful ones are the (declarative) sentences, which may be roughly described as those expressions which state some fact. We shall require that in each language, certain expressions of the language are designated as the *formulas* of the language; it is intended that these be the expressions which assert some fact.

We consider a language to be completely specified when its symbols and formulas are specified. This makes a language a purely syntactical object. Of course, most of our languages will have a meaning (or several meanings); but the meaning is not considered to be a part of the language. We shall designate the language of a formal system F by $L(F)$.

The next part of a formal system consists of its *axioms*. Our only requirement on these is that each axiom shall be a formula of the language of the formal system.

We need a third part of a formal system which will enable us to conclude theorems from the axioms. This is provided by the *rules of inference*, which we often call simply *rules*. Each rule of inference states that under certain conditions, one formula, called the *conclusion* of the rule, can be *inferred* from certain other formulas, called the *hypotheses* of the rule.

How should we define the *theorems* of a formal system F? Obviously they should satisfy the two laws:

i) the axioms of F are theorems of F;

ii) if all of the hypotheses of a rule of F are theorems of F, then the conclusion of the rule is a theorem of F.

Moreover, we want a formula to be a theorem of F only if it follows from these laws that it is a theorem. We can therefore define a theorem of F to be a formula of F which can be seen to be a theorem on the basis of laws (i) and (ii).

We can give a somewhat more explicit description of the theorems of F. Let S_0 be the set of axioms; these are the formulas which can be seen to be theorems on the basis of (i). Let S_1 be the set of formulas which are conclusions of rules whose hypotheses are all in S_0; these are some of the formulas which can be seen to be theorems on the basis of (ii). Let S_2 be the set of formulas which are conclusions of rules whose hypotheses are all in S_0 and S_1; these are also theorems on the basis of (ii). In this way, we can construct sets $S_3, S_4, \ldots$. Let S_ω be the set of formulas which are conclusions of rules whose hypotheses are all in at least one of $S_0, S_1, \ldots$; these are again theorems by (ii). We continue in this way until no new theorems can be obtained by (ii); and we then have all of the theorems.

A definition of the type just given is called a *generalized inductive definition*. A generalized inductive definition of a collection C of objects consists of a set of laws, each of which says that, under suitable hypotheses, an object x is in C. Some of the hypotheses may say that certain objects (related to x in a certain way) are in C. When we give such a definition, we always understand that an object shall be in C only if it follows from the laws that it is in C. We can then give a more explicit description of C along the above lines.

As another example, suppose that we have defined 0 and *successor*, and wish to define *natural number*. (The natural numbers are the nonnegative integers: 0, 1, 2, ... The successor of a natural number is the next larger natural number.) We can give the following generalized inductive definition:

i) 0 is a natural number;

ii) if y is a natural number, then the successor of y is a natural number.

In order to prove that every theorem of F has a property P, it suffices to prove that the formulas having property P satisfy the laws in the definition of *theorem*. In other words, it suffices to prove:

i′) every axiom of F has property P;

ii′) if all of the hypotheses of a rule of F have property P, then the conclusion of the rule has property P.

For (i′) and (ii′) imply that each member of the sets $S_0, S_1, \ldots, S_\omega, \ldots$ constructed above has property P; so every theorem of F has property P. A proof by this method is called a proof *by induction on theorems;* the assumption in (ii′) that the hypotheses of the rule have property P is called the *induction hypothesis.*

More generally, suppose that a collection C is defined by a generalized inductive definition. Then in order to prove that every object in C has property P, it suffices to prove that the objects having property P satisfy the laws of the definition. Such a proof is called a proof *by induction on objects in C.* The hypotheses in the laws that certain objects belong to C become, in such a proof, hypotheses that certain objects have property P; these hypotheses are called *induction hypotheses.* The reader will easily see that if C is the collection of natural numbers with the generalized inductive definition given above, then *proof by induction* and *induction hypothesis* have their usual meaning.

A rule in a formal system F is *finite* if it has only finitely many hypotheses. Almost all the rules which we will consider will be finite.

Let F be a formal system in which all the rules are finite. By a *proof* in F, we mean a finite sequence of formulas, each of which either is an axiom or is the conclusion of a rule whose hypotheses precede that formula in the proof. If A is the last formula in a proof P, we say that P is a proof *of A.*

We will show that a formula A of F is a theorem iff* there is a proof of A. First of all, it follows from the rules (i) and (ii) that every formula in a proof is a theorem; so if A has a proof, then it is a theorem. We prove the converse by induction on theorems. If A is an axiom, then A by itself is a proof of A; so A has a proof. Now suppose that A can be inferred from $B_1, \ldots, B_n$ by some rule of F. By the induction hypothesis, each of the B_i has a proof. If we put these proofs one after the other, and add A to the end of this sequence, we obtain a proof of A.

* We use *iff* as an abbreviation for *if and only if.*

We shall write $\vdash_F \ldots$ as an abbreviation for ... *is a theorem of F*. When no confusion results, we omit the subscript F.

The basic concepts of an axiom system will correspond to certain symbols or expressions in the associated formal system. The derived concepts, since they are defined in terms of the basic concepts, will generally correspond to more complicated expressions. If an important derived concept corresponds to a rather complicated expression, it may be desirable to introduce a new symbol as an abbreviation for that expression. We may also wish to introduce abbreviations to make certain expressions shorter or more readable.

For these reasons, we allow ourselves to introduce in any language new symbols, called *defined symbols*. Each such symbol is to be combined in certain ways with symbols of the language and previously introduced defined symbols to form expressions called *defined formulas*. Each defined formula is to be an abbreviation of some formula of the language. (In this terminology, an abbreviation does not have to be shorter than the expression which it abbreviates.) With each defined symbol, we must give a *definition* of that symbol; this is a rule which tells how to form defined formulas with the new symbol and how to find, for each such defined formula, the formula of the given language which it abbreviates.

We emphasize that defined symbols are *not* symbols of the language, and that defined formulas are *not* formulas of the language. Moreover, when we say anything about a defined formula, we are really talking about the formula of the language which it abbreviates (provided that it makes any difference). Thus the length of a defined formula is not the number of occurrences of symbols in the defined formula, but the number of occurrences of symbols in the formula which the defined formula abbreviates.

1.3 SYNTACTICAL VARIABLES

In our study of formal systems, we shall be studying expressions, just as an analyst studies real numbers. In both cases, the investigation is carried out in English augmented by certain special symbols specially suited to the investigation. We shall examine some of the special symbols used in analysis texts, and introduce analogous special symbols to be used in the investigation of formal systems.

First of all, an analysis text uses names for certain real numbers, for example, 3, $-\frac{1}{2}$, π. Similarly, we shall need names for expressions. We are in the fortunate position of being able to provide a name for each expression with one convention: each expression shall be used as a name for itself. This convention is not available to writers of analysis texts; for a name must be an expression, and a real number is not an expression.

There is, however, a danger to this convention. The expression may (in the language being discussed) be a name of some object; now it is also a name for itself. Thus *Boston* is the name of a city; according to our convention, it is also the name of an English word. We are saved from this danger because we only discuss artificial languages and because we discuss them in English. Thus when the expression occurs in a context written in the artificial language, it is a name of some object; when it occurs in a context written in English, it is a name of that expression.*

Variables are another important type of symbol used in analysis texts. Unlike a name, which has only one meaning, a variable has many meanings. In an analysis text, a variable may mean any real number; or, as we shall say, a variable *varies through* the real numbers. However, a variable keeps the same meaning throughout any one context. A formula containing variables also has many meanings, one for each assignment of a real number as a meaning to each variable occurring in the formula. For example, $x = x$ has $2 = 2$ and $\pi = \pi$ among its meanings; $x = y$ has these meanings, and also $2 = 5$. When a writer of an analysis text asserts a formula containing variables, he is claiming that all of its meanings are true.

We use *syntactical variables* in a similar manner, except that they vary through the expressions of the language being discussed instead of through the real numbers. Thus a syntactical variable may mean any expression of the language; but its meaning remains fixed throughout any one context. A formula containing syntactical variables has many meanings, one for each assignment of an expression as a meaning to each syntactical variable occurring in the formula. If we assert such a formula, we are claiming that all of its meanings are true.

To give an example of the use of syntactical variables and of expressions as names for themselves, suppose that x is a symbol of the formal system F. Suppose that it turns out that whenever we add the symbol x to the right of a formula of F, we obtain a new formula of F. If we have agreed to use $\mathbf{u}$ as a syntactical variable, we can express this fact as follows: if $\mathbf{u}$ is a formula, then the expression obtained by adding x to the right of $\mathbf{u}$ is a formula.

In an analysis text, some variables are restricted to vary through only certain real numbers. For example, it is common to restrict i and j to vary through integers only. We shall often use syntactical variables which vary through only certain expressions of the language being discussed. If we use $\mathbf{A}$ as a syntactical variable which varies through formulas, then the statement at the end of the previous paragraph can be abbreviated to: the expression obtained by adding x to the right of $\mathbf{A}$ is a formula.

In an analysis text, xy stands for the result of multiplying x by y. If $\mathbf{u}$ and $\mathbf{v}$ are syntactical variables, we shall use $\mathbf{uv}$ to stand for the expression obtained by juxtaposing $\mathbf{u}$ and $\mathbf{v}$, that is, by writing down $\mathbf{u}$ and then writing down $\mathbf{v}$ immediately after it. The same convention is used with other syntactical variables. It is also used to combine syntactical variables with names of expressions. As an example, we may shorten the statement at the end of the previous paragraph to: $\mathbf{A}x$ is a formula.

---

* To avoid any possibility of confusion, some books replace our convention by another convention: as a name for an expression, that expression enclosed in quotation marks is used.

We shall use boldface letters as syntactical variables. In particular, **u** and **v** will be syntactical variables which vary through all expressions, and **A**, **B**, **C**, and **D** will be syntactical variables which vary through formulas. Other syntactical variables will be introduced later. When we introduce a boldface letter as a syntactical variable, we shall understand that we may form new syntactical variables by adding primes or subscripts, and that these new syntactical variables vary through the same expressions as the old ones. Thus $\mathbf{A}'$ and $\mathbf{A}_1$ are syntactical variables which vary through formulas.

We add two words of caution. First, if two different syntactical variables occur in the same context, they do not necessarily represent different expressions (just as, in an analysis text, x and y do not necessarily represent different real numbers). Second, syntactical variables are *not* symbols of the language being discussed; they are symbols added to English to aid in the discussion of the language.