"Axiomatic set theory as a basis for the construction of mathematics".

A ~~dissertation~~ thesis submitted in support

of candidature for the Master of Science degree

of the

University of London

(Faculty of Science)

by

BRIAN ROTMAN

March, 1962.

ABSTRACT

It is widely known that one of the major tasks of 'Foundations' is to construct a formal system which can be said to contain the whole of mathematics.

For various reasons axiomatic set theory is a very suitable choice for such a system and it is one which has proved acceptable to both logicians and mathematicians.

The particular demands of mathematicians and logicians, however, are not the same. As a result there exist at the moment two different formulations of set theory which can be roughly said to cater for logicians and mathematicians respectively. It is these systems which are the subject of this dissertation. The system of set theory constructed for logicians is by P. Bernays. This will be discussed in chapter II. For mathematicians N. Bourbaki has constructed a system of set theory within which he has already embedded a large part of mathematics. This system will be discussed in chapter III.

Chapter I is historical and contains some of Cantor's original ideas.

The relationship between Bernays' system and (essentially) Bourbaki's system is commented upon in chapter IV.

CONTENTS

Contents continued. Page

CHAPTER I

Cantor's theory of aggregates

In this chapter we shall set forth some of Cantor's original ideas* and then briefly sketch the path of their later modifications.

It has been maintained that one of the major triumphs of Cantor's theory of aggregates is that it places the problem of 'completed infinities' (represented by cardinal and ordinal numbers) on a firm basis thus repudiating Gauss' famous dictum "I protest ......against the use of infinite magnitudes as something completed, which is never permissable in mathematics". Now although it is by no means unanimously agreed upon (among workers in 'Foundations') that Cantor's theory of aggregates does constitute such a repudiation, we shall in this chapter view the cardinal and ordinal numbers of Cantor as capable of representing 'completed infinities' and select those of Cantor's ideas which have a bearing on this notion.

Fundamental to the whole theory of aggregates is the definition of an aggregate.

'By an aggregate (Menge) we are to understand any collection into a whole (Zusammenfassung zu einem Ganzen) M of definite and separate

---

* Unless otherwise stated the quotations given in this chapter are taken from P. Jourdain 1915; our page numbers refer to the Dover reprint.

objects m of our intuition or our thought. These objects are called the elements of M. In symbols we express this thus,

$$M = \{m\}.$$

From this a <u>partial aggregate</u> or <u>part of an aggregate M</u> is defined as any other aggregate M_1 whose elements are also elements of M. Three important conceptions are implicit in this definition of an aggregate:

(a) Aggregates themselves as 'wholes' become 'definite and separate objects of our intuition' and can thus be elements of further aggregates.

(b) There is no restriction whatsoever on the formation of aggregates. Any method of collecting together different objects (e.g. by listing them, or requiring them to satisfy some condition) will yield an aggregate.

(c) All objects are elements of a single aggregate which comprehends them all.

Later in this chapter (pp.32-3) and in chapter II (pp.41-3) we shall see the important roles played by the above conceptions in the axiomatisation of Cantor's ideas.

Upon the above notion of aggregate Cantor constructed two fairly separate theories; the theory of cardinal numbers and the theory of ordinal numbers.

We shall (following Cantor) deal with the theory of cardinal numbers first.

Section 1. Cardinal numbers.

'We will call by the name of power or cardinal number of M the general concept which, by means of an active faculty of thought, arises from the aggregate M when we make abstraction of the nature of its various elements m and of the order in which they are given. We denote the result of this double act of abstraction, the cardinal number or power of M, by $\bar{\bar{M}}$.'

Thus every aggregate M has an entity $\bar{\bar{M}}$ correlated with it. Cantor conceived this entity as an aggregate, the elements of which were characterless 'units' and says of it that 'this number has existence in our mind as an intellectual image or projection of the given aggregate.'

Pursuing the notion that cardinal numbers embody the magnitude of aggregates Cantor investigates how the relations of 'equality' and 'greater than' and the operations of 'addition', 'multiplication' and 'exponentiation' can be defined for cardinal numbers.

Relation of equality

This rests on the notion of the equivalence of two aggregates. 'We say that two aggregates M and N are equivalent, in symbols $M \sim N$ or $N \sim M$, if it is possible to put them, by some law, in such a relation to one another that to every element of each one of them, corresponds one and only one element of the other.' (p.40).

Cantor then argues* that equivalence of aggregates M   N and equality of cardinal numbers $\bar{\bar{M}} = \bar{\bar{N}}$ are one and the same thing, in the sense that from either one of them the other can be deduced. Thus $\bar{\bar{M}} = \bar{\bar{N}}$ if and only if $M \sim N$.

<u>Relation of 'greater than'</u> (p. 89).

If between two aggregates M, N with cardinal numbers $a = \bar{\bar{M}}$, $b = \bar{\bar{N}}$, the following two conditions hold then <u>a is said to be less than b</u> (or b greater than a), in symbols $a < b$:

(i) There is no part of M which is equivalent to N.

(ii) There is a part N_1 of N such that $N_1 \sim M$.

Cantor proves that the relation so defined has the properties (in modern terminology) of irreflexivity (not $a < a$), asymmetry ($a < b$ implies not $b < a$) and transitivity ($a < b$ and $b < c$ implies $a < c$).

Using this definition and assuming the theorem on the comparability of cardinals in the form which asserts that if a and b are any two cardinal numbers then either $a = b$ or $a < b$ or $b < a$, Cantor is able to prove the following theorem:

If two aggregates M and N are such that M is equivalent to a part N_1 of N and N to a part M_1 of M then M and N are equivalent.

---

* We use this word advisedly since there is no question of a <u>proof</u> here from the above 'definition' of a cardinal number. However Cantor's intuition was quite in order. See chapter II p. 71 .

This theorem, which was later proved without the assumption of comparability by Schröder and Bernstein independently, is basic to the whole theory of cardinal numbers.

The addition of two cardinal numbers (p. 91).

The sum of two cardinal numbers $a = \overline{\overline{M}}$, $b = \overline{\overline{N}}$ is obtained by uniting M and N into a single aggregate (M,N) termed the union aggregate (Vereingungsmenge) and taking the cardinal number of this new aggregate:

$$a + b = \overline{\overline{(M,N)}}$$

The union aggregate (M,N) as Cantor defines it here, is only defined for disjoint sets. Thus 'we denote the uniting of many aggregates M,N,P.....which have no common elements, into a single aggregate by (M,N,P,.....). The elements of this aggregate are, therefore, the elements of M, of N, of P,.....taken together'. (p. 85).

There is however no necessity for this restriction to disjoint sets as was later shown by Zermelo who proved the general theorem: If S is any set one can form a disjointed set S' which is equivalent to S and whose members are equivalent to the members of S in a well-defined* sense.

In this connection we note that Cantor points out that the sum a + b in no way depends on the 'representative' aggregates M and N,

* see Fraenkel/Bar Hillel 1960 p. 126.

since if $M' \sim M$ and $N' \sim N$ then the result $(M', N') \sim (M,N)$ follows easily from the definition of equivalence. (However, since the time of Zermelo's axiomatisation of the ideas contained here, it has been know that the Choice-axiom is invoked in 'selecting' the representative aggregates).

Multiplication of cardinal numbers. (p. 92).

This is based on the notion of the aggregate of pairs (Verbindungsmenge) which is defined as follows:

'Any element m of an aggregate M can be thought to be bound up with any element n of another aggregate N so as to form an element (m,n); we denote by (M.N) the aggregate of all these pairs (m,n) and call it the aggregate of pairs of M and N. Thus $(M.N) = \{(m,n)\}$.'

From this, the product of a and b, a.b is defined as $a.b = \overline{\overline{(M,N)}}$. As with the sum, the product a.b is independent of the representative aggregates M and N with, of course, the same qualifying remarks as before.

Exponentiation of cardinal numbers. (p. 94).

This is based on the important notion of covering.

'By a covering of the aggregate N with the elements of the aggregate M or, more simply, by a covering of N with M, we understand a law by which with every element n of N a definite element of M is bound up, where one and the same element of M can come repeatedly into application. The element of M bound up with n is, in a way, a one-valued function of n, and may be denoted by f(n); it is called a covering function of n. The corresponding covering of N will be called f(N).' Cantor terms the totality of different coverings of N with M the covering aggregate (Belegungsmenge) of N with M and denotes it by (N/M). Thus

$$(N/M) = \{f(N)\}$$

If now $a = \overline{\overline{M}}$, $b = \overline{\overline{N}}$, then exponentiation is defined as

$$a^b = \overline{\overline{(N/M)}}$$

Using this definition Cantor obtains the following three basic index theorems:

$$a^b \cdot a^c = a^{b+c}$$

$$a^c \cdot b^c = (ab)^c$$

$$(a^b)^c = a^{bc}$$

If now $\aleph_0$ is defined as the cardinal number of the aggregate of all the integers*, i.e.

$$\aleph_0 = \{\overline{\overline{n}}\}$$

then the following two theorems are not difficult to obtain

(a) $(2^{\aleph_0})^n = 2^{\aleph_0}$

(b) $(2^{\aleph_0})^{\aleph_0} = 2^{\aleph_0}$

Now in view of our definition of covering, $2^{\aleph_0}$ can be interpreted

---

* Strictly the natural numbers should have been defined with respect to aggregates in some way for these results to have their full depth. This can be done, as we shall see below.

as the power of the continuum (this is clearly seen if the continuum is considered as the totality of all binary representations

$$x = \frac{f(1)}{2} + \frac{f(2)}{2^2} + \frac{f(3)}{2^3} + \ldots\ldots$$

where $f(n) = 0$ or 1).

The results (a), (b) thus express the facts that the n-dimensional and $\aleph_0$-dimensional continua have the same power ('same number of points') as the one-dimensional continuum.

The depth of these results bears testinomy to the fertility of the notion introduced so far. Indeed, Cantor himself remarks (with reference to the results (a), (b)) 'Thus the whole contents of my paper in Crelle's Journal (vol. 84, 1878) are derived purely algebraicly with these few strokes of the pen from the fundamental formulae of the calculation with cardinal numbers' (p. 97).

<u>Finite cardinals</u> (p. 97).

As well as providing a theory of the 'actually infinite or transfinite cardinals' Cantor recognised that his theory of aggregates could equally well be used to give a foundation to the theory of finite numbers.

'To a single thing e_0, if we subsume it under the concept of an aggregate $E_0 = (e_0)$, corresponds as cardinal number what we call "one" and denote by 1; we have

$$1 = \overline{\overline{E}}_0$$

Cantor then forms the union aggregate E_1, of E_o and a new element e_1 thus $E_1 = (E_o, e_1) = (e_o, e_1)$ (α)
and "two" is now defined as the cardinal number of E_1, thus $2 = \overline{\overline{E}}_1$, and in general

$$n = \overline{\overline{E}}_{n-1} \text{ where } E_2 = (E_1, e_2), E_3 = (E_2, e_3) \text{ etc.}$$

Having defined the entities E_o, E_1, etc. which represent the integers, Cantor then shows that the usual properties of integers are provable for them. He also proves several general theorems concerning finiteness, the most important of which (from the axiomatic standpoint) is the following:
Every finite aggregate E is such that it is not equivalent to any of its parts.

This theorem characterises, for Cantor, the idea of finiteness. It is by no means the only such characterisation, as we shall see in the next two chapters.

We shall return to Cantor's definition of finiteness when we discuss Bourbaki's definition, and for the moment we note the following points.

(a) In equation (α) Cantor fails to distinguish between the aggregate E_o whose sole element is e_o and the 'single thing' e_o itself (a distinction first made by Frege and Peano). Here, in fact, the erroneous assertion of equality $((E_o, e_1) = (e_o, e_1)$ is not necessary to Cantor's definition and can be dispensed with.

(b) the infinity of distinct objects e_v ($v = 0, 1, 2, \ldots$) used by Cantor to carry out his construction of the finite cardinals is not needed and can be replaced by one object, e_o (or 0, the null set) and an operation on sets which exploits the very distinction indicated in (a). Thus the objects of the following sequence will serve instead of e_o, e_1, etc.

$$e_o, \; [e_o], \; [[e_o]] \; , \text{etc}$$

where $[x]$ denotes the set whose sole element is x. (see chapter II p. 53).

Transfinite aggregates

Cantor defines transfinite aggregates as aggregates which are not finite (according to the definition of finiteness given above) so that the characterising property for transfinite aggregates is the following: 'Every transfinite aggregate T is such that it has parts T_1 which are equivalent to it'.

We have already encountered the cardinal numbers of two such aggregates, namely the cardinals $\aleph_o$ and $2^{\aleph_o}$, for which the relation $\aleph_o < 2^{\aleph_o}$ holds. Just as $2^{\aleph_o}$ was produced from $\aleph_o$ by exponentiation, so higher cardinals can be produced by repeating the operation*, and we would obtain $\aleph_0$, $2^{\aleph_0}$, $2^{2^{\aleph_0}}$ etc.

In general, the cardinal numbers described in this section generalise the notion of magnitude from the finite to the transfinite and the particular cardinal numbers $\aleph_0$, $2^{\aleph_0}$ etc. are certainly 'completed infinities' in this sense. There is, however, besides magnitude, another familiar aspect of the finite aggregates (i.e. numbers). This is their ordinal aspect, exemplified in their use for counting, and it is the generalisation of this aspect that is accomplished by the theory of order which we shall now describe.

---

* This is not the only way of producing higher cardinal numbers from $\aleph_0$. We shall meet another method in section 2 which produces the sequence of cardinals $\aleph_0$, $\aleph_1$, $\aleph_2$, etc. such that $\aleph_0 < \aleph_1 < \aleph_2$, etc. The connection between the two methods of generation has not yet been satisfactorily explained. (See chapter IV).

Section 2. Theory of Order

Cantor calls an aggregate M simply* ordered if 'a definite order of procedence (Rangordnung) rules over its elements m, so that, of every two elements m_1 and m_2, one takes the lower and the other the higher rank, and so that, if of three elements m_1, m_2 and m_3, m_1, say is of lower rank than m_2 and m_2 is of lower rank than m_3, then m_1 is of lower rank than m_3' (p. 110).

Now just as the cardinal number of an aggregate embodied the magnitude of that aggregate, so an entity is needed which embodies the 'orderedness' of an aggregate. Cantor proceeds as follows. 'Every ordered aggregate M has a definite ordinal type or more shortly a definite type which we will denote by $\bar{M}$. By this we understand the general concept which results from M if we only abstract from the nature of the elements m and retain the order of procedence among them. Thus the ordinal type $\bar{M}$ is itself an ordered aggregate whose elements are units which have the same order of procedence amongst one another as the corresponding elements of M.'

---

* This adjective was used by Cantor to distinguish the notion of order defined here from more complex notions of order. However these latter notions were not developed, and we shall follow Cantor and omit the word "simply".

As equivalence is the notion which allows the comparison of cardinals, so similarity is the corresponding notion for types.

'We call two ordered aggregates M and N similar (ähnlich) if they can be put into a biunivocal correspondence with one another in such a manner that, if m_1 and m_2 are any two elements of M and n_1 and n_2 the corresponding elements of N, then the relation of rank of m_1 to m_2 in M is the same as that of n_1 to n_2 is N. Such a correspondence of similar aggregates we call a mapping (Abbildung) of these aggregates on one another'. (p. 112).

We shall not pursue these ideas for in the first place notions, almost completely analogous to the ones for cardinals, can be defined whose manipulation leads to similar theorems.

Much more important than this, for our purposes, is that, whereas cardinals do completely embody magnitude and thus generalise that aspect of number, this is by no means the case with ordinal types; they are not analogous to the finite numbers considered ordinally. The reason for this is that the property of being simply ordered does not fully characterise the finite ordinals - they possess the stronger ordering property of being well-ordered.

Cantor defined an ordered aggregate F to be well-ordered if it satisfied the following two conditions.

'The elements f of F ascend in a definite succession from a lowest f_1 such that

(a) There is in F an element f_1 which is the lowest in rank.

(b) If F' is any part of F and if F has one or many elements of higher rank then all elements of F', then there is an element f' of F which follows immediately after the totality F', so that no elements in rank between f' and F' occur in F.' (p. 137).

A primary consequence of this definition is that every part F_1 of a well-ordered aggregate F has a lowest element.

It is usual now to take this consequence as the definition of a well-ordered set and to derive the properties (a) and (b) from it.

The ordinal types of well-ordered sets are entities which do indeed constitute generalisations of the finite ordinal numbers. In recognition of this fact Cantor termed these entities <u>ordinal numbers</u>.

The operations of addition, multiplication etc. for ordinals (i.e. ordinal numbers) are, of course, special cases of those defined for ordinal types in general. The consequent theory of ordinals, however, because it is a specialisation, contains much deeper and more particular results than does the general theory of ordinal types. One such result is the comparability of ordinals; to describe this we need first the idea of a segment of an ordinal-

'If f is any element of a well-ordered aggregate F which is different from the initial element f_1, then we shall call the aggregate A of all elements of F which precede f a segment (Abschnitt) of F or more fully the segment of F defined by the element f.' (p. 141).

The theorem asserting the comparability of ordinals can now be stated:

If F, G are any two well-ordered aggregates then one of the following three cases must hold.

(a) F, G are similar to one another.

(b) There is a definite B_1 of G to which F is similar.

(c) There is a definite segment A_1 of F to which G is similar.

Thus of two unequal ordinals it can always be asserted that one is 'higher' than the other. The proof of this theorem rests on a good deal of subsidiary work on the notion of similarity. The character of this work, however, has a certain transparency and the comparability theorem is a more or less straight forward consequence of the definition of an ordinal. It is this transparency to which we are calling attention here. Ordinals are much simpler entities than cardinals in spite of the stronger intuitive aspect of cardinals as sheer magnitude, and thus results for cardinals, analogous to easily obtained results for ordinals, are often difficult to obtain within the theory of cardinals.

Cantor recognised that this disparity between results for cardinals and ordinals could, to some extent, be overcome if a proof could be given that any aggregate can be well-ordered (i.e. if it could be proved that for any aggregate there exists an ordering of its elements which is a well-ordering). Cantor did not, in fact, manage to prove this theorem, and it was not until 1908 that it was proved - by Zermelo. This brings us back to comparability. For the importance of the comparability theorem for ordinals is that it, together with Zermelo's well-ordering theorem, furnishes a comparability theorem for <u>cardinals</u>. Since, given any two aggregates A, B we can, by the well-ordering theorem, assert the existence of two well-ordered aggregates A', B' such that the elements of A are the same as those of A' and similarly for B, B'; thus $A \sim A'$ and $B \sim B'$. The three mutually exclusive alternatives (a) - (c) for ordinals guaranteed by the comparability theorem for ordinals thus become three mutually exclusive alternatives for cardinals (according to the definition of ' $<$ ' on p. 4 above); so that one of the following three cases must hold.

(a)' $\overline{\overline{A}} = \overline{\overline{B}}$

(b)' $\overline{\overline{A}} < \overline{\overline{B}}$

(c)' $\overline{\overline{B}} < \overline{\overline{A}}$

The situation is a little more complicated than we have made it appear, however, since, to any aggregate A, there exist many ordinals A' with the above property and it is necessary to take the least one having this property. This point will become clearer in the next section where we shall be concerned with the particular case $\overline{\overline{A}} = \aleph_0$.

We shall finish this section by giving Cantor's account of one of the most important ideas in his theory of ordinals - the notion of a limit number.

First, Cantor calls any series of ordinals $\alpha_1, \alpha_2, \ldots\ldots$, which is similar to the series 1, 2,....., <u>a fundamental series</u> (or more precisely a fundamental ascending series of ordinals).

Now let $\beta_1, \beta_2, \ldots\ldots$(1) be any series of distinct ordinals (not necessarily a fundamental series) such that $\beta_n = \overline{G}_n$. Then Cantor proves that the aggregate $G = (G_1, G_2, \ldots\ldots)$ is a well-ordered aggregate, whose order type is thus an ordinal, and he takes* this ordinal to be the sum, β, of the series (1), so that

$$\beta = \overline{G} = \beta_1 + \beta_2 + \ldots\ldots\beta_n + \ldots\ldots$$

From the series (1) Cantor forms a new series

$$\alpha_1\ \alpha_2, \ldots\ldots, \alpha_n, \ldots\ldots(2)$$

which is a fundamental series, by taking

$$\alpha_n = \underline{\beta_1 + \beta_2 + \ldots + \beta_n}$$

i.e. $\alpha_n = (G_1, G_2 \ldots\ldots G_n)$

* In accordance with his definition of the ordinal sum of $\alpha = \overline{A}$, $\gamma = \overline{G}$ as $\alpha + \gamma = \overline{(A, G)}$.

so that $\beta_1 = \alpha_1, \ldots\ldots, \beta_{n+1} = \alpha_{n+1} - \alpha_n$

and $\alpha_{n+1} > \alpha_n$.

Cantor then expresses the relationship which holds between β and the series (2) thus (p. 157):

(a) The number β is greater than α_n for every n, because the aggregate $(G_1, G_2, \ldots\ldots G_n)$ whose ordinal is α_n is a segment of the aggregate G which has the ordinal β.

(b) If β' is any ordinal less than β, then, from a certain n onwards, we always have $\alpha_n > \beta'$. For, since $\beta' < \beta$, there is a segment B' of the aggregate G which is of type β'. The element of G which determines this segment must belong to one of the parts G_n; we will call this part G_{n_0}. But then B' is also a segment of $(G_1, G_2, \ldots\ldots G_{n_0})$ and consequently $\beta' < \alpha_{n_0}$ for $n \geqslant n_0$. Thus β is the ordinal number which follows next in magnitude after all the numbers α_n; accordingly we will call it the <u>limit</u> <u>(Grenze)</u> <u>of the number α_n for increasing n</u> and denote it by $\operatorname{Lim}_n \alpha_n$ so that we have

$$(3) \quad \operatorname{Lim}_n \alpha_n = \alpha_1 + (\alpha_2 - \alpha_1) + \ldots\ldots (\alpha_{n+1} - \alpha_n) + \ldots$$

Thus Cantor has proved the following theorem:

'To every fundamental series $\{\alpha_n\}$ of ordinals belongs an ordinal $\operatorname{Lim}_n \alpha_n$ which follows next, in order of magnitude after all the numbers α_n; it is represented by the formula (3).'

In the next section we shall see how this theorem plays a vital rôle in the construction of the second number class.

Section 3. The second number class.

If A, B are any two finite aggregates such that $A \sim B$ (i.e. $\overline{\overline{A}} = \overline{\overline{B}}$) then Cantor proves that $\overline{A} = \overline{B}$ i.e. all ordered aggregates of a given finite cardinal number have one and the same ordinal type. Thus 'finite ordinals coincide in their properties with finite cardinals'. However, 'the case is quite different with the transfinite ordinal numbers; to one and the same transfinite cardinal number α belong an infinity of ordinal numbers which form a unitary and connected system. We will call this system the number class Z (α)' (p. 159).

The simplest such number class is obtained when $\alpha = \aleph_0$, and Cantor termed this the second number class. (He understood by the first number class the aggregate $\{n\}$ of finite ordinals).

Thus the second number class is a totality of ordinals which comprises all types of well-ordered aggregates having the cardinal number $\aleph_0$.

The following four theorems (proved by Cantor) are sufficient to furnish us with a broad intuitive picture of the second number class.

Theorem 1

The second number class has a least number

$$\omega = \lim_n n$$

Theorem 2

If α is any number of the second number class, the number $\alpha + 1$ follows it as the next greater number of the second number class.

Theorem 3

If $\alpha_1, \alpha_2, \ldots\ldots;$ is any fundamental series of numbers of the first or second number class, then the number $\operatorname{Lim} \alpha_n$ following them next in order of magnitude belongs to the second number class.

Theorem 4

Every number α of the second number class is such that either

(a) it arises out of the next smaller number α_{-1} (if there is such a one) by the addition of 1: $\alpha = \alpha_{-1} + 1$

or (b) there is a fundamental series $\{\alpha_n\}$ of numbers of the first or second number class such that $\alpha = \operatorname{Lim}_n \alpha_n$.

We can now construct our picture. The following ordinals are elements of $Z(\aleph_0)$

ω by Th.1

$\omega + 1, \omega + 2, \ldots\ldots \omega + n, \ldots\ldots;$ by Th.2

$\omega.2 = \operatorname{Lim}_n \omega + n$ by Th.3

$\omega.2 + 1, \omega.2 + 2, \ldots\ldots \omega.2 + n$ by Th.2

$\omega.3 = \operatorname{Lim}_n \omega.2 + n$ by Th.3

continuing this process we obtain (amongst the ordinals generated) the ordinals:

$$\omega.4, \quad \omega.5, \ldots\ldots \omega.n, \ldots\ldots;$$

(we have omitted $\omega.4 + 1, \ldots\ldots; \quad \omega.5 + 1, \ldots\ldots$ etc).

These form a fundamental series and hence by Th. 3 have a limit ordinal

$$\omega^2 = \operatorname*{Lim}_n \ \omega.n$$

similarly $\qquad \omega^3 = \operatorname*{Lim}_n \ \omega^2.n$

In this way a new fundamental series

$$\omega^2, \ \omega^3, \ldots\ldots \omega^n, \ldots\ldots;$$

with limit ordinal $\omega^\omega = \operatorname*{Lim}_n \omega^n$ is formed; repeating this with the series

$$\omega^\omega, \ \omega^{(\omega^2)}, \ldots\ldots \omega^{(\omega^n)}, \ldots\ldots;$$ we get

$$\omega^{(\omega^\omega)} = \operatorname*{Lim}_n \omega^{(\omega^n)}$$ giving yet another

fundamental series:

$$\omega^\omega, \omega^{\omega^\omega}, \omega^{\omega^{\omega^\omega}}, \ldots\ldots;$$

The limit of this series cannot be written using a finite number of symbols ω in the exponent. It is called

ϵ_0 - the <u>first</u> <u>epsilon</u> <u>number</u>.

It has the property (not possessed by any smaller ordinal of $Z\ (\aleph_0)$) that it satisfies the equation $\omega^x = x$ (1)

(i.e. $\omega^{\epsilon_0} = \epsilon_0$).

The number ϵ_0 however is by no means the 'final number' of $Z\ (\aleph_0)$.

For the ordinals $\epsilon_0 + 1$, $\epsilon_0 + 2, \ldots\ldots$ $\epsilon_0 + n \ldots\ldots$; all belong to $Z(\aleph_0)$ and serve to start again the process of generating elements of $Z(\aleph_0)$. In fact Cantor calls all ordinals satisfying equation (1) <u>epsilon</u> <u>numbers</u> and proves the remarkable theorem that there exists an epsilon number ϵ_α for <u>every</u> ordinal α which belongs to the first or second number class (p. 197).

The epsilon numbers are all limits of fundamental series of the form

$$\gamma_1 = \omega^{\gamma}, \gamma_2 = \omega^{\gamma_1}, \ldots\ldots, \gamma_n = \omega^{\gamma_{n-1}} \ldots\ldots;$$

where γ is an ordinal of the first or second number class. By taking $\gamma = 1$ we obtain ϵ_0 as before. To obtain ϵ_1, γ is given the value $\epsilon_0 + 1$; to obtain ϵ_2, γ is given the value $\epsilon_1 + 1$ etc.

Having thus constructed the second number class Cantor, naturally, investigates its cardinal number; this he call $\aleph_1$ and proves that it is greater than $\aleph_0$ and is in fact the next cardinal number after $\aleph_0$ (hence the notation). His proof is essentially an application of his 'diagonal argument' where the diagonal 'element' is provided by Th.3. That is, on the assumption that $\aleph_1 = \aleph_0$ the elements of $Z(\aleph_0)$ can be written as

$$\beta_1, \beta_2, \ldots\ldots \beta_n, \ldots\ldots$$

From this series a fundamental series can be formed, the limit of which, is greater than any ordinal β_n and which is an element of $Z(\aleph_0)$, thus contradicting the assumption that the above series contains all the elements of $Z(\aleph_0)$.

We shall now consider an application of the second number class to the process of 'counting', which is exemplified by the notion of induction. The induction principle for the natural numbers asserts the following.

(A) If the truth of a property P for a number n implies its truth for a number n + 1 and if it is true for 1 then it is true for all natural numbers. Expressing this roughly in symbols

$$(P(1) \,\&\, (P(n) \rightarrow P(n+1))) \rightarrow P(n) \text{ for all } n.$$

This principle can be generalised to give induction over the elements of $Z(\aleph_0)$ by inserting a clause to deal with the limit numbers, thus:

$$\text{(B)} \quad (P(\omega) \,\&\, (P(\alpha) \rightarrow P(\alpha+1) \,\&\, \underset{\alpha < \beta}{P(\alpha)} \rightarrow P(\beta))) \rightarrow P(\alpha)$$

for all ordinals α of $Z(\aleph_0)$.

The principle (B) is but a special case of the more general transfinite induction principle framed for any well-ordered set rather than for $Z(\aleph_0)$. (See chapter II p. 61).

It is seen that (B) is indeed stronger than (A) by comparing their conclusions. (A) asserts the truth of a property P for $\aleph_0$ individuals whereas (B) makes this assertion for $\aleph_1$ individuals.

The importance and fruitfulness of transfinite induction cannot be exaggerated. This is so even when it is used in a weaker form than (B). Such a weakening can be obtained by restricting the induction to apply only up to a certain element of Z ($\aleph_0$). This can be illustrated as follows. Let us suppose it is desired to list the formulae of a formal system; for our purposes these can be taken as finite strings of symbols formed from an infinite alphabet of $\aleph_0$ letters. Let us suppose also that it is necessary to list all strings of n letters before any string of n + 1 letters (for n = 1, 2,..... Now it can easily be shown that any ordinal of Z ($\aleph_0$) which is less than ω^n can be expressed in the form

$$\omega^{n-1} a_1 + \omega^{n-2} a_2 + \ldots\ldots + a_n \text{ where } a_i \ (i \leqslant n) \text{ is}$$

a finite ordinal. There is thus a (1,1) correspondence between n - tuplets and ordinals of Z ($\aleph_0$) less than ω^n. Hence our list of strings will have ordinal number ω^ω, and induction up to this ordinal can be used to deal with this list; in fact Gentzen in his proof of the consistency of number theory (1936, 1938) used transfinite induction up to ϵ_0.

At first glance the weakened form of (B) illustrated above appears to be nothing more than an application of (A), since any ordinal of Z ($\aleph_0$), in particular ω^ω, is the order type of a denumerable set and by re-arrangement of this set it would be possible to apply the principle (A). Now although this is true it is not the complete story and the situation is somewhat more subtle than this*. In any case, use of a weakened form of (B) has justification in that it is very often considerably simpler than the corresponding numeral induction**.

What emerges from the above discussion of induction is that the infinite ordinals constructed by Cantor (in particular the members of $Z\ (\aleph_0)$) are certainly 'completed infinities' and that they permit the extension of counting operations beyond the finite aggregates.

---

* See Kleene 1952 pp. 476-479. The essential point is that such a rearrangement would entail a change in the character of the property P in (B).

---

* E.g. where induction occurs within the basis of another induction. See Kleene loc. cit.

Section 4. Zermelo - Fraenkel set theory.

As is well known Cantor's theory of aggregates is inconsistent; since, using modes of reasoning which Cantor* used Russell constructed an antinomy (see p. 32 below) within the theory of aggregates. The discovery of this antinomy gave the main impetus to a reformulation of Cantor's ideas on a more rigorous basis. The increasing tendency to axiomatise branches of mathematics (along the lines of Peano's axiomatisation of arithmetic) also helped to bring about this reformulation.

The original axiomatisation of Cantor's theory of aggregates was carried out by Zermelo in 1908. Since then there have been additions to and modifications of this original axiomatisation resulting in the Zermelo-Fraenkel (-Skolem) set theory which is normally referred to as ZF. It is ZF that we shall briefly describe, commenting later in this section on the difference between it and Zermelo's original version.

---

* Cantor himself formulated an antinomy which is known as Cantor's paradox. He asks whether the cardinal number of the aggregate of all aggregates is the largest cardinal number. By the very definition of such a number the answer is yes and yet the cardinal number of the aggregate of parts of the aggregate of all aggregates possesses a greater cardinal number.

ZF set theory consists of nine axioms ZF1 - ZF9 and employs the relation of membership ($\in$) as primitive. The notion of set in ZF, which formalises that of aggregate, is undefined - in the sense that no explicit definition, analogous to Cantor's definition of an aggregate is given for it. This illustrates immediately the difference between Cantor's approach and that of Zermelo (and all his successors). Cantor abstracted from a given 'world' and he thus took it as meaningful to define an aggregate in terms of notions related to this world. Zermelo on the other hand assumed nothing but his axioms (for the question of his logic, see p. 30 below) and constructed everything from these.

The basic logic of ZF is usually taken to be the predicate calculus (e.g. as set out in Hilbert/Ackermann 1938) which we shall not however describe here.

The axioms of ZF

ZF1. Extensionality.

If two sets x, y contain the same members then they are equal, in symbols*

$$(x)(y)(w)((w \in x \leftrightarrow w \in y) \rightarrow x = y$$

If, as we stated above, ' $\in$ ' is the only primitive relation in ZF then '=' must be defined in terms of it; this can be done by

---

* For the logical symbols used in this section see chapter II.

taking $x = y$ as an abbreviation for $(z)(x \in z \leftrightarrow y \in z)$. If, on the other hand, '=' is taken as a second primitive relation (as it is, in the set theories of Bernays and Bourbaki) then ZF1 must be strengthened by a further extensionality axiom (E1 Bernays) or by an addition to the basic logic (S6 Bourbaki). These strengthenings ensure that '=' has the usual properties of an equivalence relation and they are not needed when the above definition is used, since the properties in question can then be derived from it. (See Fraenkel/Bar Hillel pp. 28-33).

<u>ZF2</u>. <u>Pair set</u>.

For any two distinct sets x, y there exists a set w which contains just x, y.

$$(Ew)(z)(z \in w \leftrightarrow (z = x \vee z = y)$$

<u>ZF3</u>. <u>Power set</u>.

For any set x there exists a set y whose numbers are just all the subsets of x.

$$(x)(Ey)(z)(z \in y \leftrightarrow z \subseteq x)$$

<u>ZF4</u>. <u>Union</u>.

For any set z there exists a set y, the union of z, whose members are just the members of the members of z

$$(z)(Ey)(x)(x \in y \leftrightarrow (Ew)(x \in w \ \& \ w \in z))$$

<u>ZF5</u>. <u>Infinity</u>.

There exists an infinite set

$$(Ez)(0 \in z \ \& \ (x)(x \in z \rightarrow [x] \in z)$$

where $[x]$ denotes the set whose sole member is x.

ZF6. Sifting - Aussonderungsaxiom.

For any set z and any well-defined condition P(x) (see below) there exists a subset y of z whose members are just the members of z which satisfy the condition P(x).

$$(z)(Ey)(x)(x \in y \leftrightarrow (x \in z \;\&\; P(x)))$$

By a well-defined condition is meant any formula (with a free variable) which is built up from the 'atomic' formula $a \in b$ and the logical constants in a recursive manner. (See Bourbaki's critères formatifs or Fraenkel/Bar Hillel p. 272). It will be more convenient to comment on the nature of this axiom* in the next section.

ZF7. Choice.

For any disjoint set x, whose members are non-empty, there exists at least one subset u of the union x, with the property that u has one and only one element in common with every member of x.

$$(x)\{(y)(z)((y \in x \;\&\; z \in x) \rightarrow ((Es)(s \in y \;\&\; (\overline{Ew})(w \in y \;\&\; w \in z))))$$
$$\rightarrow (Eu)(y)(y \in x \rightarrow (Ev)(t)(t = v \leftrightarrow (t \in u \;\&\; t \in y)))\}$$

ZF8. Replacement (Substitution).

For every set t and every single-valued function F(x) which is defined for the members of t, there exists the set that contains all F(x) with $x \in t$.

---

* Axiom-schema to be precise, since only the substitution of a particular well-defined condition for P(x) will yield an axiom.

$$(x)(x')((x \in t \; \& \; x' \in t \; \& \; F(x) = F(x') \rightarrow x' = x)) \longrightarrow$$

$$(Es)(y)(y \in s \leftrightarrow (Ex)(x \in t \; \& \; y = F(x))$$

This axiom-schema in fact contains as a special case the schema ZF6 above but we include ZF6 in this enumeration of the axioms of ZF because it shows quite clearly the restrictions which have had to be imposed on the formation of sets to prevent the formulation of antinomies (see next section). The notion of a function which is employed here is interpreted in the same way as that of well-defined condition in ZF6.

<u>ZF9.</u> <u>Foundation (Fundierungsaxiom).</u>

Every non-empty set s contains a member t such that s and t have no common member

$$(s)\left\{(Ex)(x \in s) \rightarrow (Et)(t \in s \; \& \; \overline{(y)(y \in s \; \& \; y \in t))}\right\}$$

The system embodied in ZF1 - 9 (with the predicate calculus as basic logic) differs from Zermelo's original system in the following respects.

(1) Zermelo's system did not possess a fully formalised basic logic such as the predicate calculus; perhaps because of this, Zermelo was not able to give a satisfactory definition of the predicate occurring in ZF6. The definition given above of using prime formulae and building up the predicates in a recursive way is due to Skolem.

(2) Zermelo's axiom of extensionality differed from ZF1 since he was concerned with the possible existence of memberless sets (see chapter II p. 52).

(3) ZF8 was added to Zermelo's system by Fraenkel in 1921. Intuitively it permits one to replace the members of a set by any previously defined sets. It was used explicitly by Cantor in roughly the same form as is formalised in ZF8. (See Wang/McNaughton 1952 p. 18).

(4) ZF9 was added by Von Neumann in 1925. The purpose of this axiom is to exclude the so-called extraordinary sets (ensembles extraordinaires) which have the property of an infinite regression of membership or of cyclic membership. Thus the two sets x, y displaying the following properties are extraordinary sets:

$$\ldots\ldots \in x_2 \in x_1 \in x$$

$$y \in u \in v \in y$$

The existence of such sets appears not to have occurred in Cantor's ideas. They are compatible with ZF1 - 8 but their existence cannot be inferred from these axioms (see e.g. E. Specker 1957).

Section 5. The conceptions of Von-Neumann and Bernays.

The ZF system of set theory as embodied in ZF1 - ZF9 has come to be regarded as the standard axiomatisation of Cantor's ideas. This is accounted for, quite naturally, by the historical closeness of Zermelo's original paper to Cantor's theory of aggregates. In addition, set theory, is the form of ZF, has played an increasingly important role in the foundations of mathematics.

Thus newer systems of set theory tend to be regarded as modifications of, or departures from, ZF; this is in fact how we shall view the systems of Bernays and Bourbaki in chapters II and III.

In formalising conceptions (a) and (b) of Cantor's definition of an aggregate (above p. 2), there are two alternatives:

(A) To retain the freedom indicated in conception (a) whereby any collection (i.e. set) formed in accordance with the axioms could be conceived as an entity capable of being a member of some other set (i.e. an entity possessing elementhood, in the sense of Quine 1940).

(B) To retain the freedom of forming a set by abstraction, i.e. forming the set $\hat{x}P(x)$ of all the objects which satisfy any given condition $P(x)$.

In order to see that a real choice is involved here, i.e. that (A) and (B) exclude each other, we must look at Russell's paradox. This paradox poses the question:

Is the set of all sets which are not members of themselves a member of itself? In symbols, is it true that $\hat{x}(x \notin x) \in \hat{x}(x \notin x)$? As is well known, both the answers 'yes' and 'no' to this question lead to a contradiction. Now the construction of the question depends on utilising first (B) to form the set $\hat{x}(x \notin x)$ and then (A) to ensure that the question itself is meaningful i.e. the set $\hat{x}(x \notin x)$ possesses elementhood.

Thus the formulation of both (A) and (B) in a system would result in a contradiction, and some sort of choice between them must be made. Zermelo chose (A)* in that he rejected (B) by disallowing the use of unqualified abstraction in the formation of sets. He admitted a qualified form of abstraction by requiring that the objects which are to form a set by abstraction must already be members of some previously secured set (this is the set z in ZF6).

It was von Neumann who discovered that a fuller** formalisation of (B) can be achieved by admitting unqualified abstraction to form sets but excluding certain of the sets so formed from elementhood. Precisely which sets are excluded from elementhood will emerge in the next chapter where we shall be concerned with Bernays system, which embodies the discovery of von Neumann's stated above.

---

* Thus any set in ZF possess elementhood; this is shown (trivially) by the existence in ZF for any x of the unit set [x].

** In the sense that no set excluded from elementhood by von Neumann can ever occur as a set in ZF. (See p.103. below).

Chapter II.

The set theory of P. Bernays.

In this chapter we shall discuss the system of set theory constructed by P. Bernays, basing our exposition on the book Axiomatic set theory by P. Bernays and A.A. Fraenkel*. This book contains a more detailed and in some respects simplified version of the set theory put forward by Bernays in the Journal of Symbolic Logic (see bibliography). We shall use the letter B. as an abbreviation for Bernays' system as put forward in the book and make reference to the book simply by giving the page number. Differences between B. and Bernays JSL will be commented upon at the appropriate places in this chapter and in chapter IV.

B. is a modification of the ZF system; its underlying logic consists of a two-fold extension of the predicate calculus. The form of the predicate calculus which is used is a standard one (cf. Hilbert/Ackermann 1938 pp. 68-70); we shall therefore not repeat Bernays' description of it, but confine ourselves to the two extensions. These are the theory of descriptions (iota operator) and the class formalism (abstraction operator).

---

* The section contributed by Fraenkel (pp. 1-44) is an historical introduction formulated as a commentary on the ZF axioms.

Section 1. Notation and primitive symbols

The objects treated in the system are of two sorts - sets and classes; the sets are basic - set variables being the individual variables of the predicate calculus.

As set variables small roman letters a,b,c,......x,y,z,s,t,u are used. These variables can occur either free or bound.

As class variables capital roman letters A,B,C,......L,M,N, ... are used. These variables can only occur free.

As syntactic variables (i.e. meta-variables denoting objects of the formal system) german letters are used; small german letters denoting set variables and capital ones denoting class variables or formulae. For typographical reasons we shall use the equivalent english letter with a bar beneath it instead of a german letter. Thus

$\underline{a},\underline{b},\underline{c},\ldots\ldots$ denote set variables

$\underline{A},\underline{B},\underline{C},\ldots\ldots$ denote formulae

$\underline{F},\underline{G},\underline{H},\underline{K},\underline{L},\underline{M},\underline{N}..$ denote class variables

Bernays designates the logical constants in the following manner: conjunction "&", negation "$^{-}$", disjunction "$\vee$", implication "$\rightarrow$", bi-implication "$\leftrightarrow$", the universal quantifier "(x)" and the existential quantifier "(Ex)".

Bernays has two further primitive 'logical' symbols:

(i) The ι- operator $\iota_{\underline{x}}(\underline{A}(\underline{x})a)$ where a is a constant of the system. (see p. 38 below).

(ii) The abstraction operator $\{\underline{x} \mid \underline{A}(\underline{x})\}$. This is, as we shall see, essentially equivalent to Russell's class symbol $\hat{\underline{x}}\, \underline{A}(\underline{x})$.

In addition to the primitive logical symbols there are the primitive set-theoretical symbols for membership* " $\in$ " and equality (between sets) " = ".

The only other primitive symbols of B. are constants (such as 0) introduced in the axioms, which thereby constitute an implicit definition of them. We shall deal with these when we meet them.

---

* In Bernays JSL a second primitive relation of membership " η ", denoting membership of a class, was adopted. This distinction between membership of a set and of a class is now considered unnecessary.

Section 2. Descriptions

To adjoin a theory of descriptions to a formal system is to add the definite article to its logical vocabulary. The simplest way of doing this is to provide a means of constructing a name for each unique object (i.e. set) which exists in the system. Thus if $\iota_{\underline{x}} \underline{A}(\underline{x})$ denotes 'the $\underline{x}$ such that $\underline{A}(\underline{x})$' then in conformity with normal usage one would only want this expression to be formable if these existed an $\underline{x}$ satisfying $\underline{A}(\underline{x})$ and this $\underline{x}$ was furthermore unique; that is, if the following existence and uniqueness clauses were provable;

(α) $(E\underline{x})\ \underline{A}(\underline{x})$

(β) $(\underline{x})\ (\underline{y})(\underline{A}(\underline{x})\ \&\ \underline{A}(\underline{y}) \rightarrow \underline{x} = y)$*

The difficulty of this method however is that it throws the admissibility of expressions as terms onto the notion of provability for which there exists no decision procedure. This would be an undesirable weakening of the usual definiteness of what constitutes a term (particularly for metalogical investigations).

If, on the other hand, the expression $\iota_{\underline{x}} \underline{A}(\underline{x})$ is admitted as a term for any formula $\underline{A}(\underline{x})$ then trivially undecidable formulae may result.

---

* We shall adopt Bernays' order of precedence $\rightarrow$, $\leftrightarrow$, & , $\vee$ for the logical constants.

Bernays' way out of these difficulties is to introduce a certain constant, a, into the ι- term $\iota_{\underline{x}}(\underline{A}(\underline{x}),a)$ so that if either (α) or (β) is not satisfied then the ι- term assumes the value of this constant. (Cf. Quine 1940, pp. 146-151 where the same course is adopted).

These considerations are embodied in the following two <u>schemata governing the ι- operator</u>.

(1) $\underline{A}(c) \& (\underline{x})(\underline{A}(\underline{x}) \rightarrow \underline{x} = c) \rightarrow c = \iota_{\underline{x}}(\underline{A}(\underline{x}),a)$

(2) $(E\underline{x})(\underline{A}(\underline{x}) \& (\underline{y})(\underline{A}(\underline{y}) \rightarrow \underline{x} = \underline{y})) \vee \iota_{\underline{x}}(\underline{A}(\underline{x}),a) = a$

(1) asserts that if there exists a unique object c satisfying $\underline{A}(\underline{x})$ then $\iota_{\underline{x}}(\underline{A}(\underline{x}),a)$ is this object; (2) asserts that either (α) and (β) hold (in which case (1) applies) or else the ι- term reduces to a.

The application of the schemata (1), (2) is simplified when the symbol 0 (null set) is introduced through axiom A1; for then it is possible to take 0 as the constant a and to define $\iota_{\underline{x}}\underline{A}(\underline{x}) = \iota_{\underline{x}}(\underline{A}(x),0)$. The schemata (1), (2) then become

(1)' $\underline{A}(c) \& (\underline{x})(A(\underline{x}) \rightarrow \underline{x} = c) \rightarrow c = \iota_{\underline{x}}\underline{A}(\underline{x})$

(2)' $(E\underline{x})(\underline{A}(\underline{x}) \& (\underline{y})(\underline{A}(\underline{y}) \rightarrow \underline{x} = \underline{y}) \vee \iota_{\underline{x}}\underline{A}(\underline{x}) = 0$

The addition of a theory of descriptions to a logical system (in the above manner, say) clearly extends the expressive power of that system and it is natural to look for attendant disadvantages. Fortunately there appear to be none, for the ι - terms prove to be eliminable* - in the sense that any theorem proved with their aid but not containing them in its statement can be proved without them. A similar situation exists with regard to class terms which we shall discuss in chapter IV.

---

* Hilbert/Bernays 1934, p. 422 ff.

Section 3. The class formalism.

The class formalism which we shall discuss in this section is perhaps the most distinctive feature of B. Its formal embodiment in B. (as an extension of the predicate calculus) is simple enough and we shall set this down first.

The class operator $\{ \underline{x} \mid \underline{A}(\underline{x}) \}$ is primitive in B; intuitively it denotes the collection of objects (sets) which satisfy the predicate $\underline{A}(\underline{x})$. The rule which formalises this interpretation is the so-called 'Church schema' which makes explicit the connection between predicates and their extensions (classes).

Church schema. $c \in \{ \underline{x} \mid \underline{A}(x) \} \leftrightarrow \underline{A}(c)$

The predicates admissible in B. are those which can be built up from the following three prime formulae (using the logical constants and quantifiers) in the well-known manner.

(1) $\underline{a} \in \underline{b}$

(2) $\underline{a} \in \underline{K}$

(3) $\underline{a} = \underline{b}$

In these formulae $\underline{a}$, $\underline{b}$ denote set terms and $\underline{K}$ denotes a class term; the set term $\iota_{\underline{x}} \underline{A}(x)$ and the class term $\{ \underline{x} \mid \underline{A}(\underline{x}) \}$ generated by the ι - schemata and the Church schema respectively contain predicates as constituent parts and it is thus necessary to define terms and formulae by simultaneous recursion. (See for example

Bourbaki I. critères formatifs p. 17). This leads to the following classification of terms, for both set and class terms:

(a) Any free variable

(b) Any individual symbol. For instance 0, the null set, introduced in axiom A.1, is an individual set symbol.

(c) Any function symbol with terms as arguments. (See p. 47 below).

(d) Any expression formed by the application of an operation symbol to an arbitrary formula or term, of which one or more occurring free variables are bound by the operation symbol. The ι-operator is an example of this kind of operation symbol.

Another rule for classes (of less importance than the Church schema) is a substitution rule analogous to the substitution rule for sets (which can be derived from the predicate calculus). The rule is $\frac{\underline{A}(C)}{\underline{A}(\underline{K})}$ which permits the replacement of the (free) class variable C by the class term denoted by $\underline{K}$.

We shall now turn to the significance and interpretation of the class formalism. First we must recall the choice of paths that was indicated in the last chapter concerning the axiomatisation of Cantor's ideas. Implicit in Cantor's definition of an aggregate were the following three conceptions.

(a) Aggregates as 'definite and separate objects' can be elements of further more comprehensive aggregates.

(b) There is no restriction on the formation of aggregates; any condition imposable on objects determines the aggregate of objects satisfying that condition.

(c) There is an aggregate comprehending all aggregates and of which all aggregates are elements.

We saw that if 'aggregate' is interpreted in the same way throughout these conceptions and if the notion of an imposable condition is suitably characterised then it is not possible to formalise (a) and (b); Zermelo chose to formalise (a) so that all sets in ZF are capable of being elements of other sets; as for (b), he restricted set formation by means of the Aussonderungsaxiom. Thus, as far as the paradoxes were concerned, his method disallowed the formation of the 'over-comprehensive' sets upon which they are based.

von Neumann's achievement can be viewed as the discovery that all the conceptions (a), (b), (c) can be formalised provided that two different interpretations of 'aggregate' are recognised; and that no paradox results from forming any collection of objects into a whole provided that a suitable limitation is imposed on such collections as to whether they are allowed to serve as elements of further collections. von Neumann formulated his ideas in terms of functions rather than collections and it was Bernays who reformulated and simplified them into the form that we are presenting here.

Thus by the Church schema every predicate of the system determines a class which is the collection of objects (sets) satisfying that predicate. This unrestricted formation of collections by abstraction is the sense in which (b) is formalised in B. - provided that 'aggregate' is interpreted as 'class'. The paradoxes are avoided here by debarring 'over-large' classes from elementhood and allowing elementhood only to the remaining classes; classes distinguished in this way, by their ability to be elements, are called sets*. It is by interpreting 'aggregate' as 'set' that (a) is also formalised in B.; furthermore by employing both interpretations, (c) is formalised in B. since the universal class $V = \{x \mid x = x\}$ is an aggregate (class) to which all aggregates (sets) belong.

This distinction between sets and classes has substance only if the 'over-large' classes are fully characterised and criteria are subsequently established for the elementhood of classes. Precisely this was done by von Neumann who excluded from elementhood (in the terminology of B.) those classes which

---

* We have expressed this loosely to achieve intuitive clarity; strictly a class can never be an element - only the set with the same members as that class can be an element (see p.48 below).

could be mapped on to the universal class; we shall see that exactly the same situation is present in B. (see p.   below).

The division of collections into sets and classes on the criterion that their elementhood would produce antinomies appears at first sight to be, to some extent, arbitrary. Bernays points out that in fact there is very little arbitrariness here in view of a profound intuitive interpretation that can be put on the difference between sets and classes - 'This distinction between sets and classes is not a mere artifice but has its interpretation by the distinction between a set as a collection which is a mathematical thing, and a class as an extension of a predicate, which in comparison with the mathematical thing has the character of an ideal object.' (p. 56).

In view of this distinction Bernays regards 'the realm of classes not as a fixed domain of individuals but as an open universe' and regards the rules stated for class formation to be the absolute minimum. The immediate practical outcome of so regarding the realm of classes is to avoid all bound class variables and to restrict quantification to sets only. As Bernays intimates this is of supreme importance when one considers B. as a whole, since no reference can be made (within B.) to the realm of classes, thus ruling out impredicative class formation; we shall return to this point in chapter IV.

A consequence of the above view of classes (as nothing more than the extensions of predicates) is that there is no need to posit* as primitive an equality relation " $\equiv$ " between classes, such a relation can be defined thus:

$$A \equiv B \leftrightarrow (x)(x \in A \leftrightarrow x \in B)$$

It is possible, using the theory of descriptions and the class formalism, to define the basic notions concerning functions. Although these notions will be formulated for classes there will be no need to reformulate them for sets since, as we shall see, for every set there exists a class having the same members.

Df.1. $A \subseteq B \leftrightarrow (x)(x \in A \rightarrow x \in B)$

A is a subclass of B.

Df.2. $\bar{A} \equiv \{x \mid x \notin A\}$

$\bar{A}$ is the complement of A

Df.3. $A \cup B \equiv \{x \mid x \in A \vee x \in B\}$

$A \cup B$ is the union of A and B

Df.3'. $\cup A \equiv \{z \mid (Ex)(x \in A \,\&\, z \in x)\}$

$\cup A$ is the union of A

Df.4. $A \cap B \equiv \{x \mid x \in A \,\&\, x \in B$

$A \cap B$ is the intersection of A and B

Df.4'. $\cap A \equiv \{z \mid (x)(x \in A \rightarrow z \in x)$

$\cap A$ is the intersection of A

---

*Cf. the same point with respect to sets (p. 50 below).

Df.5. $V \equiv \{x \mid x = x\}$

V is the universal class

Df.6. $\Lambda \equiv \{x \mid x \neq x\}$

Λ is the null class.

The following definitions need the notion of an ordered pair $\langle a, b\rangle$ (of sets), which is not available until after axioms A1 - A3. We shall discuss it there and assume it in the following

Df.7. $\{\underline{xy} \mid \underline{A}(\underline{x},\underline{y})\} \equiv \{\underline{z} \mid (E\underline{x})(E\underline{y})(\underline{z} = \langle \underline{x},\underline{y}\rangle \;\&\; A(\underline{x},\underline{y})\}$

The class of pairs $\langle \underline{x},\underline{y}\rangle$ such that $\underline{A}(\underline{x},\underline{y})$

Df.8. $Ps(A) \leftrightarrow (x)(x \in A \rightarrow (Eu)(Ev)(x = \langle u,v\rangle))$

A is a <u>pair class</u>

Df.9. $\Delta_1(A) \equiv \{x \mid (Ey)(\langle x,y\rangle \in A)\}$

$\Delta_1(A)$ is the <u>domain</u> of the pair class A.

Df.10. $\Delta_2(A) \equiv \{y \mid (Ex)(\langle x,y\rangle \in A)$

$\Delta_2(A)$ is the <u>range</u> (converse domain) of the pair class A

Df.11. $\breve{A} \equiv \{xy \mid \langle y,x\rangle \in A\}$

$\breve{A}$ is the <u>converse class</u> of the pair class A.

Df.12. $A \mid B \equiv \{xy \mid (Ez)(\langle x,z\rangle \in A \;\&\; \langle z,y\rangle \in B)$

$A \mid B$ is the composition of the pair classes A, B.

Df.13. $A \times B \equiv \{xy \mid x \in A \;\&\; y \in B\}$

$A \times B$ is the cross product of A and B.

Df.14. $\overset{\frown}{A} \equiv \{u \mid (Ex)(Ey)(Ez)(u = \langle\langle x,y\rangle,z\rangle \;\&\; \langle x,\langle y,z\rangle\rangle \in A)$

A is the class of triplets $\langle\langle x,y\rangle,z\rangle$ such that $\langle x,\langle y,z\rangle\rangle \in A$.

Df.15. $Ft(F) \leftrightarrow Ps(F) \;\&\; (x)(y)(z)(\langle x,y\rangle \in F \;\&\; \langle x,z\rangle \in F \rightarrow y = z)$.

<u>F is a function</u>

The ι - schemata can be applied to this to give

$Ft(F) \ \& \ a \in \Delta_1 F \rightarrow (\langle a,b \rangle \in F \leftrightarrow \iota_x(\langle a,x \rangle \in F) = b)$

from which we can then define

Df.16. $F {}^{\iota} a = \iota_x(\langle a,x \rangle \in F)$ giving

$Ft(F) \ \& \ a \in \Delta_1 F \rightarrow (\langle a,b \rangle \in F \leftrightarrow b = F {}^{\iota} a)$

Thus $F {}^{\iota} a$ is the <u>value of the function F for the argument a</u>. (The ι - schema (2) will yield 0 for $F {}^{\iota} a$ if a is not in the domain of F).

$F {}^{\iota} a$ is an example of a function term with terms (here set terms) as arguments. (See p.41 above).

Df.17. $Crs(K) \leftrightarrow Ft(K) \ \& \ Ft(\breve{K})$

K is a (1,1) correspondence

Df.18. $A \underset{K}{\sqcap} B \leftrightarrow Crs(K) \ \& \ \Delta_1(K \frown (A \times B)) \equiv A \ \& \Delta_2(K \frown (A \times B)) \equiv B.$

A is in (1,1) correspondence with B by means of the mapping class K.

Definitions 1 - 18 constitute the basic class formations of B.; that they can be transferred to sets is evident from the connection between sets and classes that we shall now give (see also p. 57 below). A set a is said to <u>represent</u> a class A, in symbols Rp(A,a), if both have the same elements. Formally

Df.19. $Rp(A,a) \leftrightarrow (x)(x \in A \leftrightarrow x \in a)$

As provable formulae concerning representability we have

$Rp (\{x \mid x \in a\} , a)$

$Rp (A,a) \ \& \ Rp (A,b) \rightarrow a = b$

Thus every set represents some class.

Df.20. $Rp(A) \leftrightarrow (Ex)\, Rp(A,x)$

A is represented.

Df.21. $a^{*} \equiv \{x \mid x \in a\}$

a^{*} is the class represented by the set a

From this it follows that

$$Rp(A) \leftrightarrow (Ex)(x^{*} \equiv A)$$

Using the notion of representability it is possible now to be more precise in our account of the realm of classes than we were above. In B. no class can ever be a member of a class or a set, no matter how 'small' it is; however if a class is representable, then the set which represents it can of course be a member.

The difference between the account in the above paragraph and the one given previously is nevertheless a difference of taste only. It is perfectly possible to identify a class with its representing set and then to speak of the class as being a member - this is precisely what Gödel does in his modification of Bernays JSL (see chapter IV).

One final characteristic of the class formalism must be mentioned. This is the possibility of expressing any class generated by the Church schema in terms of one or more of the following eight basic notions:

$$a^{*},\ \bar{A},\ A \wedge B,\ A \times B,\ \Delta_1 A,\ \breve{A},\ \overset{\prec}{A},\ E$$

The first seven were defined as above, the eighth, E, is a class symbol denoting the class of pairs of sets that satisfy the $\in$ - relation, that is $E \equiv \{xy \mid x \in y.\}$

The metamathematical result which guarantees this possibility is known as the class theorem. It is proved by Bernays in (JSL 1937). (Cf. chapter IV pp. 122-5).

Section 4. The equality axioms and the axioms of general set theory.

In B. equality " = " is a primitive relation between sets governed by the following two axioms:

E1. $a = b \rightarrow a \in A \rightarrow b \in A$

E2. $(x)(x \in a \leftrightarrow x \in b) \rightarrow a = b$

From E1. by substituting $\{x \ \underline{A}(x)\}$ for $\underline{A}$ and using the Church schema, we get

$$a = b \rightarrow \underline{A}(a) \rightarrow \underline{A}(b) \quad (1)$$

The main application of E1. is through (1). An immediate consequence of E2. is the theorem

$$x = x \quad (2)$$

Theorems (1), (2) are (J_2), (J_1) respectively of Hilbert/Bernays and constitute a sufficient basis for the handling of equality; in particular from (1) we obtain the theorem

$$a = b \rightarrow (x)(x \in a \leftrightarrow x \in b) \quad (3)$$

and (3) with E2. provides the following equivalence

$$a = b \leftrightarrow (x)(x \in a \leftrightarrow x \in b) \quad (4)$$

It may be asked why Bernays does not reduce his primitive vocabulary by taking (4) as a definition of equality; E2. would then be derivable. Bernays rejects this course on grounds of taste 'we want to suggest the interpretation of equality as individual identity, whereas by (4) taken as definition, equality is introduced only as an equivalence relation'. (p. 53).

A too casual reading of Fraenkel's remarks (historical introduction p. 8) on this point gives the impression that if (4) is taken as definition then the formula x s & x = y y s cannot be proved but must be posited. This is, of course, not the case here since E1. is available (where now the " = " appearing in it is an abbreviation). Thus

$x = y \rightarrow (x \in s \rightarrow y \in s)$ taking $x \in s$ for $\underline{A}(x)$ in (1)

and therefore

$x = y \;\&\; x \in s \rightarrow y \in s$ (propositional calculus)

so that the formula in question is derivable.

A second equally plausible reason for not defining equality by (4) is that such a course would introduce the set-theoretical notion of membership, (by way of equality) into the purely logical ι- schemata.

The form of the axiom of extensionality (E2.) which is taken in a system of set theory is governed by the attitude taken towards 'individuals' (Urelemente). The existence of individuals - memberless sets - is consistent with the normal set axioms (e.g. ZF2 - ZF9); so that they* can either be excluded, posited or left as an open (consistent) possibility. The first course is adopted here, the second in some systems used for independence proofs (e.g. Fraenkel 1922) and the last was adopted by Zermelo (1930). See below.

---

* The null set is excluded from these considerations.

Thus if i, j were two individuals in B. then from the two identically true formulae $(\overline{Ex})(x \in i)$, $(\overline{Ex})(x \in j)$ we obtain, by the predicate calculus, the equivalence $(x)(x \in i \leftrightarrow x \in j)$ and hence by E2, $i = j$; thus E2. renders all individuals equal and it is therefore only possible for one individual to exist in B. - this is the null set.

Besides individuals there are other abnormal sets the existence of which is consistent with the axioms; these are so-called extraordinary sets. (See chapter I p. 31). None of the procedures outlined above for individuals has any effect on the existence of such sets and it is necessary to introduce a special axiom, the Fundierungsaxiom (ZF9) to exclude* them.

Thus all objects of B. are sets (i.e. the objects to which the individual variables of the predicate calculus refer). At first glance this is somewhat repugnant to common sense, which requires some sort of basis from which to start the generation of sets. (It was presumably to provide such a basis that Zermelo 1930 retained the possibility of an infinite number of Urelemente; however the idea of 'individuals' now finds little favour; such entities are certainly not considered necessary for the construction of mathematics within set theory. Cf. the attitude of Quine 1940 pp. 121-2 and Bernays pp. 53-4).

---

* or include them if desired - the same independence proofs using individuals (cited above) can also be carried out using extraordinary sets (Mendelson 1956).

It is now recognised that the only 'individual' which is necessary in a system of set theory is the null set. In fact, the totality of sets in B. consists of the null set 0, the set whose sole member is the null set [0], etc., also sets whose members are chosen from the thus defined totality, i.e. from 0, [0], [[0]] ,......; also sets whose members are chosen from the thus supplemented totality; and so on. This totality of sets in B. is more succinctly characterised by specifying

(1) The null set.

(2) The operation of adding a member to a given set to form a new set.

(3) The operation of forming a new set from a previous set by replacing the members of the previous set by sets which have already been defined.

Bernays formalises (1) - (3) in the axioms A1 - 3 respectively, given below.

A1. $a \notin 0$

A2. $a \in b;\ c \leftrightarrow a \in b \vee a = c$

A3. $a \in \sum_{\underline{x}} (m, t(\underline{x})) \leftrightarrow (E\underline{x})(\underline{x} \in m \ \& \ a \in \underline{t}(\underline{x}))$.

Each axiom takes the form of an implicit definition of the new primitive symbol occurring in it.

A1. introduces the null set, 0. A2. introduces " ; " as a generalised 'successor' operator, i.e. if we are given an object c and wish to add it to the set b to form a new set whose members are the members of b and the object c then b;c is just such a set. A3. introduces the 'union operator' $\sum_{\underline{\underline{x}}}(m,\underline{t}(x))$ which forms the union of the sets $\underline{t}(c)$ with $c \in m$.

The form of A1 - 3 suggests (see p. 43) an extension of Peano's axioms for arithmetic from the finite to the transfinite. A closer characterisation of A2, A3 is to be found in Cantor's two principles of formation (Erzeugungsprinzip) (Jourdain pp. 56 - 7) where A2 corresponds to Cantor's first principle which he uses to generate the finite ordinals and A3 to his second principle where he takes the limit of a sequence (of ordinals) as the next element of the sequence. We shall see in the next section how A2 - 3 do in fact yield Cantor's two principles of formation when applied to well-ordered sets. Thus in casting A2 - 3 in the form given Bernays indicates a useful and important analogy between set formation in general and the narrower and somewhat simpler formation of ordinals; and we are thus able to see the assumptions needed to build a hierarchy of sets from the null set.

From A1 - 2 we get immediate characterisations of the unit set, the unordered pair (plain set) and the ordered pair:

Unit set $[a] = 0; a$

Unordered pair $[a,b] = [a]; b$

Ordered pair $\langle a,b \rangle = [[a], [a,b]]$

By taking x for term denoted by $\underline{t}(\underline{x})$ in A3 we get

$a \in \sum_{\underline{x}} (m,x) \leftrightarrow (Ex)(x \in m \ \& \ a \in x)$

For brevity $\sum_{\underline{x}}(m,x)$ can be defined as $\sum m$ and the above theorem becomes

$a \in \sum m \leftrightarrow (Ex)(x \in m \ \& \ a \in x)$

which is the assertion of ZF4, so that $\sum m$ is the sum of the elements of m. The union $a \cup b$ of two sets is defined as $\sum [a,b]$ since

$c \in \sum[a,b] \leftrightarrow c \in a \vee c \in b.$

Recalling that x^* is the class which is represented by the set x, this last formula can be interpreted as asserting the representability of $a^* \cup b^*$ i.e.

$Rp(A) \ \& \ Rp(B) \rightarrow Rp(A \cup B)$

It is natural to ask whether the same formula holds for $A \cap B$; Bernays proves that this is so and what is more that the following stronger statement holds:

$A \ \& \ Rp(B) \rightarrow Rp(A \cap B).$

This is precisely the Aussonderungsaxiom (ZF6) since it asserts that the intersection of a set (b) with a condition (the class A) is a set. The replacement axiom (ZF8) of which ZF6 is a special case, is also derivable from A3 (as is to be expected from the intuitive description of A3 above) and Bernays' statement of it is

$Ft(F) \& Rp(\Delta_1 F) \rightarrow Rp(\Delta_2 F)$

The main idea behind the derivations of ZF6, ZF8 from A3 is one of establishing class representability; this combined with the fact that axioms A1-3 can be construed as assertions of class representability (see below) provides a framework for the proofs. Thus A1, A2 as statements on representability are as follows (p. 72)

$(x)(x \notin 0) \leftrightarrow Rp(\Lambda, 0)$

$(x)(x \in b; c \leftrightarrow x \in b \vee x = c) \leftrightarrow Rp(\{x \mid x \in b \vee x \in c\}, b;c)$

A3 can be written in a similar fashion.

This mode of proof is central in B. and Bernays regards it as important since "it makes more explicit the situation which consists with regard to the role of the logical conceptions of mathematics" (p. 56). It might be added that by his use of the class formalism Bernays effects a return (perhaps the closest possible in view of the paradoxes) to the Cantorian freedom of the unfettered formation of aggregates.

In the last section we gave the definitions of pair classes, mapping classes etc., by using the * operator it is possible to transfer these definitions directly to sets, thus:

Df.8' $Ps(a) \leftrightarrow Ps(a^*)$

9' $\Delta_1 a = \iota_x(x^* \equiv \Delta_1(a^*))$

10' $\Delta_2 a = \iota_x(x^* \equiv \Delta_2(a^*))$

11' $\breve{a} = \iota_x(x^* \equiv \breve{a}^*)$

12' $a|b = \iota_x(x^* \equiv a^*|b^*)$

13' $a \times b = \iota_x(x^* \equiv a^* \times b^*)$

15' $Ft(f) \leftrightarrow Ft(f^*)$

16' $f`a \Leftrightarrow f^*`a$

17' $Crs(f) \leftrightarrow Crs(f^*)$

18' $a \sim b \leftrightarrow (Ex)(Crs(x) \ \& \ \Delta_1 x = a \ \& \ \Delta_2 x = b)$

i.e. $a \sim b \leftrightarrow (Ex)(a^* \ \overline{x^*} \ b^*)$ by virtue of Df.18.

Thus definitions formulated for classes pass easily into definitions for sets. As to how results concerning classes pass over into results concerning sets, the key to this is that the four operations of $\cup$, $\cap$, $\times$, $\smile$ all commute with respect to the $*$ operator (i.e. with respect to representability) we saw an example of this in the theorem: $Rp(A) \ \& \ Rp(B) \rightarrow Rp(A \cup B)$ which asserts the commutivity of the union operator with respect to representability, similar theorems are obtainable for the other three operations of intersection, cross product and converse.

We have in this section considered the axioms of general set theory and shall in the next section examine the portion of set theory derivable from these axioms; this portion which Bernays terms 'general' set theory can be roughly characterised as 'constructive' since the axioms of ZF theory which are derivable from E1, 2 and A1 - 3 are ZF1, ZF2, ZF4, ZF6 and ZF8; these axioms of ZF are constructive in character in comparison with the more existential content of ZF3 (power set axiom), ZF5 (infinity), ZF7 (choice) and ZF9 (Fundierungsaxiom).

Section 5. General set theory.

Ordinals, natural numbers and recursion.

Ordinals

Cantor's way of defining ordinals (see Chapter I) was first to define order types (in modern terminology as equivalence classes with respect to similarity) and then to specialise the notion to well-ordered sets - the order types of which were ordinals. Any attempt to reproduce this definition within an axiomatic system encounters the difficulty that the sheer generality of the notion of order type (as the totality of sets similar to a given set) makes such a reproduction subject to ad hoc limitations if the antinomies are not to be introduced into the system. The character of these limitations will be discussed in the next section (pp. 71-3), where roughly the same situation occurs with the axiomatisation of Cantor's definition of cardinals.

In B. a completely different approach* is used, which is accomplished without recourse to any previously developed theory of order. The method consists of picking out the essential characteristics of an ordinal and then defining ordinals as sets possessing these characteristics. These characteristics are embodied in the following three predicates.

---

* Due to von Neumann (1923) and R.M. Robinson (1937).

---

(1) $\mathrm{Trans}(d) \leftrightarrow (x)(y)(x \in y \;\&\; y \in d \rightarrow x \in d)$

(2) $\mathrm{Alt}(d) \leftrightarrow (x)(y)(x \in d \;\&\; y \in d \;\&\; x \neq y \rightarrow x \in y \vee y \in x)$

(3) $\mathrm{Fund}(d) \leftrightarrow (x)(x \subseteq d \;\&\; x \neq \circ \rightarrow (Ey)(y \in x \;\&\; y \cap x = \circ)$

It is easily seen that (1) can be written as

$\mathrm{Trans}(d) \leftrightarrow (y)(y \in d \rightarrow y \subseteq d)$

and as such it expresses the property that any element of an ordinal is also a subset of that ordinal; (2) expresses the property that of any two ordinals, one is an element of the other (and hence by (1) an initial subset (section) of the other); (3) expresses that ordinals are 'well-founded' (wohlfundiert) in the sense that they admit no infinite regression of membership. We shall see that (3) will be in fact superfluous in view of A7 which imposes the same limitation on all sets of B. Ordinals are now defined as sets possessing all three of the properties (1) - (3) i.e.

$\mathrm{Od}(d) \leftrightarrow \mathrm{Trans}(d) \;\&\; \mathrm{Alt}(d) \;\&\; \mathrm{Fund}(d).$

Having set up this definition of ordinals, Bernays proves (pp. 80-86) that sets d satisfying it have all the required properties of ordinals. In particular that they are well-ordered with respect to the - relation, i.e. that every non-empty class of ordinals has a 'lowest' element. This theorem provides the existential content for the definition, by means of the ι-operator, of the <u>least ordinal</u> μA <u>belonging to a class A</u>, thus:

$\mu A = \iota_x(\mathrm{Od}(x) \;\&\; x \in A \;\&\; (z)(\mathrm{Od}(z) \;\&\; z \in A \rightarrow x = z \vee x \in z)).$

This can be extended to classes given in the form of a predicate with a free variable; thus

$$\mu_{\underline{x}}\underline{A}(\underline{x}) = \mu \{ \underline{x} \mid \underline{A}(\underline{x}) \}$$

From the above theorem on well-ordering (which he calls the <u>principle of the least ordinal</u>) Bernays derives the following <u>schema of transfinite induction</u>. 'If, for every ordinal $\underline{x}$, $\underline{A}(\underline{x})$ holds provided that it holds for every ordinal lower than $\underline{x}$, then $\underline{A}(\underline{x})$ holds for every ordinal' (p. 86). In symbols

$$(\underline{x})(Od(\underline{x}) \,\&\, (\underline{z})(\underline{z} \in \underline{x} \rightarrow \underline{A}(\underline{z})) \rightarrow \underline{A}(\underline{x})) \rightarrow (Od(c) \rightarrow \underline{A}(c)).$$

In the last section (p. 54) we stated that A2, A3 were generalisations of Cantor's two principles of formation; we can now indicate this more fully.

Corresponding to A1 - 3 there are three basic existential theorems:

(1) $Od(o)$

(2) $Od(c) \rightarrow Od(c;c)$

(3) $(\underline{x})(\underline{x} \in m \rightarrow Od(\underline{t}(\underline{x}))) \rightarrow Od(\sum_{\underline{x}} (m,\underline{t}(\underline{x})))$

From these theorems the following seven theorems (4)-(10) follow without much difficulty.

(4) $(x)(x \in m \rightarrow Od(x)) \rightarrow Od(\sum m)$

This enables the successor c' of an ordinal c to be defined as

$c' = c;c$

(5) $a \in a', \; a' \neq o, \; Od(a) \rightarrow Od(a')$

(6) $Od(a) \;\&\; Od(b) \;\&\; a' = b' \rightarrow a = b$

(7) $Od(c) \rightarrow o = c \vee o \in c$

(8) $Od(c) \;\&\; a \in c \rightarrow a' = c' \vee a' \in c.$

Theorems (5) and (8) characterise a' as the successor of a and we see that the operator "'" when applied to ordinals provides Cantor's first principle of formation.

(9) $(x)(x \in m \;\&\; Od(x)) \rightarrow (x)(x \in m \rightarrow x = \sum m \vee x \in \sum m)$

(10) $Od(c) \;\&\; (x)(x \in m \rightarrow Od(x) \;\&\; x \subseteq c) \rightarrow \sum m = c \vee \sum m \in c.$

Theorems (9), (10) assert that "For every set of ordinals m, $\sum m$ is as least as high an ordinal as any element of m, and it is the lowest one having that property". (p. 87). Thus if a set of ordinals m has no highest element then we see that $\sum m$ will be its limit number; so that "$\sum$" provides the limit operator of Cantor's second principle. The predicate 'c is a limit number' is defined as

$Lim(c) \leftrightarrow Od(c) \;\&\; c \neq o \;\&\; (x)(x \in c \rightarrow x' \in c).$

It is to be noted that the clarity provided by the present definition of ordinals becomes especially apparent in the proofs of the above theorem. For it is possible to construe a theorem about ordinals as the conjunction of three much simpler theorems corresponding to the conjunctive elements of the definition and establishing Od(d) thus amounts to establishing Trans(d), Alt(d) and Fund(d) separately.

From theorems (4) and (9) one infers immediately that the 'set of all ordinals' is contradictory since if m were such a set, $\sum m$ would be an ordinal higher than any element of m. Thus the class $\{x \mid Od(x)\}$ is not representable. (See section 6).

The further development of a theory of ordinals in B. progresses without difficulty constituting, for the most part, a rigorisation of the standard results and methods of the topic (e.g. Sierpinski 1957). There is however an important procedure of Bernays which requires comment - this is his method of dealing with recursive definitions (both finite and transfinite). It will be more convenient to describe this after we have set down Bernays' method for introducing the natural numbers.

Natural numbers

Number theory can be easily embedded in the above theory of ordinals by considering finite ordinals only. There is however no definition of finiteness available yet, and formally Bernays first defines natural numbers as certain special ordinals and then defines finiteness as equipollence with natural numbers. Thus he has the following definition of a natural number.

$$Nu(n) \leftrightarrow Od(n) \ \& \ (n=o \vee Suc(n)) \ \& \ (x)(x \in n \rightarrow x = o \vee Suc(x))$$

'a natural number is an ordinal such that itself and every element of it is either 0 or a successor'.

Peano's axioms for arithmetic (with the exception of the axiom of induction which is a special case of the principle of transfinite induction given above) follow from theorems (1), (5) and (6) on p. 62 above by replacing the predicate "Od" by the predicate "Nu".

Bernays points out that number theory does not need the full force of A1-3* but can be developed within the frame of A1-2 and the Aussonderungs theorem; he calls this frame the "weakened general set theory".

Finiteness can now be defined in the following manner.

$Fin(a) \leftrightarrow (Ex)(Nu(x) \& x \sim a)$

'The set a is finite if there exists a natural number equivalent to a.' From this, a class is defined to be finite if it is represented by a finite set, in symbols

$Fin(A) \leftrightarrow (Ex)(Fin(x) \& x^{*} \equiv A)$.

From the above definition of finiteness Bernays derives the usual properties of finite sets, e.g.

'The union of finitely many finite sets is again finite'. In particular he derives the theorem

$a \subset b \& (Fin(a) \vee Fin(b)) \rightarrow a \sim b$

which is Dedekind's definition of finiteness, since it asserts that a finite set cannot be equivalent to a proper subset of itself.

---

* This is to be expected in view of the purpose of A3 which was designed to extend rather than characterise Peano's axioms.

Iteration, recursion

Besides the Peano axioms one more concept is needed for a complete characterisation of arithmetic; this is the justification for introducing functions by primitive (finite) recursion. Bernays, provides this justification by establishing a schema - the schema of primitive recursion - which asserts that for every function defined in a primitive recursive manner there exists a function which can be defined explicitly. The proof of this schema is* an application of the iteration theorem which we now describe. First we need the definition of a sequence s

$Sq(s) \leftrightarrow Ft(s) \& Od(\Delta_1 s)$

Thus a sequence is a set of ordered pairs, the first members of which are ordinals (note, the elements of a sequence are the ordered pairs which constitute it, the members of a sequence are the elements of $_2s$). Next we have an iteration sequence It(s,a,F) for a function F starting from a:

$It(s,a,F) \leftrightarrow sq(s) \& Nu(\Delta_1 s) \& s \llcorner o = a \& (x)(x' \in \Delta_1 s \rightarrow \langle s \llcorner x, s \llcorner x' \rangle \in F)$

The interpretation of the right hand side of this is that s is a finite sequence starting from a, say

$\langle o,a \rangle \quad \langle 1,a_1 \rangle \quad ,\ldots\ldots \quad \langle n, a_n \rangle$

such that overlapping ordered pairs comprised of adjacent members of s belong to F i.e.

$\langle a,a_1 \rangle \in F, \quad \langle a_1,a_2 \rangle \in F, \ldots\ldots \langle a_{n-1} a_n \rangle \in F \qquad (1)$

---

* For an alternative proof which does not use the notion of iteration see Suppes 1960, pp. 142-4.

From this is defined the <u>numeral iterator J(F,a) of F on a</u>:

$$J(F,a) \equiv \{ xy \mid (Ez)\ It(z,a,F)\ \&\ \langle x,y \rangle \in z. \}$$

Intuitively $J(F,a)^{\llcorner} x$ iterates the function F applied to a just x times. This can be seen* by observing that since the pairs given in (1) are overlapping, the function value a_{r+1} for the argument a_r will be the argument producing the function value a_{r+2}, so that $a_1=F(a)$, $a_2=F(a_1)$ i.e. $a_2 = F(F(a))$ etc. and thus $J(F,a)^{\llcorner} x$ will be a_x.

The iteration theorem now states (p. 92) that "if A is a class of which a is a member and F a function mapping A into A then the iterator J(F,a) is a function H with domain $\{ x \mid Nu(x) \}$ which satisfies the following recursion equations:

$$H^{\llcorner} o = a,\ Nu(n) \rightarrow H^{\llcorner} n' = F^{\llcorner} (H^{\llcorner} n)."$$

The proof of the theorem follows almost from the definitions and consists of two parts. The first in showing that the hypothesis of the theorem, namely,

$$a \in A\ \&\ Ft(F)\ \&\ \Delta_1 F \equiv A\ \&\ \Delta_2 F \subseteq A$$

implies the existence of an iteration sequence for every n and the second in showing that the iteration sequence so guaranteed is unique; both parts follw by induction on n.

The iteration theorem provides the existential part of the schema quoted above since essentially J(F,a) is the required explicit function.

---

* Cf. the account in Keene 1961, pp. 82-3.

An immediate application of the theorem is the definition of the arithmetic functions of addition, multiplication and exponentiation, thus:

$$m + n = J\,(\{xy \mid y = x'\},\ m)^{\llcorner}\ n$$

$$m \cdot n = J\,(\{xy \mid y = x + m\},\ o)^{\llcorner}\ n$$

$$m^n = J\,(\{xy \mid y = x \cdot m\},\ o')^{\llcorner}\ n$$

We shall now pass on to the introduction of functions by transfinite recursion. Bernays' way of dealing with this does not differ essentially from the accepted approach (Cf. Bourbaki III p.42); he does however frame it as an extension of the method used for finite recursion and we shall just give an outline of the basic concepts involved.

The notion of a sequence is extended to a <u>sequential class</u> Sq (S) to cover the case where the domain of the sequence is the class of all ordinals:

$$Sq(S) \leftrightarrow Ft(S)\ \&\ ((Ex)(Od(x)\ \&\ \Delta_1 S \equiv x^*) \vee \Delta_1 S \equiv \{x \mid Od(x)\})$$

A sequential class whose members belong to a class C is called a <u>C-sequence</u>. A function F is said to <u>progress in C</u> if it assigns to every C-sequence an element of C, in symbols

$$Prog(F,C) \leftrightarrow Ft(F)\ \&\ \Delta_1 F \equiv \{x \mid Sq(x)\ \&\ \Delta_2 x^* \subseteq C\}\ \&\ \Delta_2 F \subseteq C$$

A <u>sequential class S</u> is called <u>adapted to F</u> if for each element n of its domain, $S^{\llcorner} n$ is the value of the function F when its argument is the n-segment of S:

$$Adp(S,F) \leftrightarrow Sq(S)\ \&\ (u)(u \in \Delta_1 S \rightarrow sg(S,u) \in \Delta_1 F\ \&\ S^{\llcorner} u = F^{\llcorner}\, sg(S,u))$$

where sg(S,u) denotes the segment of S determined by u, i.e. the unique subsequence of S whose domain is u.

From this comes the <u>adaptor of F</u>, AF, whose members are ordered pairs which are the elements of some sequence adapted to F:

$$AF \equiv \{xy \mid (Ez)\ Adp(z,F)\ \&\ \langle x,y\rangle \in z\}.$$

The general recursion theorem which corresponds to the iteration theorem now states: 'For any function which assigns to every sequence of elements of a class C again an element of C, we can define a function G, whose domain is $\{x \mid Od(x)\}$ and whose value for an ordinal k is that element of C, which is assigned by F to the k-segment of G.' (p. 102). In symbols

$$Prog(F,C) \rightarrow Sq(AF)\ \&\ \Delta_1 AF \equiv \{x \mid Od(x)\}\ \&\ \Delta_2 AF \subseteq C$$
$$\&\ (x)(Od(x) \rightarrow (AF)^{\llcorner} x = F^{\llcorner} sg(AF,x)).$$

The proof of this consists in showing that AF is such a function G.

A transfinite iterator I(G,a) can now be defined as the adaptor of that function which satisfies the normal recursive conditions. As before, an immediate application of the iterator is to define the (ordinal) functions of addition, multiplication and exponentiation. Thus:

$$a + b = I(\{xy \mid y = x'\},\ a)^{\llcorner}\ b$$

Similar definitions follow for multiplication and exponentiation.

The iterator provides the means of proving a schema of transfinite recursion (p. 107), since essentially I(G,a) can serve as the required explicit function.

Section 6. Cardinal numbers and the remaining axioms of B.

The comparison of powers of sets in B. is based (in the usual manner) on the notion of equivalence. The sets x, y are said to be of equal power if $x \sim y$ (p. 114). The set x is said to be of at most equal power to y if x is equivalent to a subset of y i.e.

$$x \preceq y \leftrightarrow (Ez)(z \subseteq y \ \& \ z \sim x).$$

x is of lower power than y if x is at most equal power to y but not of equal power i.e.

$$x \prec y \leftrightarrow x \preceq y \ \& \ \overline{x \sim y}.$$

The basic theorem concerning these notions is the equivalence theorem (Bernstein-Schröder)

$$x \preceq y \ \& \ y \preceq x \rightarrow x \sim y.$$

For a satisfactory construction of Cantor's theory of powers two more features are necessary. First one needs the comparability of powers, so that it can be asserted that of any two sets x,y, one of them must be of higher power, i.e. one needs the theorem

$$x \preceq y \vee y \preceq x \text{ for any sets } x, y.$$

As is well-known this theorem can be secured by the Choice-axiom used in the form of the well-ordering theorem. (Cf. Chapter I pp.15,16).

The second feature necessary for a theory of powers is Cantor's theorem that for every set x, the set of subsets of x has a higher power than x. Implicit, of course, in this theorem is that for every set there exists the set of its subsets. The closest one can get to this in B. (without postulating it) is that for every set x there exists the <u>class</u> of its subsets $\{z \mid z \subseteq a\}$. The representability of this class is the content of the Potenzmengeraxiom which takes the following form: (p. 130).

A4. $c \in \pi(x) \leftrightarrow c \subseteq x$

'The elements of $\pi(x)$ are the subsets of x.'

(A4, like A1-3, implicitly defines the set-symbol occurring in it).

By the Church schema, from A4 we have

$$\pi(x)^{*} \equiv \{z \mid z \subseteq x\}$$

so that A4 does indeed assert the representability of $\{z \mid z \subseteq x\}$.

Without this axiom, however, Bernays proves the following result.

There does not exist a function mapping x^{*} onto $\{z \mid z \subseteq x\}$ and hence there is no (1,1) correspondence between $\{z \mid z \subseteq x\}$ and x^{*} or a subset of x^{*}, in symbols:

$$Ft(F) \,\&\, \Delta_1 F \equiv a^{*} \,\&\, \Delta_2 F \equiv C \rightarrow (Ez)(z \subseteq a \,\&\, z \notin C).$$

The corresponding statement that there does not exist a (1,1) correspondence between a <u>class</u> A and the class of its subsets does not hold, as is seen from the universal class V of which every set is both an element and a subset and thus 'it appears that Cantor's paradox connected with the set of all sets is removed in our system by the distinction between sets and classes'. (p. 118).

Immediately connected with the comparison of powers is the notion of cardinal numbers. In fact the cardinal number of a set x can be considered to be that which is common to all sets having the same power as x. The difficulty, however, is in embedding such a conception in an axiomatic system. Let us consider two of the main attempts to do this.

(1) The cardinal of a set x is the set of all sets equivalent to x.

(2) Cardinal of a set x = cardinal of a set y if and only if $x \sim y$.

Definition (1) was Frege's attempt to make rigorous Cantor's notion of cardinal number (Chapter I p. 3). It is unacceptable here since forming the <u>set</u> of sets equivalent to x is not a permissible set operation; to make it so for any set x, would require an unrestricted axiom of comprehension which leads back to the antimonies. In our terminology here the class $\{ x \mid x \sim a \}$ is not representable (that this is so will be shown near the end of this section p. 78).

Definition (2) is an example of a 'working definition'. Any questions involving cardinals are changed by it into ones involving equivalence. It suffers however from two defects:

(a) It transgresses the first requisite of any definition, i.e. that the definiens should be eliminable (of course, if an <u>alternative</u> definition of cardinal number is used in a system then it is essential that (a) should be derivable - in this sense the definiens would be eliminable. (See Suppes 1960 p. 242 and also below).

(b) It does not define a cardinal number as an entity nor is there any indication that by it cardinal numbers are sets, or classes for that matter, since the definition gives no information about the nature of a cardinal number. (It is interesting to note in this context that the somewhat ideal character of cardinal numbers suggests that it might be feasible to have classes as cardinals, i.e. defining the cardinal number of x as $\{z \mid z \sim x\}$ - this certainly satisfies (2) and effects a complete return to Frege's notion of cardinal number. For this suggestion to be workable it would be necessary to ensure that none of the operations normally performed on cardinals rely on the fact that cardinals, as sets, can be members).

Bernays* takes a course completely different from either (1) or (2) and defines cardinals as certain ordinals (p. 139) specifically he defines the <u>cardinal</u> <u>of</u> <u>a</u> <u>set</u> <u>x</u> to be the <u>least</u> <u>ordinal</u> <u>in</u> <u>the</u> <u>class</u> <u>of</u> <u>sets</u> <u>equivalent</u> <u>to</u> <u>x</u>, thus:

$$\aleph(x) = \mu_z \, (z \sim x)$$

For every set x to possess a cardinal, this definition requires that the class $\{z \mid z \sim x\}$ contain at least one ordinal; this is guaranteed by the well-ordering theorem (which Bernays terms the <u>numeration</u> <u>theorem</u> p. 138). In view of our remarks concerning Bourbaki's definition of a cardinal in the next chapter, we state explicitly that Bernays' definition thus depends on the Choice-axiom (through the well-ordering theorem) to ensure that every set has a cardinal number.

From his definition Bernays[*] derives definition (2) as a theorem, i.e.

$\aleph(x) = \aleph(y) \leftrightarrow x \sim y$

Thus the usual development of cardinal arithmetic, proceeding as it does through equivalence, can be obtained.

The remaining axioms of B.

Bernays postulates the following form of the Choice-axiom

A5. $Ps(a) \rightarrow (Ey)(y \subseteq a \ \& \ \Delta_1 y = \Delta_1 a \ \& \ Ft(y))$

'For every set of pairs a there exists a subset which is a function with the same domain as a.' (p. 137).

The form of the Choice-axiom embodied in A5 is not in any way special and Bernays proves that A5 is formally equivalent to any of the usual formulations of the axiom (e.g. the multiplicative form); in particular A5 is equivalent to the following:

A5'. $(x)(x \in m \rightarrow x \neq o) \rightarrow (Ey)(Ft(y) \ \& \ \Delta_1 y = m \ \& \ (x)(x \in m \rightarrow y \text{`} x \in x))$.

'If m is a set of non-empty sets, there exists a function assigning to each element of m one of its elements'.

[*] The same course is followed in Suppes 1960

We shall return to the Choice-axiom (in the form A5') later in this section.

Bernays' <u>axiom of infinity</u> is formally equivalent to asserting the representability of the class of natural numbers. It takes the form of implicitly defining a set symbol for the set of natural numbers. Thus:

A6 $\quad a \in \omega \leftrightarrow Nu(a)$

which is equivalent to $Rep\{x \mid Nu(x)\}$, so that

$\omega^* \equiv \{x \mid Nu(x)\}$.

The predicate of being infinite is defined in the natural way, so that a set is infinite if it is not finite, in symbols

$Infin(a) \leftrightarrow \overline{Fin(a)}$

On account of this formula the connections between the numerous axioms of infinity which have been put forward are similar to the connections between the various definitions of finiteness which have been thoroughly investigated (Tarski 1924). In fact Bernays proves that A6 is equivalent to the following axioms of infinity.

Zermelo $\quad (Ex)(o \in x \;\&\; (y)(y \in x \rightarrow [y] \in x))$

Dedekind $\quad (Ex)(Ey)(y \subset x \;\&\; x \sim y)$

von Neumann $\quad (Ex)(x \neq o \;\&\; (y)(y \in x \rightarrow (Ez)(z \in x \;\&\; y \subset z)))$.

We now turn to two important strengthenings of the Choice-axiom which can be accomplished by using the class formalism. It is more convenient to consider A5' rather than A5. The first strengthening appears equivalent, in its effect, to a weakening of the hypothesis of A5' obtained by replacing the set m occurring there by V. This is however our interpretation and formally Bernays introduces a function symbol $\sigma(a, b)$ which makes $Ft(y)$ and $\Delta_1 y = m$ redundant so that we get from A5'

$$(x)(x \in m \rightarrow x \neq o) \rightarrow (c \in m \rightarrow \sigma(m,c) \in c)$$

$[c]$ is now substituted for m (making $c \in m$ redundant) to give

$$(x)(x \in [c] \rightarrow x \neq o) \rightarrow \sigma([c], c) \in c$$

defining $\sigma(c) = \sigma([c], c)$ this becomes

$$A_\sigma \quad c \neq o \rightarrow \sigma(c) \in c.$$

The interpretation of A_σ is that to every non-empty set c, a member of that set $\sigma(c)$ can be assigned; that this is stronger than A5' can be seen intuitively from the fact that we no longer require our sets, to which we assign elements, to be members of some other set m.

The motivation for constructing A_σ is to bring the Choice-axiom in line with the other axioms (A1 - 4,6) which all have a primitive symbol in them. However, whereas the other axioms contained an extensional definition of their symbols, this is not the case with the symbol $\sigma(c)$ and the nearest one can get to it is to derive the following theorem:

$$a = b \rightarrow \sigma(a) = \sigma(b) \qquad (1)$$

The second strengthening of A5' comes naturally from A_σ by allowing* c to be a class variable. This gives

$A_\sigma{}'$ $\quad a \in C \rightarrow \sigma(C) \in C$

(note that $A_\sigma{}'$ yields A_σ by defining $\sigma(a^*) = \sigma(a)$).
The interpretation of $A_\sigma{}'$ is that to every non-empty class C is assigned a member $\sigma(C)$ of that class. The analogue of (1) above is, however, not derivable and must be postulated, i.e.

$A_\sigma{}''$ $\quad A \equiv B \rightarrow \sigma(A) = \sigma(B)$

The question as to which is stronger A_σ or $A_\sigma{}'$, $A_\sigma{}''$ will be answered below, for the answer depends on the construction of a (1,1) mapping between the class of all sets, V, and the class of all ordinals $\{x \mid Od(x)\}$. This mapping is achieved by making essential use of A_σ and the Fundierungsaxiom which we now give

A7. $\quad c \neq o \rightarrow (Ey)(y \in c \;\&\; (x)(x \notin y \vee x \notin c))$

'For every non-empty set c there is an element y of c which has no element in common with c.' (p. 201). This is equivalent to postulating Fund(c) for all sets c in B. (see p. 60 above). From A7 using A_σ we get the following stronger form of the Fundierungsaxiom

---

* We are not suggesting that the passage from A_σ to $A_\sigma{}'$ is obtained by merely substituting a class variable for a set variable - there can be no such rule of substitution in B. - only that A_σ and $A_\sigma{}'$ are related intuitively in this manner. In fact $A_\sigma{}'$ must be postulated and is motivated by Hilbert's 'ϵ-formula'.

$F_\sigma \quad \sigma(c) \cap c = o$

where $\sigma(c)$ is the required element y of c. It is in the form F_σ that the Fundierungsaxiom is used with A_σ to construct the mapping between V and $\{x \quad Od(x)\}$; Bernays designates the mapping class by Θ . Using this mapping class, it is possible to give an explicit definition of the symbol $\sigma(A)$ occurring in A_σ', thus:

$\sigma(A) = \Theta^{\iota}\mu_x(Ey)(y \in A \& x = \breve{\Theta}^{\iota}y)$

We see that $\sigma(A)$ is the set corresponding (under Θ) to the least ordinal of the class of ordinals which represents A under Θ .

From this definition the axioms A ', A " are easily derivable (as can be seen from their interpretations as statements about ordinals) so that A_σ and A_σ', A_σ'' are therefore equivalent.

The major importance of the construction of Θ lies however in a different direction from the derivation of results concerning Choice-axioms. For, using Θ, it is possible to show that von Neumann's criterion for 'over-large' classes is present in B. We recall that for von Neumann a class was 'over-large' and therefore non-representable if and only if it was mappable onto V. To show that this is true for the classes of B. we must use the following theorem (proved within the theory of ordinals p. 129).

'For any class C of ordinals either C is represented by a set c.... or else for every ordinal k C there is a higher one in C and then there is a (1,1) correspondence between $\{x \mid Od(x)\}$ and C.....'

Hence by Θ any class K corresponds to some class of ordinals C. If C is representable then so is K and K cannot be mapped onto V. If C is not representable then neither is K and C can be mapped onto $\{x \quad Od(x)\}$ i.e. K can be mapped onto V.

Precisely why von Neumann chose the above criterion for representability is not clear. It is certainly a necessary condition for the exclusion of the antinomies; proving its sufficiency, however, would amount to a proof of consistency for set theory. That it is necessary follows from the fact that if there existed a set c in B. such that $c \sim V$ then the formation of $\pi(c)$, the power set of c, would reintroduce Cantor's paradox (p.26 above). This is not a very strong justification for the criterion, however, since the same effect could be achieved by inserting a clause in A4 (power set axiom) disallowing the formation of $\pi(x)$ for $x \sim V$.

We can now prove the assertion that $\{x \mid x \sim a\} \sim V$ for any set a, as was stated on p.71 above. Since a is a set we have, by the mapping Θ, $\overline{a \sim V}$ and hence $\overline{a} \sim V$ where $\overline{a}$ is the complement of a with respect to V.

Now let c be any member of a, and consider the set

$$y = a^{-}[c]; x$$

(i.e. the set formed from a by replacing c by x).

It follows immediately that $y \sim a$.

Now if $x \in \bar{a}$ then

$$\{y \mid y = a^{-}[c]; x \ \& \ x \in \bar{a}\} \sim V$$

by virtue of $\bar{a} \sim V$

and thus $\{x \mid x \sim a\} \sim V$ since

$$\{y \mid y = a^{-}[c]; x \ \& \ x \in \bar{a}\} \subseteq \{x \quad x \sim a\} .$$

We turn now to Bernays' sketch of the possibility of constructing mathematics within B. This is divided into two main sections, the development of classical analysis and the associated structures of finite and infinite vector spaces etc. and the development of the arithmetic of cardinal and ordinal numbers. Both sections amount to formalisations of procedures, accepted by now as standard; we shall ignore the last section (cf. p. 63 above) noting only that it has been secured in weaker systems, ZF, than B. (See Suppes 1960).

The fundament of classical analysis is the definition of a real number. Bernays' achieves this after the method of Dedekind by defining real numbers as sets of rationals. Formally instead of rationals, we have fraction triplets $\langle\langle a,b\rangle, c\rangle$ where a, b, c are natural numbers with c different from 0. The formal definition of fraction triplets is as follows:

$Ftp(d) \leftrightarrow (Ex)(Ey)(Ez)(Nu(x) \ \& \ Nu(y) \ \& \ Nu(z) \ \& \ z \neq 0 \ \& \ d = \langle\langle x,y\rangle, z\rangle)$

Intuitively $\langle\langle a,b\rangle ,c\rangle$ will be interpreted as $\frac{a-b}{c}$. The operations of sum, sifference, product are then defined for these triplets, e.g.

$$p +^{r} q = \iota_t(\mathrm{Ftp}(p) \,\&\, \mathrm{Ftp}(q) \,\&\, (x)(y)(z)(u)(v)(w)(p = \langle\langle x,y\rangle, z\rangle$$
$$\& \; q = \langle\langle u,v\rangle , w\rangle \rightarrow t = \langle\langle (w.x) + (z.u), (w.y) + (z.v)\rangle , z.w\rangle))$$

which formalises

$$\frac{x - y}{z} + \frac{u - v}{w} = \frac{wx + zu - (wy + zv)}{zw}$$

similar definitions follow for $p -^{r} q$, $p .^{r} q$.

The triplet $\langle\langle a,b\rangle ,c\rangle$ is positive, negative or a null triplet according as $b \in a$, $a \in b$ or $a = b$. From this the predicates of equality ($p =^{r} q$), greater than ($p >^{r} q$) and less than ($p <^{r} q$) are defined according as $p -^{r} q$ is a null triplet a postive or negative triplet.

It is seen of course that signed fraction come easily from triplets by having either $a = o$ or $b = o$.

A real number is now defined as a special set of fraction triplets, specifically as a non-void initial section without greatest element in the ordered set of fraction triplets, in symbols

$$\mathrm{Re}(c) \leftrightarrow c^{*} \subset \{x \;\; \mathrm{Ftp}(x)\} \,\&\, c \neq o \,\&\, (x)(x \in c \rightarrow (Ey)(y \in c \,\&\, x <^{r} y))$$
$$\& \; (x)(y)(y \in c \,\&\, (x =^{r} y \vee x <^{r} y) \rightarrow x \in c)$$

(That real numbers are <u>sets</u> can be seen from the fact that $\{x \mid \mathrm{Ftp}(x)\}$ is a subclass of the <u>set</u> $(\omega \times \omega) \times \omega$. The computation laws for real numbers follow the usual Dedekind pattern; equality between real numbers is just set-theoretic equality and the operations of addition and multiplication (and their inverses) are defined in terms of these operations on fraction triplets which are the members, e.g. the arithmetic sum p $\#$ q of two real numbers p,q is defined as the set of triplets which are the sum of an element of p and an element of q, thus

$$p \# q = \iota_t(Re(p) \;\&\; Re(q) \;\&\; t^* \equiv \{ z \mid (Ex)(Ey)(x \in p \;\&\; y \in q \;\&\; z = x +^r y)\})$$

A real number is positive if it has some positive fraction triplet as an element; it is negative if there is some negative triplet which is not an element of it. The real number null is thus the set of all negative triplets. A real number p is called rational if there exists a triplet t such that p represents $\{x \mid x <^r t.\}$

Bernays proves the property of continuity for the real numbers by deriving the theorem of the least upper bound:

$$A \neq \wedge \;\&\; (x)(x \in A \rightarrow Re(x)) \;\&\; (Ey)(Re(y) \;\&\; (x)(x \in A \rightarrow x \subseteq y)$$
$$\rightarrow (Ez)(Re(z) \;\&\; (u)(z \subseteq u \leftrightarrow (x)(x \in A \rightarrow x \subseteq u)).$$

'For every non-void class of real numbers A which has an upper bound, there exists a real number which is the least upper bound'. (p. 161). Bernays proves this by observing that the sum of the elements of A, being a class of triplets of the required form, is a real number which moreover has the property of being the least upper bound.

Bernays does not go beyond the above theorem in his sketch* of the possibility of constructing classical analysis in B.; indeed there is little need to go beyond the above theorem if one's main interest is to secure the <u>possibility</u> of constructing analysis rather than actually constructing it - this point will be developed in the next chapter.

---

* For a fuller and more comprehensive construction of the real numbers within an axiomatic system of set theory which is similar in many ways to B. see Suppes 1960 (pp. 159-194).

CHAPTER III.

The set theory of N. Bourbaki.

This chapter is an exposition of the set theory put forward by Bourbaki in*

"Theorie des Ensembles:

Chapitre 1. Description de la Mathématique Formelle.

2. Théorie des Ensembles.

3. Ensembles Ordonnés, Cardinaux - Nombres Entiers".

which constitute the first three chapters of Book 1 of "Les Structures Fondamentales de l'Analyse".

Reference will be made to chapter and page numbers thus: II. 86 for page 86 of chapter II. To avoid confusion we point out that the pages of chapter III are numbered from 1 again. Lastly Bourbaki's system as a whole will be referred to as Bou.

Bou. is a variant of the ZF set theory outlined in chapter I, but departing essentially from this in its use of a selection operator.

---

* The system described here is different from (and has now superseded) the provisional set of axioms published by Bourbaki in 'Foundations of Mathematics for the working mathematician'. J.S.L. 14, pp. 1-8, 1949.

Section 1 - Basic Logic

The logic upon which the specific axioms of set theory in Bou. are superimposed is a first order functional calculus whose primitive symbols (signes logiques) are $\square$, τ , $\vee$, $\neg$ and a denumberable alphabet of letters x,y,z,...; X,Y,Z.

The primitive symbols of Bourbaki's set theory are = , $\in$, $\supset$.

Bourbaki formulates his logic in terms of assemblages. An assemblage is a finite string of the seven basic symbols and also letters.

For the sake of clarity we shall give the intuitive interpretation of the basic symbols before proceeding with our description of the logic.

$\vee$, $\neg$ are the usual signs denoting disjunction and negation respectively; from these the sign of implication $\Rightarrow$ is defined as an abbreviation for the assemblage $\vee\neg$.

τ is a selection operator and is identical to Hilbert's ϵ - operator. As we shall see, it is the only means of binding variables in Bou.

$\square$ stands in place of a bound variable and is joined to the appropriate binding operator with a bar. Thus if we wish to write down the assemblage which is designated by $\tau_x(A)$ we first join all occurrences of x in A to τ by a bar then replace all these occurrences by $\square$.

E.g. The term $\tau_x(x \neq y)$ is first written as

$\tau \neg = xy$ then $\overline{\tau \neg = x}y$ and finally as

$\overline{\tau \neg = \Box}\, y$. We note that the assemblage designated by

$\tau_x(A)$ does not contain x.

The intuitive interpretations of the signes relationnels $\in$, $=$ are the usual ones of membership and equality respectively; the signe substantifique* $\supset$ denotes an ordered pair.

The symbols $\vee$, $\in$, $=$, $\supset$ all have a scope of two letters.

In an assemblage the letters to which a primitive symbol applies are written to the right of that symbol. Thus the following are assemblages:

(i) $\supset xy$

(ii) $\in xy$

(iii) $\vee \in xy = xx$

(iv) $\overline{\tau \in \Box}\, y$

(v) $\overline{\tau \vee \neg \in \Box}\, y \in \Box\, z$

(vi) $\neg \vee \in xy = xy$

Such a method of writing down expressions of the system would, however, lead to unmanageable typographical difficulties so (i) - (vi)

---

* By signe substantifique Bourbaki means a term-generating sign; the term generated here being an ordered pair.

would be 'abbreviated' as follows:

(i)' (x,y)

(ii)' x ∈ y

(iii)' $x \in y \vee x = x$

(iv)' $\tau_x(x \in y)$

(v)' $\tau_x(x \notin y \vee x \in z)$

(vi)' $\neg(x \in y \vee x = y)$

These abbreviations are effected by applying simple metamathematical rules (I. 12, 13, 30). Bourbaki also uses metavariables which can be classified as follows:

(1) As metavariables denoting set variables of the system we have **x, y, ...; X, Y, Z.**

There is no syntactical difference between the use of **x** and **X** etc.

(2) As metavariables denoting relations* we have

(a) **A, B, C, R, S**

and as metavariables denoting terms we have

(b) **T, U.**

Thus all metavariables are in heavy print as opposed to the ordinary set variables of the system which are x, y, z,...; X, Y, Z (where again there is no syntactical difference between the use of x and X etc.).

---

*The precise definitions of relations and terms are given below.

Although Bourbaki never confuses variables of the system with metavariables it must be added that the distinctions between cases 2(a) and 2(b) and even between cases (2) and (1) are not always adhered to; for typographical reasons we shall use a bar beneath a variable to indicate that it is a metavariable.

We shall now complete our account of the intuitive content of the primitive symbols by giving Bourbaki's account of the τ-operator (I. 16).

"If $\underline{B}$ is an assertion, $\underline{x}$ a letter then $\tau_{\underline{x}}(\underline{B})$ is an object; if we consider the assertion $\underline{B}$ as expressing a property of the object $\underline{x}$, then if there exists an object possessing the property in question $\tau_{\underline{x}}(\underline{B})$ represents one such object (un objet privilégié). If there does not exist any such object then $\tau_{\underline{x}}(\underline{B})$ represents an object about which nothing can be said (dont on ne peut rien dire)."

To delineate the 'well-formed' assemblages Bourbaki sets up metamathematical rules under the heading constructions formatives (I. 15); any assemblage which satisfies these rules is 'well-formed'. At the same time he divides the class of all assemblages into two species:

Species 1. Assemblages beginning with τ, with a signe substantifique or reducing to a single letter.

Species 2. All other assemblages.

Assemblages satisfying the constructions formatives which belong to species 1 are called terms and those which belong to species 2 are called relations. Having thus defined terms and relations with respect to assemblages, Bourbaki proves several Critères formatifs (I. 17) which recursively define the notions of term and relation with respect to their abbreviated forms rather than assemblages. The outcome of the constructions formatives and the critères formatifs is that Bourbaki's notion of a relation coincides* with the usual notion of a well-formed formula built up from atomic $\in$ - statements.

It should be noted that Bourbaki frames the constructions formatives and hence the critères formatifs in an open manner; that is to say he retains the possibility of enlarging his notion of relation and of term by the introduction of new (mathematical) primitive symbols. Thus critère formatif 4 asserts:

CF. 4. If $\underline{A}_1, \underline{A}_2, \ldots \underline{A}_n$ are terms and $\underline{s}$ is a signe relationel (respectively substantifique) of scope n then $s\, \underline{A}_1\, \underline{A}_2 \ldots \underline{A}_n$ is a relation (term). (Where, of course $\in$, = are signes relationels and $\supset$ a signe substantifique of scope 2).

---

* With the proviso that his inclusion of τ - terms somewhat widens this usual notion of formula.

We can now indicate fully Bourbaki's logic. It is embodied in the schemata S1. - S7. which we now state;

S1. If $\underline{A}$ is a relation then the relation $\underline{A} \vee \underline{A} \Rightarrow \underline{A}$ is an axiom.

S2. If $\underline{A}$, $\underline{B}$ are relations then the relation $\underline{A} \Rightarrow (\underline{A} \vee \underline{B})$ is an axiom.

S3. If $\underline{A}$, $\underline{B}$ are relations then the relation $\underline{A} \vee \underline{B} \Rightarrow \underline{B} \vee \underline{A}$ is as axiom.

S4. If $\underline{A}$, $\underline{B}$ are relations then the relation
$\underline{A} \Rightarrow \underline{B} \Rightarrow ((\underline{C} \vee \underline{A}) \Rightarrow (\underline{C} \vee \underline{B}))$ is an axiom.

S5. If $\underline{R}$ is a relation, $\underline{T}$ a term, $\underline{x}$ a letter then the relation
$(\underline{T}|\underline{x})\underline{R} \Rightarrow (\exists \underline{x})\underline{R}$ is an axiom.
Where $(\underline{T}|\underline{x})\underline{R}$ denotes the result of replacing all occurrences of $\underline{x}$ in $\underline{R}$ by $\underline{T}$.

S6. If $\underline{x}$ is a letter, T, U are terms and $R\{x\}$ a relation then
$\underline{T} = \underline{U} \Rightarrow (\underline{R}\{\underline{T}\} \Leftrightarrow \underline{R}\{\underline{U}\})$ is an axiom.

S7. If $\underline{R}$, $\underline{S}$ are relations and $\underline{x}$ a letter then
$(\forall \underline{x})\ (\underline{R} \Leftrightarrow \underline{S})\ ) \Rightarrow (\ \tau_{\underline{x}}(\underline{R}) = \tau_{\underline{x}}(\underline{S})\ )$ is an axiom.

The quantifiers occurring in S5., S7. are not primitive but are defined by means of the τ- operator as follows:

$(\ \exists\ \underline{x})(\underline{R})$ is an abbreviation for $(\ \tau_{\underline{x}}(\underline{R}) | \underline{x})\underline{R}$ and

$(\ \forall\ \underline{x})(\underline{R})$ is an abbreviation for $\neg (\ \exists\ \underline{x}) \neg (\underline{R})$ i.e. for
$\neg (\ (\tau_{\underline{x}}(\underline{R}) | \underline{x}) \neg \underline{R})$.

Now S1. - S4. constitute a standard* formalisation of the propositional calculus; we shall therefore just indicate how S5. - S7. yield the remaining axioms of the predicate calculus.

The predicate calculus** is obtainable in Bou. if the formulae (a) and (b) given below are derivable and if the rule of Modus Ponens and suitable rules of substitution are available.

(a) $(\forall \underline{x})\ \underline{R}(\underline{x}) \Rightarrow \underline{R}(\underline{y})$

(b) $\underline{R}\ (\underline{y}) \Rightarrow (\exists\, \underline{x})\ \underline{R}(\underline{x})$

Now (a) is derivable from S5. (I. 38) and (b) is identical to S5. since R(T) is an abbreviation for $(T \mid x)R$.

Bourbaki postulates Modus Ponens as a rule of proof (I. 21-2) and sets up suitable rules for substitution on pages I. 18-20. Furthermore S6. completely characterises the relation of equality.

Thus S1. - S7. certainly contain a standard formalisation of the predicate calculus. We shall return to S5. and S7. when we discuss the form of the Choice-axiom derivable in Bou.

Hilbert and Ackermann (1938)

* p. 27.

** p. 86.

Section 2 - Set-theoretic terminology

A relation R is collectivisante in $\underline{x}$ (II. 63) when the formula $(\exists \underline{y})(\forall \underline{x})(\underline{x} \in \underline{y} \Leftrightarrow \underline{R})$ is provable. Bourbaki abbreviates this formula to $\mathrm{Coll}_{\underline{x}}\, \underline{R}$.

Intuitively the provability of $\mathrm{Coll}_{\underline{x}}\, \underline{R}$ means that there exists a set $\underline{y}$ such that all the objects satisfying $\underline{R}$ are precisely the members of $\underline{y}$. Thus, here, as in Bernays' system, we have the recognition (formulated within the system) of the idea that certain classes (here predicates) are not capable of yielding sets, if consistency is to be preserved.

Once we have proved $\mathrm{Coll}_{\underline{x}}\underline{R}$ it is permissible to define a symbol denoting the extension of R. This is $\mathcal{E}_{\underline{x}}(\underline{R})$ which is an abbreviation for $\tau_{\underline{y}}((\forall \underline{x})\ ((\underline{x} \in \underline{y}) \Leftrightarrow R)$; as a provable formula we have (II. 64):

$$\mathrm{Coll}_{\underline{x}}\, \underline{R} \Rightarrow (\forall x)(x \in \mathcal{E}_{\underline{x}}(\underline{R}) \Leftrightarrow \underline{R})$$

We note that, like the assemblage designated by $\tau_{\underline{x}}(\underline{R})$, the assemblages designated by $\mathrm{Coll}_{\underline{x}}(\underline{R})$ and $\mathcal{E}_{\underline{x}}(\underline{R})$ do not contain x.

There is clearly a connection between relations collectivisantes and the representable classes in Bernays. We shall exhibit such a connection between these two concepts after we have discussed the τ - operator in greater detail.

The relations $x \notin x$, $(\forall x)(x \in X)$ are easily proved (II. 67) to be non-collectivisantes in x, X respectively.

A relation R is said to be functional ("fonctionel") in x if there exists one and only one x such that R(x)*.

The relation $(\forall x)(x \notin X)$ is functional in X (II. 67) and from this $\emptyset$ (the null set) is defined as $\tau_x((\forall x)(x \notin X)$. Written out as an assemblage this is $\tau \neg\neg\neg \in \tau \neg\neg \in \Box \Box \Box$

A set G is a graph ("graphe"), if all its members are couples (ordered pairs). The name pair-set is used by Bernays.

If the relation $(\exists G)(G$ is a graph and $(\forall x)(\forall y)(R \Leftrightarrow ((x,y) \in G))$ is provable, then R is said to admit a graph with respect to x,y.

A correspondence between sets A, B is a triplet $\Gamma = (G,A,B)$ where G is a graph such that $pr_1 G \subset A$, $pr_2 G \subset B$. (see below). G is termed the graph of Γ, A the departure set and B the arrival set.

$pr_1 z$, $pr_2 z$ are the première et deuxième projections de z i.e. terms denoting the first and second members of a couple or the set of first and second members of a set of couples.

---

* When a relation R is functional in x, $\tau_x(R)$ is a unique object - the x such that R. In this case the τ- operator becomes identical to the ι- operator and thus yields the definite article.

Formally $pr_1\, z$ designates $\tau_x((\exists y)\ (z = (x,y))$

and $pr_2\, z$ designates $\tau_y((\exists x)\ (z = (x,y))$.

A graph F is called functional graph if for any x there exists at most one object y, corresponding to x by F. (i.e. there is at most one couple of the form (x,y) in F).

A correspondence f = (F,A,B) is a function if its graph F is functional and if its departure set A is pr_1 F i.e. f is a function if, for all $x \in A$, the relation $(x,y) \in F$ is functional in y (see p.92 above); the unique object y corresponding to a given x by F is termed the value of f (for x in A) and is written as f(x) or f_x or F(x) or F_x .

The functions f and g are said to coincide in a set E if E is contained in the departure sets (sometimes called sets of definition) of f and of g and if f(x) = g(x) for all $x \in E$.

If for two functions f = (F,A,B), g = (G,C,D) we have $B \subset D$ then we say that g is a prolongation ("prolongement") of f to C or simply that g prolongs f to C.

In a similar way the reverse of this notion, that of restriction is defined so that a function prolongs each one of its restrictions.

Finally the much used notions of one to one, onto and one to one onto are termed injection, surjection and bijection respectively.

Section 3 - Derivation of the ZF axioms

We shall in this section derive the axioms ZF1 - 8 in Bou; first we must state the one schema and five axioms of Bourbaki's set theory.

S8. $\underline{R}$ is a relation, $\underline{x}$, $\underline{y}$ letters, $\underline{X}$, $\underline{Y}$ letters distinct from $\underline{x}$, $\underline{y}$ and not figuring in $\underline{R}$ then the following relation is an axiom (- schema).

$$(\forall \underline{y})(\exists \underline{X})(\forall \underline{x})(\underline{R} \Rightarrow (\underline{x} \in \underline{X})) \Rightarrow (\forall \underline{Y}) \text{ Coll}_{\underline{x}} ((\exists \underline{y})((\underline{y} \in \underline{Y} \ \& \ \underline{R})$$

this is known as Le schéma de sélection et réunion.

A1. $(\forall x)(\forall y)(x \subset y \ \& \ y \subset x) \Rightarrow (x=y)$

Axiom of extensionality

A2. $(\forall x)(\forall y) \text{ Coll}_z \ (z=x \vee z=y)$

Axiom of pairs (L'axiome de l'ensemble a deux éléments)

A3. $(\forall x)(\forall x')(\forall y)(\forall y')(((x,y) = (x',y')) \Rightarrow (x=x' \ \& \ y=y')$

Axiome du couple

A4. $(\forall X) \text{ Coll}_y \ (Y \subset X)$

Power set axiom (L'axiome de l'ensemble des parties)

A5. "Il existe un ensemble in fini"

The nature of this axiom will be discussed in this section.

The derivation of ZF1 - 8 proceeds as follows:

ZF1. If we replace the subset relation by its definition in terms of $\in$, then A1. reads:

$$(\forall x)(\forall y)(\forall z)(((z \in x \Rightarrow z \in y) \& (z \in y \Rightarrow z \in x)) \Rightarrow x=y)$$

i.e. $(\forall x)(\forall y)(\forall z)((z \in x \Leftrightarrow z \in y) \Rightarrow x = y)$

which is ZF1.

ZF2. Replacing $\text{Coll}_z R$ by its definition in A2. we get

$$(\exists w)(\forall z)(z \in w \Leftrightarrow z = x \vee z = y)$$

ZF3. Replacing $\text{Coll}_Y (Y \subset X)$ by its definition is A4. we get

$$(\forall X)(\exists y)(\forall Y)(Y \in y \Leftrightarrow Y \subset X)$$

which is ZF3.

ZF4. Bourbaki derives a theorem equivalent to this axiom, namely:

$\text{Coll}_x (\exists i)(i \in I \& x \in X_i)$ (II. 88)

this can be rewritten as

$$(\exists y)(\forall x)(x \in y \Leftrightarrow (\exists i)(i \in I \& x \in X_i) \tag{1}$$

we recall that ZF4. asserted

$$(\exists y)(\forall x)(x \in y \Leftrightarrow (\exists w)(w \in z \& x \in w) \tag{2}$$

The equivalence of (1) and (2) can be seen by noting that the notion of a family of sets $\{X_\iota\}_{\iota \in I}$ is based on the identity transformation (<u>application identique</u> II. 77-8) $i \rightarrow X_i$ so that $(\forall x)(\exists i)(i \in I \ \& \ x \in X_i) \Leftrightarrow (\forall x)(\exists i)(i \in I \ \& \ x \in i)$.

ZF5. Corresponding to this axiom is A5. (III. 90) which is

<u>There exists an infinite set</u>.

Whether this axiom is stronger than ZF5. or not depends on Bourbaki's definition of finiteness, which we have not yet discussed. For the sake of completeness, however, we shall state the relevant result which is that Coll $_x$ (x is an integer) is formally equivalent to the proposition that there exists an infinite (i.e. non-finite) set. Thus both ZF5. and A5. assert the existence of a set which (intuitively) has cardinality $\aleph_0$

ZF6. A theorem(C. 51. II. 65) equivalent to this axiom is an immediate consequence of S8. This is of course to be expected since ZF6. is a special case of ZF8. for the derivation of which S8. was set up. The theorem is

Coll $_{\underline{x}}$ $(\underline{P} \ \& \ \underline{x} \in \underline{A})$

i.e. $(\exists \underline{y})(\forall \underline{x})(\underline{x} \in \underline{y} \Leftrightarrow \underline{x} \in \underline{A} \ \& \ \underline{P})$.

ZF7. We shall derive a theorem, from which, according to Wang 1954 a Choice-axiom follows in any system containing the ϵ - operator. The theorem is

$$(\forall x)(\exists y)(\forall z)\{z \in y \Leftrightarrow (\exists w)[w \in x \& z = \epsilon_u(u \in w)]\} \quad (W)$$

This can be rewritten as

$$(\forall x)\ \mathrm{Coll}_z\{(\exists w)[w \in x \& z = \tau_u(u \in w)]\}$$

Now we recall that S8. asserted

$$(\forall \underline{y})(\exists \underline{X})(\forall \underline{x})(\underline{R} \Rightarrow (\underline{x} \in \underline{X})) \Rightarrow (\forall \underline{Y})\ \mathrm{Coll}_{\underline{x}}\ (\exists \underline{y})(\underline{y} \in \underline{Y} \& \underline{R})$$

Let us make the following substitution in S8.

x for $\underline{Y}$, w for $\underline{y}$, z for $\underline{x}$,

and $z = \tau_u(u \in w)$ for $\underline{R}$; then we obtain (W) as the right hand member of the main implication of S8. Hence (W) is derivable in Bou. if the left hand member is derivable.

i.e. if $(\forall w)(\exists X)(\forall z)\ [z = \tau_u(u \in w) \Rightarrow z \in X]$ is derivable. Now $(\forall z)(z = T \Rightarrow z \in \{T\})$ is a theorem. Hence $(\exists X)(\forall z)(z = T \Rightarrow z \in X)$ is a theorem. From this follows

$$(\forall w)(\exists X)(\forall z)\ [z = \tau_u(u \in w) \Rightarrow z \in X]$$

by taking $\tau_u(u \in w)$ as the term T.

Thus (W) is derivable and a Choice-axiom follows in Bou. The carrying out of the above derivation is not in fact essential to our purposes here, which is to show that a Choice-axiom follows in Bou. For we shall now see (using the analysis given by Bernays. see

Chapter II. p. 76.) that S5., S7. are equivalent to an extremely strong form of the Choice-axiom.

Let us rewrite S5. in terms of the τ- operator.

$$(\underline{T} \mid \underline{x})\underline{R} \Rightarrow (\tau_{\underline{x}}\underline{R} \mid \underline{x})\underline{R}$$

i.e. $\underline{R}(\underline{T}) \Rightarrow \underline{R}(\tau_{\underline{x}}(\underline{R}))$ (H)

Now (H) is precisely Hilbert's '$\in$-formula' (with of course τ replacing $\in$). Bernays' form of this is

$$a \in C \Rightarrow \sigma(C) \in C \qquad A'_{\sigma}$$

which can be obtained from (H) by writing the class variable C for the predicate variable $\underline{R}$ and $\sigma(C)$ for $\tau_{\underline{x}}(\underline{R})$.

Again by writing class variables A, B for the predicate variables $\underline{R}$, $\underline{S}$ in S7. and by writing $\sigma(A)$, $\sigma(B)$ for $\tau_{\underline{x}}(\underline{R})$, $\tau_{\underline{x}}(\underline{S})$ respectively we obtain

$$A \equiv B \Rightarrow \sigma(A) = \sigma(B) \qquad A''_{\sigma}$$

Thus S5., S7. appear to be equivalent to a powerful form* of the Choice-axiom. We can not, of course, formally prove this equivalence within Bou. since we have no class formalism at our disposal to effect the necessary change of variables; however it follows that S5. names an object - $\tau_{\underline{x}}(\underline{R})$ satisfying the non-empty predicate $\underline{R}$, and as such is equivalent to A'_{σ} which names an object

---

*For a powerful form of the Choice-axiom, for sets, deducible from our formula (H), (see Sievpinski, 1957, p.93).

$\sigma(C)$ of the non-empty class C.

In the above considerations (especially in the derivation of (W)) we have perhaps laboured the point that, the τ- operator embodied in S5., S7. is a generalised Choice-axiom; this is not without reason for the contrary opinion is expressed in Wang (loc. cit.) and in Fraenkel-Bar Hillel 1958, p.184, where we find the words "There clearly exists a close connection between the ϵ- formula and the axiom of choice. This connection should however not be overstated, as it is occasionally done, in the form that the ϵ- formula is a kind of logical (or generalised) axiom of choice. Indeed the ϵ - formula allows for a single selection only, while the axiom of choice allows for a simultaneous selection from each member of an (infinite) set of sets and guarantees the existence of the set comprising the selected entities".

There appears to be here a confusion between a choice, as a statement of existence, and a choice, by the <u>naming of a representative</u>; this is embodies in the dichotomy between 'a single selection' and 'a simultaneous selection'. What in fact the τ- operator does is not to make a single selection or choice from a set but to <u>name</u> for any set S a member of that set. Thus if S is any (infinite) set and y is any member of S then $\tau_x(x \in y)$ denotes an object of the set y. The existence of the set of all such named objects (one for each $y \in S$) is guaranteed by the schema S8., that is to say, by the replacement axiom under the mapping ($\tau_x(x \in y)$, y) (See ZF8. below) .

The τ - operator, in so far as its use here distinguishes Bourbaki's system from the normal ZF system, is extremely important and we shall return to the above point later in this chapter (see section 4 below p. 109).

ZF8. This axiom can be obtained from the following two theorems, which were both derived from S8.

C. 53 (II. 66) $\mathrm{Coll}_{\underline{y}}\ (\exists \underline{x})\ (\underline{x} \in \underline{A}\ \&\ \underline{y} = \underline{T})$

This asserts that all the sets $\underline{y}$ which can be put in the form $\underline{T}(=\underline{T}(\underline{x}))$ for $\underline{x} \in \underline{A}$ form a set.

C. 54 (II. 78) The formula $\underline{x} \in \underline{A}\ \&\ \underline{y} = \underline{T}$ admits a functional graph $\underline{F}$ with respect to $\underline{x}$ and $\underline{y}$ such that $pr_1\ \underline{F} = \underline{A}$, $pr_2\ \underline{F} = \mathcal{E}_{\underline{y}}\{\underline{x} \in \underline{A}\ \&\ \underline{y} = \underline{T}\}$ and for all $\underline{x} \in \underline{A}$ one has $\underline{F}(\underline{x}) = \underline{T}$.

From C. 54 we get

$(\forall x)(\forall x')(x \in A\ \&\ x' \in A\ \&\ F(x) = F(x') \Rightarrow x' = x)$

Call this formula α, then from C. 53 we get

$\alpha \Rightarrow (\exists s)(\forall y)(y \in s \Leftrightarrow (\exists x)(x \in A\ \&\ y = F(x))$ which is ZF8.

ZF9. This appears to be unobtainable in Bou. We say this in view of the fact that Bourbaki seems to have chosen his axioms so as to yield ZF1 - 8 and no more and that ZF9. is known* to be independent of these.

---

*See for example Thiele E-J 1955.

Again this only reference to the occurrence of the so called extraordinary sets is in the Note Historique p. 106 where one is told to see II. 63 where it is proved that non $\mathrm{Coll}_x (x \notin x)$ is a theorem; but clearly this has no bearing either way.

Thus Bourbaki's system certainly contains the usual Zermelian set theory (omitting ZF9.) built on the predicate calculus.

One axiom of Bourbaki's which has not figured in our considerations is A3. which defines an ordered pair (couple) i.e. sets up the property of the primitive symbol $\supset$.
The normal way of dealing with ordered pairs is to use Kuratowski's method of defining the ordered pair (a, b) in terms of the plain sets $\{a,b\}$ and $\{b\}$. Thus:
Df. $(a, b) = \{\{a,b\}, \{b\}\}$

We can see no reason why this course is not adopted here as follows:

$\{a,b\}$ is an abbreviation for $\mathcal{E}_z(z=a \vee z=b)$

i.e. $\{a,b\}$ is an abbreviation for $\tau_y(\forall z)((z \in y) \Leftrightarrow z=a \vee z=b)$

and $\{b\}$ is an abbreviation for $\tau_y(\forall z)((z \in y) \Leftrightarrow z=b)$

Thus (a,b) is an abbreviation for $\{\{a,b\}, \{b\}\}$

i.e. for $\mathcal{E}_u(u = \tau_y(\forall z)(z \in y) \Leftrightarrow z=a \vee z=b) \vee u = \tau_y(\forall z)((z \in y) \Leftrightarrow z=b)$

and finally (a,b) is an abbreviation for $\tau_x(\forall u)((u \in x) \Leftrightarrow$

$$u = \tau_y(\forall z)((z \in y) \Leftrightarrow z=a \vee z=b) \vee u = \tau_y(\forall z)((z \in y) \Leftrightarrow z=b)).$$

The proof that (a,b) as defined here has the property of an ordered pair (as given in A3.) is simple. In fact Bourbaki sets this proof as an exercise (II. 70).

Now that we have seen the full extent of Bourbaki's system it is possible to return to the connection between the concepts of <u>relations collectivisantes</u> and <u>representable classes</u> of Bernays.

We shall first indicate informally that Bou. is a subsystem of $\bar{B}'$; where $\bar{B}'$ is the system obtained from $\bar{B}$ by replacing A5. by A_σ (note that A_σ', A_σ'' are derivable in $\bar{B}'$ on a suitable definition of σ).

S1 - 4. and S6.* are contained in Bernays' formulation of the predicate calculus.

S5., S7. are A_σ', A_σ'' respectively, as was discussed earlier in this chapter.

S8. which combines the sum axiom and replacement axioms (ZF4., ZF8) can be subsumed under A3. of $\bar{B}'$ which is a powerful combination of the same axioms.

---

* There is a slight difficulty here which would have to be cleared up in a formal approach. The class of terms in Bou. is more comprehensive than in $\bar{B}$, since the τ - terms have no cognate in Bernays' logical apparatus.

A1. of Bou. is precisely E2. of B'.

A2. of Bou. is a special case of A2. of B'.

A4. of Bou. is precisely A4. of B'.

A5. of Bou. is equivalent to Coll $_{\underset{\sim}{x}}$ (x is an integer) and as such is equivalent to Bernays' theorem that $\{x \mid Nu(x)\}$ is representable.

Thus it appears that Bou is contained in B'. From this a connection between the two concepts stated above is evident.

For let $\underline{R}(\underline{x})$ be any relation in Bou., then Coll $_{\underset{\sim}{x}}$ $\underline{R}(\underline{x})$ implies Rep $\{y \mid \underline{R}(y)\}$. For if not (i.e. if there exists a relation $\underline{R}(\underline{x})$ such that Coll $_{\underset{\sim}{x}}$ $(\underline{R}(\underline{x}))$ but not Rep $\{y \mid R(y)\}$) then by the mapping $\ominus$ (chapter II p. 77) the class $\{y \mid R(y)\}$ is in (1,1) correspondence with V. This cannot be so for by virtue of Coll $_{\underset{\sim}{x}}$ $\underline{R}(\underline{x})$ the class $\{y \mid R(y)\}$ is precisely the set $\mathcal{E}_{\underset{\sim}{x}}(\underline{R}(\underline{x}))$ and thus V would be a <u>set</u> by the replacement axiom.

This is of course only a partial result and the reverse implication would be desirable. We shall touch upon this in the next chapter.

Section 4 - The development of set theory

Since ZF1 - 8. are derivable in Bou. it follows that any development of set theory from these axioms which has so far been put forward* could be embodied here.

In fact Bourbaki's development of set theory does (essentially) follow the accepted pattern; it will therefore be sufficient for us to enlarge upon divergences from this pattern; or upon those notions, the handling of which in the past has presented difficulties.

We have seen in section 1 that the τ- operator is an essential part of Bourbaki's logical apparatus. In the last section we have seen that it is equivalent to a generalised Choice-axiom. The questions therefore arise whether <u>implicit</u>** application of this axiom can come about, and if so at what points in Bou. they are generated.

---

* E.g. P. Suppes, 1960.
cf. Bernays/Fraenkel Historical introduction.

** By this we mean applications of the axiom, which are not obvious and which are often difficult to detect.

It is perhaps not too surprising that these implicit applications of the Choice-axiom occur at the very places where Bourbaki's treatment is novel; that is wherever he makes essential use of the τ- operator. viz. in the definitions of cardinals, order types (hence ordinals) and finiteness.

The introduction of these notions appears to have been formulated with Cantor's original conceptions in mind. We shall therefore preface each example with Cantor's corresponding notion.

Definition of Cardinals (III. 85).

Inherent in Cantor's definition* of the cardinal number of a set a was the 'set of sets equivalent to a'. This intuitively simple definition cannot be mirrored here because of the unwarranted set of sets.....'. The collection of sets equivalent to the set a is non-collectivisante. Bourbaki's way out of this is first to define the dyadic predicate of equivalence Eq(X,Z) - 'X and Z are equivalent' and then to define the cardinal number of the set a as the representative (objet privilégie) of the collection of sets equivalent to a. Thus:

Card (a) is an abbreviation for $\tau_z(\text{Eq}(a,z))$

Hence cardinal numbers are terms and thus sets.

---

*In fact Cantor did not define cardinal numbers but used a working definition (Jourdain p. 86, cf. Chapter I). The definition quoted above was Frege's 'rigorisation' of Cantor's notion, which however was invalidated by Russell's criticism.

Note: This definition is not possible using a weaker selection operator formulated for sets only.

From the above definition we see that if a and b are equivalent i.e. if Eq(a,b) then $\tau_z(Eq(a,z) = \tau_z(Eq(b,z)$

i.e. card (a) = card (b).

The converse can also be shown; and thus questions involving cardinals can be reduced to questions involving equivalence in the normal manner.

Consider now the sentence - 'Let $(a_i)_{i \in I}$ be a family of cardinals' (III. 58). (For any set A, Bourbaki defines the family of sets $(X_\iota)_{\iota \in A}$ by means of the identity transformation such that $(X_\iota)_{\iota \in A}$ is identical to A). Then if I is infinite, the assertion of the existence of such a set of cardinals requires the Choice-axiom. To illustrate this let I be denumerable.

$$I = \{a_1, a_2, a_3 \ldots\ldots\}$$

$$\text{and } a_i = \text{card } E_\iota = \tau_x(Eq(E_\iota, x))$$

$$\text{then } I = \{\tau_x(Eq(E_1, x)), \quad \tau_x(Eq(E_2, x)) \ldots\ldots\}$$

To secure the existence of I we are effectively applying the Choice-axiom to the set $E = \{E_1, E_2 \ldots\ldots\}$

Thus an infinite number of selections have been made and the result of all these has been combined into a set. We shall for the moment delay comment on this until after we have considered our other examples.

Definition of order types (III. 50).

Again Frege's 'rigorisation' of Cantor's original notion (cp. cit. p. 112) is inadmissable here and Bourbaki utilises the τ- operator as before.

First the dyadic predicate of similarity Is(x,y) - 'There exists an order - preserving isomorphism between (the ordered sets) x and y' is defined. Then the order type of the set a is defined as the representative of the predicate Is(a,x) i.e.

Order (a) is an abbreviation for τ_x(Is (a, x))

Ordinal numbers can now be defined (precisely as Cantor did cp. cit. p. 137) as the order types of well-ordered sets.

A similar situation, concerning the Choice-axiom, occurs here when an infinite number of ordinals are combined into a set.

Definition of finiteness. (III. 65).

Cantor (cp. cit. pp. 97-9) deals with the concept of finiteness by generating a sequence of sets, each one obtained by adding a new element e_v (for v = 1,2,...) to the last one; starting from a single element e_o. He calls the sets so obtained "finite cardinal numbers" and proves all the usual properties of natural numbers for them. Thus in Cantor the notions of finite set and natural number are fused and the two terms are synonymous.

Now this fusion, if carried out consciously and rigorously is intuitively* satisfying and it is, to some extent, reproduced by Bourbaki.

In fact Bourbaki inverts the order of Cantor's procedure by using a provable consequence of this procedure, namely, that no two cardinal numbers thus generated are equal, (Cf. Dedekind's definition of finiteness) as his definition. Thus

<u>A cardinal c is finite if</u> $c \neq c + 1$

A set is finite if its cardinal is finite. From this definition the terms of the sequence

$0, \quad 0 + 1, (0 + 1) + 1, \ldots\ldots$

are finite cardinals i.e. <u>natural</u> <u>numbers</u>, which is the alternative name given to them by Bourbaki.

To prove that the finite cardinals in the above sequence do in fact characterise the natural numbers, it is sufficient to show that they satisfy Peano's axiom. Now the induction principle is proved by reductio ad absurdum (III. 67) and the remaining axioms

---

*Not all definitions of finiteness have this intuitive clarity; e.g. a set is finite if, and only if, it is a double well ordered set. See Bernays (1958) p. 151.

follow from theorems that Bourbaki proves for all cardinals (III. 59-60). Also from these theorems Bourbaki proves the result that

$$(\forall x)(x \subset E \ \&\ E \text{ is finite} \Rightarrow \text{Card}(x) < \text{Card}(E))$$

and its converse. From this it is seen that Bourbaki's definition of a finite set is equivalent to Dedekind's definition of a finite set as one not equivalent to a proper subset of itself. However it is known (Tarski 1924) that the definition of finiteness as equipollence to natural numbers requires the Choice-axiom to prove it equivalent to Dedekind's definition. The proof of the above result (III. 66) as well as the theorems from which it is derived make no explicit use of this axiom - it has thus been used implicitly.

The three examples given above illustrate where and in what manner the Choice-axiom is used implicitly in Bou. Gandy (1959) in his review of Bourbaki's system, sees in these implicit applications the most severe drawback to Bourbaki's system. By building in the τ- operator Bourbaki has formalised a 'completely classical' approach to set theory; an approach which 'is too ambiguous to give a definite answer to some problems of analysis (e.g. the continuum problem)'.

In addition to the generation of implicit applications of the Choice-axiom there is another aspect of the τ- operator which is relevant here. Even when we know that a particulat theory or portion of set theory (e.g. the theory of cardinals) is based essentially on the τ- operator it is not at all easy to discern just how strong[*] the choice assumption is; either at the outset (e.g. in the actual definition of a cardinal) or at various stages of the theory. Again the problems of the dependence (or independence) of the Choice-axiom from the other axioms of set theory and even of the various forms of the axiom itself have not yet been solved[**]. This, together with apparent paradoxical results (e.g. Banach-Tarski, 1924) arising from the Choice axiom would suggest that a close surveillance of its use and examination of its strength in particular cases is very necessary. (Cf. the attitude in Sierpinski 1957, p. 96).

We shall complete our discussion of the τ- operator in the next section; there we shall see that its use is to some extent justified when considered in the broader context of Bourbaki's programme.

---

[*] i.e. in relation to a conventional form of the axiom formulated for sets.

[**] That is, not in systems of the type ZF; but results have been obtained in systems employing an infinite number of individuals or employing extraordinary sets. See Fraenkel 1922, Mendelson 1956.

Section 5 - Bourbaki's programme

It has been our purpose in this essay to assess two systems of set theory which have been put forward to encompass the whole of mathematics. The character of this assessment has been one of exposition and comment on the system, qua systems. The manner in which the theories and concepts of mathematics are to be (or have been) axiomatised within these systems is relevant only if it affects the character of the systems. We shall be content therefore to set down some general remarks.

Any single system set up to encompass mathematics must be a compromise between opposing tendencies.

On the one hand the aim of achieving maximum rigour with the minimum of existential assumptions is of primary importance. This often calls for investigations of a metamathematical kind such as attempts to eliminate (or replace) impredicative definitions, axiom schemata etc. from the system. These investigations can only be carried out for systems which have been framed with meticulous care; that is, where attention has been paid to such points as the status of definitions, the use and mention of signs etc.

On the other hand the practical aspect of encompassing mathematics in a single system lays its own claims on such a system. The ability to reproduce all modes of mathematical reasoning as faithfully as possible is essential; so too is the ability to translate easily from mathematics to set theory if the project is not to become outdated. Both these considerations require a system to possess a logical apparatus whose expressive power is as comprehensive as consistency permits.

The needs of the working mathematician moreover must also be considered. For him attainment of rigour which involves the laying aside of well tried, but logically vague* procedures in favour of safer but perhaps more artificial ones is stultifying. Of much more use to him is the exhibition of the interaction and mutual connection between seemingly diverse branches of mathematics. He seeks "to find the common ideas of these theories, buried under the accumulation of detail properly belonging to each of them, to bring these ideas forward and to put them in their proper right". (Bourbaki, (1950) p. 223).

The systems of set theory constructed by Bernays and Bourbaki respectively, illustrate extremely well these two tendencies.

B. was constructed from the point of view of a logician. Bernays aim was to set up a theory capable of encompassing mathematics and then to proceed with a metalogical investigation of this theory; that is to consider "the further axiomatical questions of eliminability, relative consistency and independence". Bernays (1958) p. 43.

---

* E.g. Applying an infinite number of dependent choices. See Sierpinski (op. cit.) pp. 129 - 31.

Thus Bernays has given substance to the suggestion that the foundations of mathematics and of set theory are one and the same thing, and his approach to the problem of "Foundations" is an intensive investigation of set theory.

Bourbaki's programme has altogether a more practical character. It is to enable mathematicians to gain an insight into the assumptions and directions particular to their own branches in perspective to the whole edifice of mathematics. Such a purpose cannot be achieved by simply setting up a set theory capable of yielding all of mathematics; the development must be realised. That is the whole of mathematics must be translated into (or in some cases reconstructed within) the language of set theory.

A cogent and very lucid account of the guiding principles behind this translation is to be found in Bourbaki 1950. On this we shall base a short account* of the dominant idea, which is that of structure.

A structure consists of (a) Objects (sets as understood here).

(b) Certain relations into which these objects can enter.

(c) Axioms which these relations satisfy.

---

* Cf. J.A.P. Hall (1960), L. Felix (1960) and M. Colmez (1961).

"To set up the axiomatic theory of a given structure, amounts to the deduction of the logical consequences of the axioms of the structure, excluding every other hypothesis on the elements under consideration (in particular, every hypothesis as to their own nature)".

Using this notion Bourbaki thinks it possible to examine the whole of mathematics. "The organising principle will be the concept of a hierarchy of structures, going from the simple to the complex, from the general to the particular".

"At the centre of our universe we have the three great types of structure" or <u>mother structures</u> which are

I. Algebraic structures in which the relations are "laws of composition" (i.e. relations between three elements determining the third uniquely in terms of the other two).

II. Order structures where the relations are order relations.

III. Topological structures which "furnish an abstract mathematical formulation of the intuitive concepts of neighbourhood, limit and continuity, to which we are led by our ideas of space".

"A considerable diversity exists in each of these types; one has to distinguish between the most general structure of the type under consideration, with the smallest number of axioms, and those which are obtained by enriching the type with supplementary axioms, from each of which comes a new harvest of consequences".

In addition to the three primary types of structure there are also multiple structures. These involve two or more of the primary structures combined "organically by one or more axioms which set up a relation between them".

Finally "further along we come to the theories properly called particular. In these the elements of the sets under consideration, which, in the general structures have remained entirely indeterminate, obtain a more definitely characterised individuality. At this point we merge with the theories of classical mathematics.....".

Thus Bourbaki's programme is to clarify and simplify* the whole edifice of mathematics by the axiomatic method. Within this programme set theory has the role of a basic language (or logic, for Bourbaki the terms are synonymous, Cf. I. 4-9.) which contains all the vocabulary and syntax necessary to conduct mathematics. The originality that this programme possesses and its unquestionable value to mathematics make it a worthwhile contribution to "Foundations"

---

*Discussion of the pedagogic aspect of Bourbaki's programme as well as amplification of some of the points raised in this section are to be found in L. Felix, 1960.

---

and indeed set against this, the valid criticisms of vagueness and ambiguity engendered by the use of the τ - operator take on a different aspect. Within the bounds of a logician's attempt at reconstructing the Foundations of Mathematics these criticisms are valid. In Bou, however, they are of secondary importance and perhaps a necessary evil, to be expected in view of the scope of Bourbaki's programme.

Nevertheless it could be asserted that Bourbaki's uncritical acceptance of all present day mathematics is an act of faith, and that "Foundations" exists as a subject to obviate such acts of faith. There is considerable truth in this assertion and Bourbaki's answer would be that his programme is being created in the same spirit that all mathematics has been created - in the hope that posterity (i.e. logicians) will provide justification for its 'doubtful' regions.

Chapter IV

Gödel's system and some results concerning it

In the preceding two chapters we examined two different systems of set theory. Both these systems were designed to deal with the construction of the whole of mathematics. In the first part of this chapter we shall describe (briefly) a system of set theory which was constructed for an entirely different purpose; namely, to deal with difficulties arising in set theory itself. By far the most important and outstanding of these difficulties is the Continuum Hypothesis. The only real progress that has been made towards a solution of this problem was made by Gödel in 1939. He proved that the Continuum Hypothesis is consistent with (i.e. cannot be refuted within) a certain system of set theory. It is this system that we shall describe now, basing our description on Gödel 1940.

Section 1. Gödel's system of set theory.

We shall call this system G. The system G. and the system B. considered in chapter II are both modifications of the system of set theory expounded by Bernays in the Journal of Symbolic Logic (Bernays JSL).

G. was set up very much earlier than B. and is considerably closer to Bernays JSL. We shall be commenting on the relationship between B. and G. throughout this section.

The logical basis for G. is the restricted predicate calculus with equality ' = '.

The relation of membership, denoted by ' $\in$ ', is primitive, as are the notions of class 'Cls(A)' and set 'M(A)'.*

Large latin letters X, Y, Z,....; are used as class variables, small ones x, y. z,....; as set variables.

In contrast to the membership relation in B. (whose left hand side was restricted to set variables) the membership relation in G. can occur in any of the four following contexts:

$X \in Y$, $X \in y$, $x \in Y$, $x \in y$.

---

* Where Gödel uses a german letter, we shall use the equivalent english letter with a bar underneath. However for the sake of convenience we shall use Bernays' symbols for the logical constants and quantifiers (as used in chapter II).

Gödel divides his axioms into five groups A,B,C,D,E.

Group A.

A1. $\underline{Cls}(x)$

A2. $X \in Y \rightarrow \underline{M}(X)$

A3. $(u)(u \in X \leftrightarrow u \in Y) \rightarrow X = Y$

A4. $(x)(y)(Ez)(u \in z \leftrightarrow (u = x \vee u = y)$

A1. asserts that all sets are classes. Thus Gödel identifies a class and a set which have the same members rather than speaking of a class being 'represented' by a set.

A2. asserts that classes possessing elementhood are sets. A class which does not possess elementhood and which is thus not a set is called a proper class:

$\underline{Pr}(X) \leftrightarrow \overline{\underline{M}(X)}$

A3. is the axiom of extensionality. Using A1, a special case of A3 is the following:

$(u)(u \in x \leftrightarrow u \in y) \rightarrow x = y$

which is E2 in B.

Thus the relation of equality is primitive for sets <u>and</u> classes in G. as opposed to B. where, as was explained, there is no need for a primitive relation of equality between classes.

A4. is the pair set axiom which posits the existence of the plain (unordered) set $\{x,y\}$ whose only members are x and y.

Thus

$$u \in \{x,y\} \leftrightarrow (u = x \quad u = y)$$

(The set $\{x,y\}$ is unique by A3.).

From A1 - 4 come the usual notions of ordered pairs $\langle xy \rangle$, inclusion $\subseteq$ and the emptiness of a class:

$$\underline{Em}\ (X) \leftrightarrow (u)(u \notin X)$$

<u>Group B.</u>

The following axioms are concerned with the existence of classes and constitute a list of the basic* conditions which determine classes.

B1. 'Axiom of $\in$ - relation'

$$(EA)(x)(y)(\langle xy \rangle \in A \leftrightarrow x \in y)$$

B2. 'Axiom of intersection'

$$(A)(B)(EC)(u)(u \in C \leftrightarrow u \in A \ \& \ u \in B)$$

B3. 'Axiom of the complement'

$$(A)(EB)(u)(u \in B \leftrightarrow (u \notin A))$$

B4. 'Axiom of domain'

$$(A)(EB)(x)(x \in B \leftrightarrow (Ey)(\langle yx \rangle \in A))$$

---

* For an intuitive account of the motivation for choosing the particular classes in B1 - 8 see A. Borgers (1948).

B5. 'Axiom of direct product'

(A)(EB)(x)(y)($\langle yx \rangle \in B \leftrightarrow x \in A$)

B6. (A)(EB)(x)(y)($\langle xy \rangle \in B \leftrightarrow \langle yx \rangle \in A$)

B7. (A)(EB)(x)(y)(z)($\langle xyz \rangle \in B \leftrightarrow \langle yzx \rangle \in A$)

B8. (A)(EB)(x)(y)(z)($\langle xyz \rangle \in B \leftrightarrow \langle xzy \rangle \in A$)

Axioms 6,7,8 are 'axioms of inversion'.

As Gödel points out, the class A in axiom B1 and the class B in axioms B5 - 8 are not uniquely determined, since nothing is said about those sets which are not pairs (or triples) i.e. whether or not they belong to A (or B). This is not the case, however, with the classes C in axiom B2, and B in axioms B3, B4 which are uniquely determined (by A3). This uniqueness makes possible the following definitions.

Intersection $x \in A.B \leftrightarrow x \in A \;\&\; x \in B$

Complement $x \in -A \leftrightarrow x \notin A$

Domain $x \in \underline{D}(A) \leftrightarrow (Ey)(\langle yx \rangle \in A)$

We note the following close correspondence between the classes posited in B1 - 8 and the 'primary constituents' of B. (see chapter II p. 48);

B1. E

B2. $A \cap B$

B3. $\overline{A}$

B4. $\Delta_1 A$

B5. $A \times B$

B6. $\breve{A}$

B7. } $\overset{\smile}{A}$, $\overset{\prec}{A}$
B8. }

Group C.

These axioms are concerned with the existence of sets only. Disregarding slight differences of formulation these axioms are identical to ZF5 (infinity), ZF4 (Union), ZF3 (Power set) and ZF8 (Replacement) respectively.

C1. $(Ea)\,\boxed{\underline{Em}}\,(a)\ \&\ (x)(x \in a \rightarrow (Ey)(y \in a\ \&\ x \subset y\,)))$

C2. $(x)(Ey)(u)(v)(u \in v\ \&\ v \in x \rightarrow u \in y)$

C3. $(x)(Ey)(u \subseteq x \rightarrow u \in y)$

C4. $(x)(A)(\underline{Un}\,(A) \rightarrow (Ey)(u)(u \in y \leftrightarrow (Ev)(v \in x\ \&\ \langle uv \rangle \in A)))$

Where in C4

$\underline{Un}\,(X) \leftrightarrow (u)(v)(w)((\langle vu \rangle \in X\ \&\ \langle wu \rangle \in X) \rightarrow v = w)$

i.e. 'X is single-valued'.

Group D.

$\underline{Em}\,(A) \rightarrow (Eu)(u \in A\ \&\ \underline{Em}\,(u.A))$

This is the Fundierungsaxiom (ZF9) in precisely the same form as A7 of B.

Group E.

$(EA)(\underline{Un}\,(A)\ \&\ (x)\,\boxed{\underline{Em}}\,(x) \rightarrow (Ey)(y \in x\ \&\ \langle yx \rangle \in A)))$.

This is the Choice-axiom in a strong form since 'it provides for the simultaneous choice by a single relation, of an element from each set of the universe under consideration' (p. 6).

The main difference between G. and B. is in their respective treatments of classes. Even this however is, as we shall see, one of approach rather than substance.

In B. the concepts of class and set are kept quite distinct and are related only through the notion of representability, so that a set x, having the same members as a class X, is said to represent that class. In G., on the other hand, any set x is a class (by A1), and (by A3) would be equal to X if the two had the same members.

Of much more importance than this however is the difference between the <u>formation</u> of classes in G. and B.

In B. any well-formed predicate P(x) has, as its extension, the class of sets $\{x \mid P(x)\}$ satisfying it. Such a formation of classes is achieved by means of the 'Church schema';

$$a \in \{x \mid P(x)\} \leftrightarrow P(a)$$

Having thus formed a domain of classes in strict correspondence to the domain of predicates, Bernays proves* that any class of this domain can be constructed from a small number of basic class notions ('primary constituents' see above).

---

* Bernays I 1937, VII 1954. This is his famous <u>class theorem</u>.

In G. we start with a small number of basic class notions (B1 - B8) as the 'definition' of a class, and build up all our available classes from these. It would therefore be desirable to relate these classes to the predicates of the system. This is precisely what Gödel does in a series of metatheorems, M1 - M6. We shall describe the basic one M1 and its generalisation M2 - the others being modifications of these.

For the statement of M1 we need the notion of a primitive propositional function (ppf). Characterised intuitively a ppf is a well-formed formula containing quantification over set variables only, more precisely:

Let Π, Γ denote variables or particular classes (see below) then

(1) $\Pi \in \Gamma$ is a ppf

(2) If ϕ, ψ are ppf then so are $\bar{\phi}$, $\phi \& \psi$.

(3) If ϕ is a ppf then so is $(Ex)\phi$, and any result of replacing x by another set variable is a ppf.

(4) Only formulae obtained by (1) - (3) are ppf.

M1 asserts that the extension of any ppf is a class, formally: If $\phi(x_1 \ldots\ldots x_n)$ is a ppf containing no free variables other than $x_1 \ldots\ldots x_n$ (but not necessarily all these) then there exists a class A, such that for any sets $x_1 \ldots\ldots x_n$

$$\langle x_1 \ldots\ldots x_n \rangle \in A \leftrightarrow \phi(x_1 \ldots\ldots x_n)$$

The proof of this metatheorem is by cases, using induction on the length of the formula.

It should be noted that the axiom of extensionality (A3) does not guarantee the uniqueness of A, since nothing is said in M1 about sets which are not n - tuplets being members of A. (This is of course the same situation as in axioms B1 and B5 - 8). On this point, Gödel proves a later metatheorem (M4) which restricts membership of A to n - tuplets (A thus becomes an n - ary relation); this measure, combined with a generalised A3 applying to n - tuplets, secures uniqueness for the class A.

Whilst the notion of a ppf is a wide one, there is an important class of formulae that it does not cover; this is the class of formulae containing symbols introduced by definition, Gödel classifies these into four types as follows:

(1) Particular classes 0, V,......

(2) Notions M(X), Pr(X), X ⊆ Y,......

(3) Operations -X, D(X), X.Y,......

(4) Kinds of variables x, X,......(these are defined by notions).

These defined symbols are usually introduced by expressions which make use of quantification over classes and of previously defined symbols. To characterise such expressions Gödel extends the idea of a ppf to a propositional function (pf); essentially a pf is a ppf which may include previously defined symbols and quantifiers applied to classes. Thus the following would be the 'kind of definition' for special (particular) classes.

'A special class A is introduced by a defining postulate $\phi(A)$, where ϕ is a pf containing only previously defined symbols and it has to be proved first that there is exactly one class A, such that $\phi(A)$.' Similar 'kinds of definition' follow for defined notions, operations and variables.

Gödel next introduces the property of <u>normality</u>:

A notion $\underline{B}$ is normal if there is a ppf ϕ such that

$\underline{B}(X_1 \ldots\ldots X_n) \leftrightarrow \phi(X_1 \ldots\ldots X_n)$.

An operation $\underline{A}$ is normal if there is a ppf ϕ such that

$Y \in \underline{A}(X_1 \ldots\ldots X_n) \leftrightarrow \phi(Y, X_1 \ldots\ldots X_n)$

A variable is normal if its range consists of the elements of a class. From these a pf is normal if it contains normal notions, normal operations and normal bound variables.

Thus a pf is normal if it can be reduced to (or replaced by) a ppf. It therefore follows that M1 applies to any pf ϕ provided ϕ is normal. This assertion is the content of M2. It is clear that the whole apparatus of normality is designed to prevent the formation of classes as the extensions of pf which contain non-eliminable definitions or non-eliminable class quantification.

Leaving aside the question of defined symbols, it is seen that the idea of a ppf with a single free variable* coincides with the usual idea of a well-defined condition. This being so the formation

---

* If the ppf ϕ in M1 has a <u>single</u> free variable then the problem of the uniqueness of A disappears - A will be unique by A3.

of classes in G. (through M1) and in B. (through the Church schema) are identical (i.e. the same domain of classes is produced).

In view of the above remarks the systems G., B. are clearly equivalent and differ in form[*] rather than substance. Thus results which have been proved in relation to G. can be taken as holding for B. without qualification and we shall make no distinction between the systems when discussing such results in the next section.

---

[*] This difference in form is, of course, important when it allows for the simplification of metamathematical results. E.g. the proof of M1 is considerably simpler than Bernays' class theorem which achieves the same purpose from the opposite direction.

Section 2. The relationship between G. and ZF.

It is evident from the considerations of Chapters I - III that the Bernays-Gödel set theory is more comprehensive than (i.e. contains) the set theory of Zermelo-Fraenkel; the following question therefore presents itself.

Does the greater comprehensiveness of G. (embodied in the class formalism) entail any inconsistency in G. not existing in ZF; more generally, precisely how much stronger is G. than ZF? This question is completely answered by the following results*:

(1) If the ZF system of set theory is consistent then so is the system G.

(2) Any theorem of G. involving set variables only is a theorem of ZF. (i.e. the only theorems provable in G. which are not obtainable in ZF. are theorems essentially involving classes).

We shall give a brief account of the Rosser/Wang proof of result (1). (Cf. Wang/McNaughton 1953 pp. 21 - 2).

The starting point for this proof is the Löwenheim-Skolem theorem which asserts that any consistent system is satisfiable in the domain of natural numbers.

---

* (1) was proved first by I.M. Novak (1948) and then (more generally) by Rosser/Wang (1950). (2) is evidently due to Mostowski (see Fraenkel/Bar Hillel p. 122 footnote). It is also obtained by Rosser/Wang (op. cit).

Thus, on the hypothesis that the system* ZF is consistent, we can assert that it has a denumerable model having the following character.

To every set x of ZF. there can be assigned a number m and to the relation $\in$ of ZF there can be assigned an arithmetical relation $\in^*$ such that

$x \in y$ if and only if $m \in^* n$,

where m and n are the numbers assigned to the sets x and y. The next step, which is the substance of the proof, is to show how the additional objects of G. (i.e. the classes) can be included in the model by assigning to each of them a natural number. This is done by considering axioms A1, B1 - B8 of G. (p. 120 above) as nine operations which yield classes when applied to certain sets and classes. The numbers to be assigned are all of the form $2^m . 3^n . 5^k$ where m takes the values 1 to 9 (according to which operation is being considered) and n, k take the values indicated below).

---

* ZF here refers to Z1 - Z8 in Wang (1949) which does not differ essentially from our ZF1 - 8; ZF9 or axiom D in G. is left out of these considerations - it is, in any case, independent.

A1. $\underline{Cls}(x)$

If n is assigned to x then $2^1 . 3^n . 5^k$ is assigned to the class with the same members as x; k is arbitrary.

B1. $(EA)(x)(y)(\langle xy \rangle \in A \leftrightarrow x \in y)$.

To the class A the number $2^2 . 3^n . 5^k$ is assigned; n, k are arbitrary.

B2. $(A)(B)(EC)(u)(u \in C \leftrightarrow u \in A \,\&\, u \in B)$

To the class C the number $2^3 . 3^n . 5^k$ is assigned where n, k have been assigned to the classes A, B respectively.

B3. $(A)(EB)(u)(u \in B \leftrightarrow (u \notin A))$

To the class B the number $2^4 . 3^n . 5^k$ is assigned where n has been assigned to A; k arbitrary.

The axioms B4 - 8 can be treated in the same manner. On completing the list A1, B1 - 8 one must repeat the process in order to be able to assign numbers to those classes provided by B2 (say) which depend on classes provided by a later axiom B6 (say).

In this way every class constructed by means of A1, B1 - 8 has a number assigned to it in such a manner that no two different classes have the same number assigned to them.

A new arithmetical relation $\bar{\in}$ can now be defined such that if m is assigned to the set x and n to the class Y then

$m \bar{\in} n$ if and only if $x \in Y$.

Now the remaining axioms of G. (i.e. the set axioms) do not differ essentially from their counterparts in ZF; hence the above construction has provided a model for G. as an extension of the original ZF model guaranteed by the Löwenheim-Skolem theorem. Thus G. is consistent relative to ZF.

It is clear from this result and from result (2) above that the essential difference between G. and ZF is one of expressive power rather than strength* in the accepted use of the word for mathematical theories, and that the class formalism of G. does not enable one to prove any stronger results about sets than those already available in ZF (one naturally expects results concerning the relationship between sets and classes to be provable in G).

There is an obvious similarity between result (2) and Hilbert's ι - theorem (chapter II p. 39); and the class formalism can thus be likened (in its effect) to a theory of descriptions which increases the richness of a system's vocabulary without increasing the assumptions of the system.

---

* After giving their proof Rosser/Wang state 'It thus appears that despite the appearance to the contrary the systems of ZF, G. are of essentially equal strength'.

In conclusion we mention the important fact that G. is a weak extension of ZF in that the class formalism of G. is predicative. This is achieved in G. (Cf. chapter II p. 44) by not allowing bound class variables to figure in the predicates which determine classes. Such variables are allowed, however, provided that they are eliminable - which is precisely what 'normality' for a formula means. This point is seen more easily in B. where there are no bound class variables at all and thus the predicate occurring in the Church schema obviously cannot contain them. This fact is important in the context of relative consistency, for it is doubtful* whether a system comprising ZF plus an impredicative class formalism could be proved consistent relative to ZF. Certainly the Rosser/Wang proof outlined above would not carry through for this type of extension since entities (classes) defined in terms of a totality of which they were members could not be assigned numbers in the same manner as the classes of G.

---

* See Fraenkel/Bar Hillel pp. 326 - 8.

Section 3. Conclusion.

If we view the systems of ZF and G. (i.e. effectively B.) in the light of the comments of the last section we see that there is little to choose between them if we are seeking a formal system which can be said to contain all of mathematics. The additional ability of G. to deal with proper classes, important as it is, is not essential for mathematics. This could have been discerned from Bernays' interpretation of classes as ideal objects as opposed to sets as mathematics things. (Chapter II p. 44). In fact the sets necessary for mathematics are in no way as comprehensive as the proper classes of G. but are, for the most part, subsets of sets wasily securable in G.

A consequence of this is that the antinomies of Russell and Burali-Forti, depending as they do on a questionable manipulation of proper classes, do not enter the orbit of normal mathematical reasoning.

It is thus clear that the ZF set theory, which is the essential portion of either Bourbaki's or Bernays' system, is certainly capable of containing all known mathematical reasoning; and that the task of reshaping Cantor's notions to avoid the antinomies and at the same time producing a system comprehensive enough to contain mathematics can be completed in either of the systems discussed in chapters II and III. If a choice has to be made between these systems then this will depend (as we saw in chapter III) on additional factors - i.e. on one's total programme in 'Foundations'. If this programme is concerned with the overall logical assumptions inherent in mathematics then Bernays' system is more suitable. On the other hand, the greater flexibility of Bourbaki's system is clearly an advantage if one is concerned with the actual construction of mathematics and with the logical relations between its parts.

Having said this we shall turn now to some general observations on set theory made by Gödel in 1947. These observations were made in connection with Gödel's result concerning the Continuum Hypothesis.

We recall that the result is that the Continuum Hypothesis cannot be refuted within any of the set theories considered in this essay. Gödel proves this result by first considering a certain hypothesis - A, about the nature of all sets in G., namely that they are 'constructible' and then showing that from A the Continuum Hypothesis can be derived and that A is consistent with the axioms of G. (provided that these axioms themselves are consistent). Intuitively a set is 'constructible' if it is 'definable in terms of ordinal numbers..... by means of transfinite recursions, the primitive terms of logic and the ε - relation, admitting, however as elements of sets and of ranges of quantifiers only previously defined sets'. The sets defined in this way form a model for set theory in which the hypothesis A, and therefore the Continuum Hypothesis, is true.

Consideration of the Continuum Hypothesis shows that there are three alternatives - it is provable, disprovable or undecidable. The complete failure to obtain any results about the cardinal of the continuum* over the past seventy years would indicate that there is little hope of proving the Hypothesis. On the other hand Gödel's result rules out the possibility of disproving the Hypothesis on the basis of the present axioms. The third alternative is the one which Gödel thinks is most likely and he interprets his result as evidence in favour of it.

To say that an hypothesis is undecidable on the basis of a system of axioms is to say that these axioms constitute an insufficient characterisation of their subject matter; it is thus to call for new axioms. This is what Gödel does, arguing in the following manner.

---

* By Zermelo's theorem the continuum can be well-ordered - its cardinal is therefore an aleph - $\aleph_\beta$. Not only has it not been settled whether $\beta = 1$ but it has not even been established what kind of ordinal it is; neither has an upper bound, however large, been assigned to β (see Sierpinski 1957 p. 412).

In the first place the two sets $Z(\aleph_0)$ and the continuum, whose wquivalence is asserted in the Hypothesis, are of a totally different kind. The first set is built up by an iterative procedure and is 'constructible' in the sense explained above. As opposed to this, the continuum is composed of sets 'in the sense of arbitrary multitudes, irrespective of, if, or how they can be defined'. The fact that both these sets can be formed from the same system of axioms is a clear indication for Gödel that these axioms do not contain a complete description of the 'well-defined reality' in which the sets occur.

The second argument for the undecidability of the Hypothesis rests on a survey of the facts (as known in 1947) regarding the truth of Hypothesis. Not only are there no facts available to support the Hypothesis but there are many* which would have seemingly paradoxical results if adjoined to the Hypothesis. Now since Gödel's result makes impossible a refutation of the Hypothesis on the basis of the present axioms there appears to be a strong case for supplementing

---

* 'E.g. The existence of certain properties of point sets..... for which one has succeeded in proving the existence of undenumerable sets having these properties, but no way is apparent by means of which one could expect to prove the existence of examples of the power of the continuum'. (p. 523).

---

the known axioms in order to obtain such a refutation. Gödel does not of course set down any possible new axioms but does indicate a direction in which these should be sought. This direction is that of the definability of sets. Thus referring to the characteristic difference between $Z(\aleph_0)$ and the continuum as one of definability, and to the fact that using this notion (in his own consistency proof) he has obtained a partial result he states 'It is plausible that the continuum problem will not be solvable by means of the axioms set up so far, but, on the other hand, may be solvable by means of a new axiom which would state or at least imply something about the definability of sets'.

BIBLIOGRAPHY

P. BERNAYS, JSL.
A system of axiomatic set theory.
Journal of Symbolic Logic.
1937. I. 2, pp. 65-77: 1941. II. 6, pp. 1-17: 1942. III. 7, pp. 65-89:
IV. pp. 133-145: 1943. V. 8, pp. 89-106: 1948. VI. 13, pp. 65-79:
1954. VII. 19, pp. 81-96.

P. BERNAYS, 1958.
Axiomatic set theory. Amsterdam. (226 pp.).

A. BORGERS, 1948.
Development of the notion of set and of the axioms of sets.
Synthese. 7, pp. 374-390.

N. BOURBAKI, 1950.
The Architecture of mathematics.
American Mathematical Monthly, 57, pp. 221-232.
1954.
Théorie des Ensembles:
chapitre I, chapitre II.
Actualités Scientifiques et Industrielles 1212.
1956.
Théories des Ensembles:
chapitre III.
Actualités Scientifiques et Industrielles 1243.

G. CANTOR.
Contribution to the founding of the theory of transfinite numbers,
translated by P. JOURDAIN (Dover Reprint).

M.J. COLMEZ, 1961.
La structure des mathématiques modernes.
Association des Professeurs de Mathématiques de l'Enseignement
Public. No. 217, pp. 3-10.

L. FELIX, 1960.
The modern aspects of mathematics.

A. FRAENKEL, 1922.
Über den Begriff "definit" und die Unabhangigkeit des Auswahlaxioms.
Sitzungberichte der Preussischen Akademie der Wissenschaften, pp. 253-257.

FRAENKEL/BAR HILLEL, 1958.
Foundations of set theory. Amsterdam. (415 pp.).

R.O. GANDY, 1959.
Review of BOURBAKI, 1954.
Journal of Symbolic Logic, 24, pp. 71-73.

K. GÖDEL, 1940.
Continuum Hypothesis.
Annals of Mathematical Studies, 3, Princeton (66 pp.).

K. GÖDEL, 1947.
What is Cantor's continuum problem?
American Mathematical Monthly, 54, pp. 515-525.

J.A.P. HALL, 1960.
"N. BOURBAKI"
Mathematical Gazette, 350, pp. 250-253.

HILBERT/ACKERMANN, 1938.
Principles of mathematical logic.
English Edition, Chelsea, New York, (1950).

HILBERT/BERNAYS.
Grundlagen der Mathematik.
Vol. I. 1934, Berlin, (471 pp.).

P. JOURDAIN, 1915.
(See G. Cantor).

G.B. KEENE, 1961.
Abstract sets and finite ordinals.
Pergamon Press. (105 pp.).

S.C. KLEENE, 1952.
Introduction to Metamathematics.
Amsterdam. (550 pp.).

E. MENDELSON, 1956.
The independence of a weak axiom of choice.
Journal of Symbolic Logic. 21, pp. 350-366.

J. von NEUMANN, 1923.
Zur Einfuhrung der transfiniten Zahlen.
Acta Szeged, 1, pp. 199-208.

I.M. NOVAK, 1948.
A construction for models of consistent systems.
Fundamenta Mathematicae 37, pp. 87-110 (1951).

W.V.O. QUINE, 1940.
Mathematical logic. Cambridge, Mass. (346 pp.).

R.M. ROBINSON, 1937.
The theory of classes. A modification of von NEUMANN'S system.
Journal of Symbolic Logic, 2, pp. 69-72.

ROSSER/WANG, 1950.
Non-standard models for formal logics.
Journal of Symbolic Logic, 15, pp. 113-129.

W. SIERPINSKI, 1957.
Cardinal and Ordinal numbers. Warsaw. (487 pp.).

E. SPECKER, 1957.
Zur Axiomatic der Mengenlehre (Fundierungsund Auswahlaxiom).
Zeitschrift für mathematische logic und Grundlagen der Mathematik.
3, pp. 173-210.

P. SUPPES, 1960.
Axiomatic set theory. Van Nostrand. (265 pp.).

A. TARSKI, 1924.
Sur les ensembles finis.
Fundamenta Mathematicae, 6, pp. 45-95.

E.J. THIELE, 1955.
Ein axiomatisches System der Mengenlehre nach Zermelo und Fraenkel.
Zeitschrift für mathematische Logic und Grundlagen der Mathematik.
1, pp. 173-195.

H. WANG, 1949.
On Zermelo's and von Neumann's axioms for set theory.
Proc. of the Academy of Sciences, U.S.A., 35, pp. 150-155.

H. WANG, 1954.
Mathematical interpretations of formal systems, pp. 57-84.
Amsterdam.

WANG/McNAUGHTON, 1953.
Les systèmes axiomatiques de la théorie des ensembles. Paris and
Louvain. (55 pp.).

E. ZERMELO, 1930.
Uber Grenzzahlen und Mengenbereiche.
Fundamenta Mathematicae, 16, pp. 29-47.