WHITEHEAD AND RUSSELL'S THEORY OF DEDUCTION AS A MATHEMATICAL SCIENCE*

BY B. A. BERNSTEIN

1. *Introduction.* Whitehead and Russell develop in their *Principia Mathematica* a theory of deduction for "elementary" propositions.† On this theory almost the entire *Principia* rests. From the primitives (the undefined ideas and the postulates) of this theory and from a few other primitives the authors proceed to derive all logic and all mathematics. It is then of great importance to the mathematician and to the logician to understand clearly the mathematical make-up of the theory of deduction. But such understanding is not easy. The novel views of the *Principia*, its strange symbolism, its elaborate structure, its long informal discussions, all combine to hide the real mathematical nature of the theory from the average man interested in mathematics and in logic. The purpose of this paper is to examine the theory critically and to exhibit clearly its true nature as a mathematical science.

My discussion will be based on the formal, the "official" development of the theory as found in the first edition of the *Principia* and as modified by the second edition. The informal discussions will be ignored unless they furnish needed information not obtainable from the formal account.‡

I begin with a brief account of the theory.

2. *Brief Account of the Theory.* The theory of deduction is "the calculus of propositions" (p. 88); it is "the theory of how one proposition can be inferred from another" (p. 90). As a mathematical science, as "the most elementary part of mathematics" (p. 115), the theory consists of a certain set of primitive

* Presented to the Society, June 20, 1929, as part of a paper entitled *On Whitehead and Russell's theory of deduction.*

† Whitehead and Russell, *Principia Mathematica*, vol. 1, second edition, pp. 87–126. In later references to pages, this volume of the *Principia* will be understood.

‡ There are some discrepancies between the informal and the formal accounts of the theory.

propositions and of theorems derived from the primitive propositions. The authors prove neither the consistency nor the independence of their primitive propositions, because "the recognized methods of proving independence" (and impliedly also consistency) "are not applicable, without reserve, to fundamentals" (p. 91, footnote).

There are two editions of the *Principia*. In the first edition the primitive ideas are six in number, as follows:

(1) *Elementary proposition*, denoted by $p, q, r, \cdots$;

(2) *Elementary propositional function of* x, denoted by "ϕx";

(3) *Assertion* of an elementary proposition p, denoted by "$\vdash \cdot p$";

(4) [*Assertion of an elementary propositional function* ϕx, denoted by "$\vdash \cdot \phi x$"];

(5) *Negation* of an elementary proposition p, denoted by "$\sim p$";

(6) *Disjunction* of two elementary propositions p and q, denoted by "$p \vee q$".

In the second edition the authors tell us (p. xiii) to drop notion (4) from the list of primitives. To indicate this fact, I have placed this notion in brackets.

The authors explain (pp. 88–93) their primitive ideas. "Elementary propositions" are "such as contain no reference, explicit or implicit, to any totality" (p. 88). An "elementary propositional function ϕx" is an expression containing a variable x such that when x is determined, ϕx becomes an elementary proposition. The symbol "$\vdash$" means "it is true that," and $\vdash \cdot p$ is to be distinguished from plain p. The symbol "$\sim p$" our authors read: " 'not-p' or 'p is false'." The disjunction "$p \vee q$" they read: " 'p or q', that is, 'either p is true or q is true'."

Before stating the primitive propositions the authors introduce a definition of the "implicative function" $p \supset q$. The definition is:

$$*1\cdot 01. \qquad (p \supset q) = (\sim p \vee q)\ Df.$$

The authors read this: " 'p implies q' is to be defined to mean 'either p is false or q is true'."

In the first edition the primitive propositions are ten in number. They are (with the original numbering retained and with the Peano dot-parentheses replaced by ordinary parentheses) propositions *1·1—*1·72 below. In the second edition we are told to omit *1·11 from the list of primitive propositions. I have placed this proposition in brackets. Following each of the propositions *1·2—*1·6 is the authors' interpretation of the symbols. The primitive propositions follow.

*1·1. Anything implied by a true elementary proposition is true.

Our authors say "we cannot express this principle symbolically."

*1·11. [When ϕx can be asserted, where x is a real variable, and $\phi x \supset \psi x$ can be asserted, where x is a real variable, then ψx can be asserted, where x is a real variable.]

*1·2. $\vdash \cdot (p \vee p) \supset p.$

"If either p is true or p is true, then p is true."

*1·3. $\vdash \cdot q \supset (p \vee q).$

"If q is true, then 'p or q' is true."

*1·4. $\vdash \cdot (p \vee q) \supset (q \vee p).$

" 'p or q' implies 'q or p'."

*1·5. $\vdash \cdot \{p \vee (q \vee r)\} \supset \{q \vee (p \vee r)\}.$

"If either p is true, or 'q or r' is true, then either q is true or 'p or r' is true."

*1·6. $\vdash \cdot (q \supset r) \supset \{(p \vee q) \supset (p \vee r)\}.$

"If q implies r, then 'p or q' implies 'p or r'."

*1·7. If p is an elementary proposition, $\sim p$ is an elementary proposition.

"1.71. If p and q are elementary propositions, $p \vee q$ is an elementary proposition."

*1·72. If ϕp and ψp are elementary propositional functions which take elementary propositions as arguments, $\phi p \vee \psi p$ is an elementary propositional function.

In the first edition much is made of the distinction between "real" and "apparent" variables, but in the second edition we are instructed (p. xiii) to abandon this distinction and to read a proposition of the form " $\vdash \cdot \phi p$" as if it were written " $\vdash \cdot (p) \cdot \phi p$," that is, "$\phi p$ is true for all p's."

The authors define the product $p \cdot q$ of two propositions p and q by:

$$*3\cdot01. \qquad (p\cdot q) = \{\sim(\sim p \vee \sim q)\}\ Df.$$

" 'p and q are both true' is defined to be 'it is false that either p is false or q is false'."

The *equivalence* of p and q, denoted by "$p\equiv q$," is defined by:

$$*4\cdot01. \qquad (p \equiv q) = (p \supset q)(q \supset p)\ Df.$$

"p and q are said to be *equivalent* when p implies q and q implies p" (p. 7).

The theorems derived from the primitive propositions are all, with the single exception of $*3\cdot03$,† of the form $\vdash\cdot\phi(p, q, r, \cdots)$, where $\phi(p, q, r, \cdots)$ is an elementary proposition built up from $p, q, r, \cdots$ by means of the operations "$\sim$" and "$\vee$," that is, the theorems are all, with the single exception of $*3\cdot03$, of the type of propositions $*1\cdot2-*1\cdot6$. In proving a theorem of the form $\vdash\cdot\phi(p, q, r, \cdots)$ the authors restrict themselves to two general methods. One is to find a *known* proposition $\vdash\cdot f(p, q, r \cdots)$ of which $\vdash\cdot\phi(p, q, r, \cdots)$ is a particular case got from $\vdash\cdot f(p, q, r\cdots)$ by substituting particular forms of elementary propositions for the general propositions $p, q, r, \cdots$. The other method is to find a function $f(p, q, r, \cdots)$ such that we have both $\vdash\cdot f(p, q, r, \cdots)$ and $\vdash\cdot f(p, q, r, \cdots) \supset \phi(p, q, r, \cdots)$, and then apply $*1\cdot11$. The demonstrations of the early theorems are given in full, except that reference to $*1\cdot7$, $*1\cdot71$, $*1\cdot72$ is generally tacit. The demonstrations are given in symbols, the authors employing for this purpose the symbols of the theory and other symbols.

This concludes my report of the theory of deduction.‡ In this report I tried to give a mere account of the theory as presented by the authors, an account without comment. What now shall we say of the mathematical nature of the theory? To answer this question I shall first give a brief characterization of a mathematical science, a characterization which an analysis of any orthodox mathematical science seems to me to justify.

---

† Proposition $*3\cdot03$ states that given $\vdash\cdot\phi p$ and $\vdash\cdot\psi p$ then we have $\vdash\cdot\phi p\cdot\psi p$.

‡ In the second edition, the authors give a great deal of space to a restatement of their theory into a language in which Sheffer's operation "$|$" is made basic. I have given no account of this restatement in my report, because it is irrelevant to the discussion of the theory as a mathematical science.

3. *Nature of a Mathematical Science.* A mathematical science, a science in the sense of a pure deductive theory, is a body of propositions consisting of *postulates* and *theorems*. The postulates are the propositions which cannot be derived from one another; the theorems are the propositions which are derivable from the postulates. The propositions of a mathematical science concern a certain totality of things and certain connections among the things; they give information about a certain *class* of *elements* and about certain *operations* or *relations* among the elements. The classes, operations and relations constitute the *ideas* of the science. Some of the ideas are *primitive*, that is, not definable in terms of one another; the rest are definable in terms of the primitive ones. Since the propositions of the science give information *about* its ideas, every proposition must contain, beside the ideas *belonging to* the science, also ideas that are *outside* the science. The latter are the ideas of general language by means of which the information is given. Likewise, since the theorems are derived from the postulates, the science must use, beside the propositions belonging to it, also propositions which are outside it. These are the principles of logic which give the theorems as conclusions from the postulates as premises.†

A mathematical science, in the sense of a pure deductive theory, is *abstract*. If, for example, a *class* K, a *binary operation* $\oplus$, and a *dyadic relation* $\circledcirc$ constitute the primitive ideas of a mathematical science, then K, $\oplus$, $\circledcirc$ have no properties other than that K is some class of elements, $\oplus$ some binary operation, and $\circledcirc$ some dyadic relation. K is simply some set of things, denoted, say, by $a, b, c, \cdots$; $\oplus$ is merely some rule which states for every pair of K-elements a, b what K-element $a \oplus b$, if any, corresponds to a, b; $\circledcirc$ is merely a rule which states for every pair of K-elements a, b whether or not the proposition $a \circledcirc b$ is true. All other properties of K, $\oplus$, $\circledcirc$ are given by the postulates. The science is the *system* $(K, \oplus, \circledcirc)$ satisfying the postulates. It is the blank form, the logical skeleton, for a number of concrete sciences each of which is obtainable from $(K, \oplus, \circledcirc)$

† On the necessity for distinguishing between the ideas and propositions belonging to a science and the ideas and propositions that are outside it, compare my *Sets of postulates for the logic of propositions*, Transactions of this Society, vol. 28, p. 472, last footnote. See also my review of the *Principia*, this Bulletin, vol. 32, pp. 711–713.

by substituting for K, $\oplus$, $\otimes$ respectively a concrete class, a concrete operation, and a concrete relation so as to make the postulates true propositions. It is this abstract nature of a mathematical science that makes possible the ordinary methods of proving the consistency and the independence of its postulates: the postulates of the above system (K, $\oplus$, $\otimes$) are consistent if K, $\oplus$, $\otimes$ can be so interpreted as to make all the postulates true; a postulate P is independent of the other postulates if K, $\oplus$, $\otimes$ can be so interpreted as to contradict P and not contradict any of the rest.

How does the theory of deduction of the *Principia* meet the requirements of a mathematical science just outlined? A brief consideration of the primitives will show that the theory falls quite short of these requirements both with regard to the primitive ideas and with regard to the primitive positions.

4. *Shortcomings of the Theory as a Mathematical Science.* As to the primitive ideas of the theory, we observe that the ideas denoted by p, $\sim p$, $p \vee q$ fully conform to the demands made on the primitive ideas of a mathematical science: p is a member of the class of elementary propositions, the totality with which the theory is concerned; $\sim p$ and $p \vee q$ denote results of operations in the class of elementary propositions. But the ideas represented by "ϕx" and "$\vdash \cdot p$" do not satisfy the above demands. ϕx can be *defined* in terms of the primitive ideas p, $\sim p$, $p \vee q$. For ϕx is used in the theory simply as an expression involving an elementary proposition x built up from x by means of "$\sim$" and "$\vee$", and hence is an *elementary proposition*, by *1·7 and *1·71. The idea "$\vdash \cdot p$" does not conform to the above demands because it does not stand for any *class* or *operation* or *relation*. Moreover, nothing is said *about* it, no condition is imposed on it, by the primitive propositions.

As to the primitive propositions of the theory, we observe that *1·7 and *1·71 are the only propositions that meet our requirements. *1·7 and *1·71 are the only primitive propositions which contain both ideas of the theory, about which information is given, and also ideas outside the theory, which give the information. Proposition *1·1 does not have the necessary requirements because it *has no symbols*, hence, impliedly, has no ideas belonging to the theory, hence says nothing about the ideas of the theory. Propositions *1·2—*1·6 have not the requirements because they *contain only symbols denoting ideas of the theory*, and hence do not say anything about these ideas. Finally, proposition *1·72 does not satisfy the requirement of a primitive proposition because *it can be derived as a theorem* by means of *1·7 and *1·71, since, as we have seen, ϕx and ψx are elementary propositions.

The presence of *1·1—*1·6 among the primitive propositions removes the theory from the category of abstract sciences, and would account for the authors' view that the recognized methods of proving independence are not applicable to their theory.

To the above defects in the primitives, as found in the first edition and retained in the second, must be added one introduced in the second edition. In this edition we are told to read every proposition of the form $\vdash \cdot \phi(p, q, \cdots)$ as if it were written $\vdash \cdot (p, q, \cdots) \cdot \phi(p, q, \cdots)$. But we are nowhere told whether the symbol $(p, q, \cdots) \cdot \phi(p, q, \cdots)$ denotes a primitive idea or not. In the first edition this symbol stands for a primitive idea carefully relegated to the theory of "apparent" variables.

5. *The Theory Transformed into a Mathematical Science.* But serious as are the shortcomings of the theory as a mathematical science, its primitive propositions, when the symbols are given appropriate interpretations, nevertheless give facts in logic, and the theorems are, indeed, as can be verified, logical consequences of the primitive propositions. It seems, therefore, that it should be possible, by merely changing the logical structure of the theory, to transform it into a mathematical, an abstract science. This indeed, is the case. And this transformation I wish now to carry out. The transformation will bring out very clearly the true mathematics underlying the theory.

We have seen that the ideas denoted by p, $\sim p$, $p \vee q$ and the propositions *1·7 and *1·71 are quite suitable as primitives for a mathematical theory, but that the notion ϕx and the proposition *1·72 are to be rejected as primitives on the ground that ϕx is definable by means of p, $\sim$, $\vee$, and that *1·72 is derivable from *1·7 and *1·71. In order to change the theory into a mathematical science, therefore, we must so dispose of the remaining primitives, the idea $\vdash \cdot p$ and the propositions *1·1—*1·6, as to free them from the objections raised against them, without changing the facts of logic embodied in the present symbolism. The necessary changes are simple.

The symbol $\vdash \cdot p$ stands for the proposition "p is true." This proposition implies that there exists a certain proposition, "1" say, characterized as "true," to which p is equivalent. Let "$p=q$" denote the proposition "p and q may be interchanged." Then, if p and q are propositions of our theory, "$p=q$" is the proposition "p and q are equivalent," that is, "p and q are both true or both false." And so the proposition "p is true" may be expressed by the symbolism "$p=1$," wherein, it is to be noted, p and 1 are ideas of our theory, but "$=$" is not. The new symbolization of "p is true" thus removes $\vdash \cdot p$ from the list of primitives. The new symbolization of "p is true" at the same time properly disposes of the propositions $*1\cdot 2 - *1\cdot 6$. For each of these propositions is changed to the form $\phi(p, q, \cdots)=1$, in which $\phi(p, q, \cdots)$ and 1 are ideas of the theory and "$=$" is not, and so is free from the criticism expressed in the preceding section.

As to proposition $*1\cdot 1$, though the authors say that they cannot express it symbolically, they *use* the proposition as if it were written "From '$\vdash \cdot p$' and '$\vdash \cdot p \supset q$' follows '$\vdash \cdot q$'," that is, "From $p=1$ and $\sim p \vee q=1$ follows $q=1$." And so $*1\cdot 1$, too, can be stated in a form free from the above criticism.

I have now indicated how to translate the theory into the language of an abstract mathematical science. Let me actually perform the translation. In this translation, I shall, for the sake of easy comparison of the theory with Boolean logic, replace $p \supset q$ by $\sim p \vee q$, in accordance with $*1\cdot 01$, and shall write p' for $\sim p$ and $p+q$ for $p \vee q$. For the sake of easy reference, the original numbering of the propositions will be retained, except that the sign "*" before the numbers will be omitted. The theory restated as a mathematical science follows.

Consider an undefined *class* K of elements $p, q, r, \cdots$. Let p' be the result of an undefined *unary operation* on a K-element p, and $p+q$ the result of an undefined *binary operation* on the K-elements p, q. The theory of deduction is the *system* $(K, ', +)$ satisfying the postulates $1\cdot 1 - 1\cdot 71$ below. In postulates $1\cdot 1 - 1\cdot 6$ there is implied the supposition that the indicated elements in each proposition are K-elements. In each of the propositions $1\cdot 2 - 1\cdot 6$ there is also implied the supposition that the element 1 of postulate $1\cdot 1$ exists and is unique. The postulates follow.

$1\cdot 1$. There exists a K-element 1 such that from $p=1$ and $p'+q=1$ follows $q=1$.

$1\cdot 2$. $(p+p)'+p=1$.

$1\cdot 3$. $q'+(p+q)=1$.

$1\cdot 4$. $(p+q)'+(q+p)=1$.

$1\cdot 5$. $[p+(q+r)]'+[q+(p+r)]=1$.

$1\cdot 6$. $(q'+r)'+[(p+q)'+(p+r)]=1$.

$1\cdot 7$. If p is a K-element, p' is a K-element.

$1\cdot 71$. If p and q are K-elements, $p+q$ is a K-element.

To complete the statement of the theory in the new language, I add the authors' definitions of $p \supset q$, $p\cdot q$ (or pq), and $p\equiv q$:

Def. $1\cdot 01$. $(p\supset q)=p'+q$.

Def. $3\cdot 01$. $pq=(p'+q')'$.

Def. $4\cdot 01$. $(p\equiv q)=(p\supset q)\,(q\supset p)=(p'+q)\,(q'+p)$.

We have now the "mathematicized" form of the theory of deduction. In this form the propositions give the same facts of logic as the propositions of the old form, but there is no confusion in it between the ideas belonging to the theory and the ideas that are outside it, and the ideas and propositions in it are all of the type of the ideas and propositions found in any orthodox mathematical theory.

6. *Conclusion.* The foregoing discussions make clear the mathematical nature of Whitehead and Russell's theory of deduction. The theory as a mathematical science is the system $(K, ', +)$ satisfying the postulates $1\cdot 1-1\cdot 71$ of the preceding section. This system is the theory of the *Principia* cleared of obvious redundancies in the primitives and written in the notation of Boolean logic. The two forms of the theory give the same facts of logic; but while in the Boolean form the ideas that are outside the theory are separated from those that belong to the theory, in the *Principia* form this separation is not made. It is the failure to distinguish between these two sets of ideas that makes the authors say that "the recognized methods of proving independence is not applicable, without reserve, to fundamentals." The theory, as seen from the Boolean form, is not more fundamental than any other mathematical theory and is subject to postulational investigations applicable to all other mathematical sciences.

THE UNIVERSITY OF CALIFORNIA