Bernays Project: Text No. 12

Foundations of Mathematics Vol. 1 (1934)

Paul Bernays

(Grundlagen der Mathematik, Vol. 1)

Translation by: *Ian Mueller*

Comments:

Volker Peckhaus, par. 2

---

2. Elementary number theory.—Finite inferring and its limits.

The question raised at the end of the previous paragraph, whether we couldn't found arithmetic directly by a method independent of axiomatics and make a special proof of consistency superfluous, gives us reason to recall that the method of rigorous axiomatics, especially existential inference, presupposing a fixed domain of individuals, is by no means the original procedure of mathematics.

Geometry was indeed built up axiomatically from the beginning. But Euclid's axiomatic system is intended to be contentual and intuitive. There is no abstraction from the intuitive meaning of figures in it. Moreover, the axioms are not in existential form. Euclid does not presuppose that points and lines constitute any fixed domain of individuals whatsoever. And that is why he does not formulate any existence-axioms but only construction-postulates.

An example of such a postulate is: one can join two points with a straight line; furthermore, one can draw a circle around a given point with a prescribed radius.

This methodological standpoint can only be carried out if the postulates are looked on as the expression of known facts or of immediate evidence. As is well known, the question of the range of validity of the geometrical axioms is a very awkward and controversial one; and it is indeed an essential advantage of formal axiomatics that it makes the foundation of geometry independent of deciding this question.

In the domain of arithmetic we are free of these problems which are connected with the special character of geometrical knowledge; in fact we find in this domain, as well as in the disciplines of elementary number theory and algebra, the standpoint of direct contentual deliberation carried out without axiomatic assumptions in its purest form. The mark of such a standpoint is that the deliberation is carried out in the form of *thought experiments* involving objects assumed to be *concretely given*. In number theory we are concerned with numbers assumed to be given, in algebra with expressions with letters with given number coefficients.

We wish to consider the procedure here more closely and make the principles somewhat more precise in methodological respects. In number theory we have an initial object and a process of progressing. We must determine both intuitively in a certain way. The particular determination is inessential here, but once the choice is made it must be maintained for the whole theory. We choose the numeral 1 as the initial object and the addition of 1 as the process of progressing.

The things which we obtain by applying the process of progressing, beginning from the numeral 1, for example,

$$1, 11, 1111$$

are figures of the following kind: they begin with 1 and end with 1; on every 1 which is not already the end of the figure there follows an adjoined 1. They are obtained by applying the process of progressing and, i. e., by a *construction* which concretely terminates; and this *construction* can therefore be reversed in terms of a step-by-step *decomposition*.

These figures constitute a kind of numeral; we want to use the word "*numeral*" here simply to designate *these* figures.

As is usual we imagine that a certain amount of latitude is allowed concerning the exact shape of the numerals; that is, small differences in the realization such as the shape of the 1 or its size or the distance at which the 1 is put on paper, will not be taken into consideration. What we require as essential is only that we have both in 1 and the affixing of 1 an intuitive object, which can always be recognized in an unambiguous way, and that we can always survey the discrete parts from which a numeral is constructed.

In addition to numerals we introduce further signs, signs "for communication." These signs have to be distinguished on principle from the numerals, which constitute the *objects* of number theory.

In itself a sign for communication is also a figure; and we presuppose that it can always be recognized in an unambiguous way and that small differences in its realization are irrelevant. However, within the theory itself it is not taken as an object of consideration; but it is only a means for formulating facts, assertions, and assumptions concisely and clearly.

In number theory we use the following kinds of signs for communication:

1. Small German letters to designate any indeterminate numeral;

2. the usual number-signs abbreviating definite numerals, e.g., 2 for 11, 3 for 111;

3. Signs for certain formation processes and calculating operations that we perform to get from given numerals other ones. These can be applied either to certain or to indeterminate numerals, like, e. g., in $\mathfrak{a} + 11$;

4. the sign $=$ to indicate coincidence with respect to shape, the sign $\neq$ to indicate the difference between two figures: the signs $<$, $>$ to indicate a relation of magnitude between numbers which has still to be explained.

5. Parentheses as signs for the order of processes when there is a possibility of ambiguity.

How the introduced signs are manipulated and how contentual deliberations are carried out becomes clearest if we develop number theory in its basic features to a small extent.

The first thing we determine for numerals is the relationship of magnitude. Let $\mathfrak{a}$ be a numeral different from a numeral $\mathfrak{b}$. Let us consider how this is possible. Both begin with 1, and the construction continues in the same way for both $\mathfrak{a}$ and $\mathfrak{b}$, unless one of the numerals comes to an end while the construction of the other continues. This case must occur at some time, and so the one numeral coincides with a *segment* of the other; or, in more precise terms: the construction of the one numeral coincides with an initial segment of the construction of the other.

If a numeral $\mathfrak{a}$ coincides with a segment of $\mathfrak{b}$, we say that $\mathfrak{a}$ is smaller than $\mathfrak{b}$ or that $\mathfrak{b}$ is larger than $\mathfrak{a}$; and for this we apply the designation

$$\mathfrak{a} < \mathfrak{b}, \quad \mathfrak{b} > \mathfrak{a} \ .$$

Our consideration shows that for a numeral $\mathfrak{a}$ and a numeral $\mathfrak{b}$, one of the relations

$$\mathfrak{a} = \mathfrak{b}, \quad \mathfrak{a} < \mathfrak{b}, \quad \mathfrak{b} < \mathfrak{a}$$

must always hold; and on the other hand it is obvious from the intuitive meaning that these relations exclude one another. Similarly it is an immediate consequence that if $\mathfrak{a} < \mathfrak{b}$ and $\mathfrak{b} < \mathfrak{c}$ then $\mathfrak{a} < \mathfrak{c}$ also always holds.

Addition is very closely bound up with the relationship of magnitude. If a numeral $\mathfrak{b}$ coincides with a segment of $\mathfrak{a}$, the remainder is again a numeral $\mathfrak{c}$; one gets the numeral $\mathfrak{a}$ by affixing $\mathfrak{c}$ to $\mathfrak{b}$ in such a way that the 1 with which $\mathfrak{c}$ begins is attached to the 1 with which $\mathfrak{b}$ ends in conformity with the process of continuing. This kind of concatenation of numerals we call *addition*, and we use the sign + for it.

We conclude directly from this definition of addition: if $\mathfrak{b} < \mathfrak{a}$, then from the comparison of $\mathfrak{b}$ with $\mathfrak{a}$ one gets a representation of $\mathfrak{a}$ in the form $\mathfrak{b} + \mathfrak{c}$, with $\mathfrak{c}$ again a numeral. And if one starts on the other hand with any numerals $\mathfrak{b}$, $\mathfrak{c}$, then addition produces another numeral $\mathfrak{a}$, such that

$$\mathfrak{a} = \mathfrak{b} + \mathfrak{c} \ ;$$

in this case we have

$$\mathfrak{b} < \mathfrak{a}.$$

In general

$$\mathfrak{b} < \mathfrak{b} + \mathfrak{c}$$

holds as well.

The significance of numerical equalities and inequalities such as

$$2 < 3, \quad 2 + 3 = 5$$

is clear from the above definitions. $2 < 3$ says that the numeral 11 coincides with a segment of 111;

$2 + 3 = 5$ says that the numeral 11111 results from the affixing of 111 to 11.

In both of these cases we have the representation of a correct assertion, whereas $2 + 3 = 4$ is the representation of a false assertion.

We now have to determine that the calculational laws hold for addition defined intuitively.

These laws will be here conceived as propositions about arbitrarily given numerals and understood in terms of intuitive deliberation.

The associative law, according to which if $\mathfrak{a}, \mathfrak{b}, \mathfrak{c}$ are any numerals,

$$\mathfrak{a} + (\mathfrak{b} + \mathfrak{a}) = (\mathfrak{a} + \mathfrak{b}) + \mathfrak{c} \ ,$$

is immediately inferred from the definition of addition. The commutative law, which says that

$$\mathfrak{a} + \mathfrak{b} = \mathfrak{b} + \mathfrak{a}$$

always holds, is not so directly given. We need here the method of proof by *complete induction.* We first make clear how this kind of inference is to be understood from our elementary point of view: Consider any assertion about a numeral which has an elementary intuitive content. The assertion holds for 1, and one knows that if it holds for a numeral $\mathfrak{n}$ then in every case it also holds for the numeral $\mathfrak{n}+1$. One infers that the assertion holds for every given numeral $\mathfrak{a}$.

In fact the numeral $\mathfrak{a}$ is constructed by applying the process of adjoining 1 beginning from 1. If one establishes that the assertion under consideration holds for 1 and, according to the presupposition, for every adjoining of 1 for the new numeral resulting, then, with the completion of the construction of $\mathfrak{a}$ one determines that the assertion holds for $\mathfrak{a}$.

We are then not dealing with an independent principle but with a consequence taken from the concrete construction of numerals.

Using this method of inference we can now show in the usual way that for every numeral

$$1 + \mathfrak{a} = \mathfrak{a} + 1 \ ,$$

and from this that one always has

$$\mathfrak{a} + \mathfrak{b} = \mathfrak{b} + \mathfrak{a} \ .$$

We will now sketch briefly the introduction of multiplication, division and the concept formations connected with them.

Multiplication can be defined in the following way: $\mathfrak{a}\cdot\mathfrak{b}$ means the numeral which one gets from the numeral $\mathfrak{b}$ when one replaces in the construction every 1 with a numeral $\mathfrak{a}$; thus one first constructs $\mathfrak{a}$ and affixes $\mathfrak{a}$ in every case where 1 is affixed in the formation of $\mathfrak{b}$.

The associative law for multiplication as well as the distributive law according to which

$$\mathfrak{a} \cdot (\mathfrak{b} + \mathfrak{c}) = (\mathfrak{a} \cdot \mathfrak{b}) + (\mathfrak{a} \cdot \mathfrak{c})$$

are immediate from the definition. The other distributive law according to which

$$(\mathfrak{b} + \mathfrak{c}) \cdot \mathfrak{a} = (\mathfrak{b} \cdot \mathfrak{a}) + (\mathfrak{c} \cdot \mathfrak{a})$$

is seen to hold based on the laws of addition with the help of complete induction as described above. By this method of proof one also gets the commutative law of multiplication.

In order to get at division we must first introduce some preliminary considerations. The construction of a numeral is such that the adjunction of 1 always produces a new numeral. The formation of a numeral $\mathfrak{a}$ then involves the formation of a concrete row of numerals beginning with 1 and ending with $\mathfrak{a}$ in which every numeral arises from its predecessor through the adjoining of l. One also sees immediately that this row contains except for $\mathfrak{a}$ itself only numerals which are $< \mathfrak{a}$ and that a numeral which is $< \mathfrak{a}$ must occur in this row. We call this succession of numerals "the row of numerals from 1 to $\mathfrak{a}$" for short.

Now let $\mathfrak{b}$ be a numeral different from 1 which is $< \mathfrak{a}$. $\mathfrak{b}$ has the form $1 + \mathfrak{c}$; thus

$$\mathfrak{b} \cdot \mathfrak{a} = (1 \cdot \mathfrak{a}) + (\mathfrak{c} \cdot \mathfrak{a}) = \mathfrak{a} + (\mathfrak{c} \cdot \mathfrak{a}),$$

and therefore

$$\mathfrak{a} < \mathfrak{b} \cdot \mathfrak{a}.$$

If we multiply $\mathfrak{b}$ successively with the numerals in the row from 1 to $\mathfrak{a}$, in the resulting row of numerals

$$\mathfrak{b} \cdot 1,\ \mathfrak{b} \cdot 11,\ \ldots, \mathfrak{b} \cdot \mathfrak{a}$$

the first of these is $< \mathfrak{a}$, and the last $> \mathfrak{a}$. Let us go now through this row until we first come upon a numeral $> \mathfrak{a}$; its predecessor (call it $\mathfrak{b} \cdot \mathfrak{q}$) is then either $= \mathfrak{a}$ or $< \mathfrak{a}$, while

$$\mathfrak{b} \cdot (\mathfrak{q} + 1) = (\mathfrak{b} \cdot \mathfrak{q}) + \mathfrak{b} > \mathfrak{a} \ .$$

Therefore either

$$\mathfrak{a} = \mathfrak{b} \cdot \mathfrak{q}$$

or we have a representation

$$\mathfrak{a} = (\mathfrak{b} \cdot \mathfrak{q}) + \mathfrak{r} \ ,$$

with

$$(\mathfrak{b} \cdot \mathfrak{q}) + \mathfrak{r} < (\mathfrak{b} \cdot \mathfrak{q}) + \mathfrak{b} \ ,$$

and so

$$\mathfrak{r} < \mathfrak{b} \ .$$

In the first case $\mathfrak{a}$ is "divisible by $\mathfrak{b}$" ("$\mathfrak{b}$ divides $\mathfrak{a}$"), and in the second case there is division with a remainder.

In general we say $\mathfrak{a}$ is divisible by $\mathfrak{b}$ if the numeral $\mathfrak{a}$ occurs in the row

$$\mathfrak{b} \cdot 1, \mathfrak{b} \cdot 11, \ldots, \mathfrak{b} \cdot \mathfrak{a} \ .$$

This occurs if $\mathfrak{b} = 1$ or if $\mathfrak{b} = \mathfrak{a}$ or in the first case just described.

From the definition of divisibility it follows immediately that if $\mathfrak{a}$ is divisible by $\mathfrak{b}$, the determination that it is yields a representation

$$\mathfrak{a} = \mathfrak{b} \cdot \mathfrak{q}.$$

But the converse also holds; the divisibility of $\mathfrak{a}$ by $\mathfrak{b}$ (in the defined sense) follows from an equation $\mathfrak{a} = \mathfrak{b} \cdot \mathfrak{q}$ since the numeral $\mathfrak{q}$ must belong to the row of numerals from 1 to $\mathfrak{a}$.

If $\mathfrak{a} \neq 1$ and no divisor of $\mathfrak{a}$ other than 1 and $\mathfrak{a}$ occurs in the row of numerals from 1 to $\mathfrak{a}$, then every product $\mathfrak{m} \cdot \mathfrak{n}$ in which $\mathfrak{m}$ and $\mathfrak{n}$ belong to the row of numerals from 2 to $\mathfrak{a}$ is distinct from $\mathfrak{a}$; in such a case we call $\mathfrak{a}$ a *prime number.*

If $\mathfrak{n}$ is a numeral different from 1, then there is a first numeral in the row from 1 to $\mathfrak{n}$ with the property of being distinct from 1 and a divisor of $\mathfrak{n}$. It is easy to show that this "least divisor of $\mathfrak{n}$ distinct from 1" is a prime number.

Now we can also prove in the same way as EUCLID did the theorem that for any numeral $\mathfrak{a}$, a prime number $> \mathfrak{a}$ can be determined: One multiplies together the numbers from the row from 1 to $\mathfrak{a}$, adds 1, and then takes the least divisor $\mathfrak{t}$ distinct from 1 of the numeral thus obtained. This is a prime number, and one sees easily that $\mathfrak{t}$ cannot occur in the row of numbers from 1 to $\mathfrak{a}$ and so is $> \mathfrak{a}$.

The further development of elementary number theory is clear; only one point still requires fundamental discussion, the method of *recursive definition.* We recall the nature of this method: A new function-sign, say, φ, is introduced, and the definition of the function takes place through two equations which in the simplest case have the form

$$\begin{aligned} \varphi(1) &= \mathfrak{a} \\ \varphi(\mathfrak{n}+1) &= \psi(\varphi(\mathfrak{n}), \mathfrak{n}) \,. \end{aligned}$$

Here $\mathfrak{a}$ is a numeral and ψ is a function constructed from functions already known so that $\psi(\mathfrak{b}, \mathfrak{c})$ can be calculated for given numerals $\mathfrak{b}$, $\mathfrak{c}$ and again provides a numeral as a value.

For example the function

$$\varrho(\mathfrak{n}) = 1 \cdot 2 \ldots \mathfrak{n}$$

can be defined by the equations

$$\begin{aligned} \varrho(1) &= 1 \\ \varrho(\mathfrak{n}+1) &= \varrho(\mathfrak{n}) \cdot (\mathfrak{n}+1). \end{aligned}$$

What sense this method of definition has is not self-evident. To explain it we must first make precise the concept of function. We understand a *function* to be an intuitive instruction on the basis of which a numeral can be assigned to a given numeral, or a pair, a triple ... of numerals. A pair of equations of the above kind—we call it a "recursion"—, we interpret as an *abbreviated indication* of the following instruction:

Let $\mathfrak{m}$ be some numeral. If $\mathfrak{m} = 1$, then $\mathfrak{m}$ is to be paired with the numeral $\mathfrak{a}$. Otherwise $\mathfrak{m}$ has the form $\mathfrak{b} + 1$. First one writes schematically

$$\psi(\varphi(\mathfrak{b}), \mathfrak{b}) \ .$$

If $\mathfrak{b} = 1$, then one replaces $\varphi(\mathfrak{b})$ here with $\mathfrak{a}$; otherwise $\mathfrak{b}$ has the form $\mathfrak{c} + 1$ and one replaces $\varphi(\mathfrak{b})$ with

$$\psi(\varphi(\mathfrak{c}), \mathfrak{c}).$$

Again, either $\mathfrak{c} = 1$ or $\mathfrak{c}$ is of the form $\mathfrak{d} + 1$. In the first case one replaces $\varphi(\mathfrak{c})$ with $\mathfrak{a}$ and in the second case with

$$\psi(\varphi(\mathfrak{d}), \mathfrak{d}) \ .$$

The continuation of this procedure terminates in every case. For the numerals

$$\mathfrak{b}, \ \mathfrak{c}, \ \mathfrak{d}, \ldots \ ,$$

which we obtain in sequence arise from the *decomposition of the numeral* $\mathfrak{m}$; and this, like the construction of $\mathfrak{m}$, must terminate. When the decomposition reaches 1, $\varphi(1)$ is replaced with $\mathfrak{a}$; the resulting configuration no longer contains the sign φ; rather only ψ occurs as a function sign, perhaps with multiple overlapping occurrences, and the innermost arguments are numerals. We have then obtained a calculable expression; for ψ is supposed to be an already known function. This calculation has to be done from the inside out; the numeral obtained is the numeral to be paired with the numeral $\mathfrak{m}$.

The content of this instruction shows us first that in principle it can be carried out for any case for a given numeral $\mathfrak{m}$ and that the result is unambiguously fixed. At the same time we see that for a given numeral $\mathfrak{n}$ the equation

$$\varphi(\mathfrak{n}+1) = \psi(\varphi(\mathfrak{n}), \mathfrak{n})$$

is satisfied, if we replace in it $\varphi(\mathfrak{n})$ and $\varphi(\mathfrak{n}+1)$ with the numerals paired with $\mathfrak{n}$, and $\mathfrak{n}+1$ according to our instruction and then substitute for the known function ψ its definition.

The somewhat more general case in which one or more undetermined numerals occur as "*parameters*" in the function being defined is handled in much the same way. In the case where there is one parameter $\mathfrak{t}$ the recursion equations have the form

$$\begin{aligned} \varphi(\mathfrak{t}, 1) &= \alpha(\mathfrak{t}) \\ \varphi(\mathfrak{t}, \mathfrak{n}+1) &= \psi(\varphi(\mathfrak{t}, \mathfrak{n}), \mathfrak{t}, \mathfrak{n}); \end{aligned}$$

here both α and ψ are known functions. For example, the function $\varphi(\mathfrak{t}, \mathfrak{n}) =$

$\mathfrak{t}^{\mathfrak{n}}$ is defined by the recursion

$$\begin{aligned}\varphi(\mathfrak{t}, 1) &= \mathfrak{t} \\ \varphi(\mathfrak{t}, \mathfrak{n}+1) &= \varphi(\mathfrak{t}, \mathfrak{n}) \cdot \mathfrak{t} \ .\end{aligned}$$

With definition by recursion one is again not dealing with an independent principle of definition; within the framework of elementary number theory, recursion has only the meaning of a convention for abbreviating the description of certain formation-processes through which one gets from one or more given numerals another numeral.—

We take as an example to indicate that we can carry out *proofs of impossibility* in the framework of intuitive number theory the assertion expressing the irrationality of $\sqrt{2}$: There cannot be two numerals $\mathfrak{m}$, $\mathfrak{n}$ such that[1]

$$\mathfrak{m} \cdot \mathfrak{m} = 2 \cdot \mathfrak{n} \cdot \mathfrak{n} \ .$$

As is well known, the proof proceeds as follows: One shows first that every numeral is either divisible by 2 or of the form $(2 \cdot \mathfrak{k}) + 1$; therefore $\mathfrak{a} \cdot \mathfrak{a}$ is divisible by 2 only if $\mathfrak{a}$ is divisible by 2.

If a pair of numbers $\mathfrak{m}$, $\mathfrak{n}$ satisfying the above equation were given, we could examine all number-pairs $\mathfrak{a}$, $\mathfrak{b}$ with

$$\begin{gathered}\mathfrak{a} \text{ belonging to the row } 1, \ldots, \mathfrak{m} \ , \\ \mathfrak{b} \text{ belonging to the row } 1, \ldots, \mathfrak{n} \ ,\end{gathered}$$

and then determine whether or not

$$\mathfrak{a} \cdot \mathfrak{a} = 2 \cdot \mathfrak{b} \cdot \mathfrak{b} \ .$$

[1] We use the customary procedure following the associative law of multiplication in writing products involving several factors without parentheses.

We choose from among the pairs of values satisfying the equation the one in which $\mathfrak{b}$ has the smallest value. There can only be *one* such; call it $\mathfrak{m}'$, $\mathfrak{n}'$. In accordance with our previous remarks, it follows from the equation

$$\mathfrak{m}' \cdot \mathfrak{m}' = 2 \cdot \mathfrak{n}' \cdot \mathfrak{n}'$$

that $\mathfrak{m}'$ is divisible by 2:

$$\mathfrak{m}' = 2 \cdot \mathfrak{k}' \ ;$$

therefore we obtain

$$\begin{aligned} 2 \cdot \mathfrak{k}' \cdot 2 \cdot \mathfrak{k}' &= 2 \cdot \mathfrak{n}' \cdot \mathfrak{n}' \ , \\ 2 \cdot \mathfrak{k}' \cdot \mathfrak{k}' &= \mathfrak{n}' \cdot \mathfrak{n}' \ . \end{aligned}$$

But then $\mathfrak{n}'$, $\mathfrak{k}'$, would be a pair of numbers satisfying our equation, and at the same time it would be the case that $\mathfrak{k}' < \mathfrak{n}'$. This however is inconsistent with the way in which $\mathfrak{n}'$ was determined.

Of course, the sentence just proved can be expressed positively: If $\mathfrak{m}$ and $\mathfrak{n}$ are any two numerals, $\mathfrak{m} \cdot \mathfrak{m}$ is different from $2 \cdot \mathfrak{n} \cdot \mathfrak{n}$.

Let this much suffice as a characterization of the elementary treatment of number theory. We have developed it as a theory of numerals, i.e., of a certain kind of especially simple figure. The significance of this theory for knowledge depends upon the relation of numerals to the ordinary *concept of cardinal number* [*Anzahl-Begriff*]. We obtain this relation in the following way:

Imagine a concrete (i.e., in any case finite) collection of things. One considers the things in the collection successively and correlates them in a row with the numerals 1, 11, 111, ... as numbers. When no thing is left a certain numeral $\mathfrak{n}$ has been reached. This numeral is for the time being assigned as the *ordinal number* of the collection of things taken in the sequence chosen.

But now we easily convince ourselves that the resulting numeral $\mathfrak{n}$ is in no way dependent on what sequence is chosen. For let

$$a_1, a_2, \ldots, a_{\mathfrak{n}}$$

be the things of the collection in the sequence chosen,

$$b_1, b_2, \ldots, b_{\mathfrak{k}}$$

the things in some other sequence. We can then go from the first enumeration to the second by a series of interchanges of numbers in the following manner: If a_1, is different from b_1, then we interchange the number $\mathfrak{r}$ which the thing b_1 has in the first enumeration with 1, that is to say we correlate the thing $\mathfrak{a}_{\mathfrak{r}}$ with the number 1, the thing a_1 with the number $\mathfrak{r}$. In the resulting enumeration the thing b_1, has the number 1; following it and correlated to the number 2 is the thing b_2 unless this thing has here some other number $\mathfrak{s}$ which is in any case distinct from 1; in this case we then exchange this number $\mathfrak{s}$ with 2 in the enumeration; the result is an enumeration in which the thing b_1 has the number 1, b_2 the number 2. b_3 has either the number 3 or some other, t, in any case distinct from 1 and 2; we exchange the latter with 3.

This procedure must terminate; for with every interchange the enumeration of the collection considered is brought at least one step closer to a correspondence with the enumeration

$$b_1, b_2, \ldots, b_{\mathfrak{k}} \ ;$$

as a result one will eventually get the number 1 for b_1, the number 2 for b_2, ..., the number $\mathfrak{k}$ for $b_{\mathfrak{k}}$; and then there is no other thing left. On the other hand the stock of numbers used remains exactly the same with every interchange executed; for all that happens is that the number of one thing is exchanged with that of another. Therefore in every case the enumeration goes from 1 to $\mathfrak{n}$, and as a result we have

$$\mathfrak{k} = \mathfrak{n} \ .$$

Thus the numeral $\mathfrak{n}$ is assigned to the collection under consideration independently of any sequence; in this sense we can correlate $\mathfrak{n}$ with the collection as its *cardinal number* [*Anzahl*].[2] We say that the collection consists of $\mathfrak{n}$ things.

If a concrete collection has a common cardinal number with another one, we get from an enumeration of each a one-one correlation of the things in one collection with those in the other. On the other hand, if we have such a correlation between two given collections of things the two have the same cardinal number; this is an immediate consequence of our definition of cardinal number.

From the definition of cardinal number we pass now through contentual considerations to the principles of the *theory of cardinal numbers*, e.g., to the theorem that the unification of two collections of cardinality $\mathfrak{a}$ and $\mathfrak{b}$ and without a common element gives rise to a collection of $\mathfrak{a} + \mathfrak{b}$ things.—

After the presentation of elementary number theory we would like to indicate briefly the character of the elementary contentual point of view in *algebra*. We shall deal with the elementary theory of rational functions of one or more variables with integers as coefficients.

[2]This consideration was put forward by von Helmholtz in his essay "Zählen und Messen" (1887). (Hermann v. Helmholtz, Schriften zur Erkenntnistheorie. Berlin: Julius Springer 1921. See pp. 80–82.)

The objects of the theory are again certain figures, "polynomials;" they are constructed from a determinate stock of letters, $x, y, z, \ldots$, called "variables" and the numerals with the help of the signs $+, -, \cdot$ and parentheses. In this case then the signs $+, \cdot$ are not to be construed as communication-signs as in elementary number theory, but as objects of the theory.

We again use small German letters as communication-signs, not just for numerals but also for arbitrary polynomials.

The construction of polynomials out of the signs indicated above follows the following formation rules:

A variable and also a numeral can be considered in itself a polynomial.

From two polynomials $\mathfrak{a}$, $\mathfrak{b}$ the polynomials

$$\mathfrak{a} + \mathfrak{b}, \quad \mathfrak{a} - \mathfrak{b}, \quad \mathfrak{a} \cdot \mathfrak{b}$$

can be formed; from a polynomial $\mathfrak{a}$ $(-\mathfrak{a})$ can be formed. The usual rules for setting parentheses hold here. As communication-signs we also introduce:

the numbers 2, 3, ..., as in elementary number theory;

the sign 0 for $1 - 1$;

the usual signs for powers: for example, if $\mathfrak{z}$ is a numeral $x^{\mathfrak{z}}$ signifies the polynomial which results from $\mathfrak{z}$ when x is put in place of every 1 and the sign "$\cdot$" is put between every two successive x's;

the sign = indicates the mutual *substitutability* of two polynomials.

Substitutability is determined by the following contentual rules:

1. The associative and commutative laws for "+" and "$\cdot$".

2. The distributive law

$$\mathfrak{a} \cdot (\mathfrak{b} + \mathfrak{c}) = (\mathfrak{a} \cdot \mathfrak{b}) + (\mathfrak{a} \cdot \mathfrak{c}) \ .$$

3. Rules for "$-$":

$$\begin{aligned} \mathfrak{a} - \mathfrak{b} &= \mathfrak{a} + (-\mathfrak{b}) \ , \\ (\mathfrak{a} + \mathfrak{b}) - \mathfrak{b} &= \mathfrak{a} \ . \end{aligned}$$

4. $1 \cdot \mathfrak{a} = \mathfrak{a}$.

5. If two polynomials $\mathfrak{m}$, $\mathfrak{n}$ are free of variables and of $-$, and if the equality $\mathfrak{m} = \mathfrak{n}$ holds *in the sense of the interpretation of elementary number theory*, then $\mathfrak{m}$ can be substituted by $\mathfrak{n}$.

These rules of substitutability relate to polynomials occurring as *parts* of other polynomials. From them further assertions about substitutability are derived, which constitute the "identities" and theorems of elementary algebra. As examples some of the most simple provable identities we mention:

$$\begin{aligned} \mathfrak{a} + 0 &= \mathfrak{a} & -(\mathfrak{a} - \mathfrak{b}) &= \mathfrak{b} - \mathfrak{a} \ , \\ \mathfrak{a} - \mathfrak{a} &= 0 & -(-\mathfrak{a}) &= \mathfrak{a} \ , \\ \mathfrak{a} \cdot 0 &= 0 & (-\mathfrak{a}) \cdot (-\mathfrak{b}) &= \mathfrak{a} \cdot \mathfrak{b} \ . \end{aligned}$$

Of the theorems which can be established from contentual considerations we mention the following fundamental assertions:

a) If $\mathfrak{a}$, $\mathfrak{b}$ are two mutually substitutable polynomials of which at least one contains the variable x and if the polynomials $\mathfrak{a}_1$, $\mathfrak{b}_1$ result from $\mathfrak{a}$, $\mathfrak{b}$ when the variable x is replaced at all places where it occurs with one and the same polynomial $\mathfrak{c}$, then $\mathfrak{b}_1$ is substitutable for $\mathfrak{a}_1$.

b) Substitution of numerals for variables in a correct equation between polynomials yields a correct numerical equation in the sense of number theory (if we suppose that calculation with negative numbers is introduced into number theory).—The meaning of this assertion b) may be illustrated by a simple example: The equation

$$(x + y) \cdot (x + y) = x^2 + 2 \cdot x \cdot y + y^2$$

says for the time being nothing except that according to our determinations $x^2 + 2 \cdot x \cdot y + y^2$ is substitutable for $(x + y) \cdot (x + y)$. On the basis of the assertion b), however, we can infer from this that if $\mathfrak{m}$ and $\mathfrak{n}$ are number signs $(\mathfrak{m} + \mathfrak{n}) \cdot (\mathfrak{m} + \mathfrak{n})$ and $\mathfrak{m} \cdot \mathfrak{m} + 2 \cdot \mathfrak{m} \cdot \mathfrak{n} + \mathfrak{n} \cdot \mathfrak{n}$ coincide in the number-theoretic sense.

c) Every polynomial can be substituted for either by 0 or by a sum of different powers of products of variables—(the polynomial 1, as well, counts as such in this case)—, each conjoined with a positive or negative numerical factor.

Using this normal form we obtain a procedure for deciding for two given polynomials whether or not they are mutually substitutable. For the following assertion holds:

d) 0 is not substitutable for a polynomial which is the sum of different products of powers with numerical factors, and two such polynomials are only mutually substitutable if they coincide in their products of powers and their numerical factors when the order of the summands and the order of the numerical factors are disregarded.

The second part of this assertion follows from the first; and the first can be proved with the help of b) by considering suitable substitutions of numerals.

The following is a particular consequence of d):

e) if a numeral $\mathfrak{n}$ is substitutable for a numeral $\mathfrak{m}$, conceived as a polynomial, $\mathfrak{m}$ and $\mathfrak{n}$ coincide.

In methodological respects the following should be noted concerning these assertions: The substitutability of polynomials assumed in a), e) is to be understood as the assumption that the substitutability of the one polynomial by the other has been determined according to the rules. In sentence c) the assertion of substitutability is more closely determined by giving a procedure being described in the proof of the theorem.

And so we are here just as much as in elementary number theory in the domain of elementary contentual inferences. And the same is true of the other assertions and proofs of elementary algebra.—

The consideration of the elementary foundations of number theory and algebra just presented has served to show us the application and use of direct contentual inference carried out in thought experiments performed on intuitively imagined objects and free from axiomatic assumptions. We will call this kind of inference "*finite*" inference in order to have a short expression; likewise we shall call the attitude underlying this kind of inference the "finite" attitude or the "finite" point of view. We will speak of finite concept formations and assertions in the same sense; in using the word "finite" we convey the idea that the consideration, assertion or definition in question remains within the limits of objects that it is in principle possible to imagine and of processes that it is in principle possible to realize; thus that it is carried out in the framework of concrete consideration.

To characterize the finite point of view further we stress certain general aspects relating to the use of the logical forms of judgment in finite thinking; we consider assertions about *numerals* as examples.

A *universal* judgment about numerals can only finitely be interpreted in a hypothetical sense; that is to say, as an assertion about any given numeral. Such a judgment expresses a law which must be verified in every particular case.

An *existential sentence* about numerals, i.e., a sentence of the form "there is a numeral $\mathfrak{n}$ with the property $\mathfrak{A}(\mathfrak{n})$" is to be construed finitely as a "partial judgment," that is, an incomplete communication of a more precisely determined assertion, which consists either in a direct indication of a numeral with the property $\mathfrak{A}(\mathfrak{n})$ or in the indication of a procedure to obtain such a numeral — it belongs to the indication of a procedure that the series of operations to be executed has a definite bound.

Judgments in which a universal assertion is conjoined with an existential one are to be interpreted in a corresponding way. So, for example, a sentence of the form "for every numeral $\mathfrak{k}$ with the property $\mathfrak{A}(\mathfrak{k})$ there is a numeral $\mathfrak{l}$ such that $\mathfrak{B}(\mathfrak{k}, \mathfrak{l})$" is constructed finitely as an incomplete communication of a procedure which for any given numeral $\mathfrak{k}$ with the property $\mathfrak{A}(\mathfrak{k})$ makes possible the finding of a numeral $\mathfrak{l}$ which stands to $\mathfrak{k}$ in the relation $\mathfrak{B}(\mathfrak{k}, \mathfrak{l})$.

The application of *negation* demands special attention.

Negation is unproblematic in the case of "elementary" judgments, which involve a question decidable by a direct intuitive determination (a "finding"). For example, if $\mathfrak{k}$, $\mathfrak{l}$ are particular numerals it can be directly determined whether or not

$$\mathfrak{k} + \mathfrak{k} = \mathfrak{l},$$

i.e., whether or not $\mathfrak{k} + \mathfrak{k}$ coincides with $\mathfrak{l}$ or is different from $\mathfrak{l}$.

The negation of such an elementary judgment says simply that the result of the respective intuitive decision diverges from the state of affairs asserted to obtain in the judgment; and without hesitation there is for an elementary judgment the alternative that either it or its negation is correct.

By contrast it is not immediately clear what should count as the negation of a universal or existential judgment in the finite sense.

As a result, we consider first existential assertions. The assertion that there is no numeral $\mathfrak{n}$ with the property $\mathfrak{A}(\mathfrak{n})$ might mean in an imprecise sense that one has no numeral with this property at one's disposal for indication. But such an assertion has no objective meaning because of its connection to an accidental epistemological condition. If, however, one wishes to maintain the unavailability of a numeral $\mathfrak{n}$ with the property $\mathfrak{A}(\mathfrak{n})$ independently of epistemological conditions, he can do it in a finite sense only with an assertion of impossibility, which says that a numeral $\mathfrak{n}$ *can* not have the property $\mathfrak{A}(\mathfrak{n})$.

In this way we arrive at a *sharpened* negation; however, it is not exactly the contradictory opposite of an existential assertion, "there is a numeral $\mathfrak{n}$ with the property $\mathfrak{A}(\mathfrak{n})$", which (as a partial judgment) points to a known numeral with this property or to a procedure which we possess for obtaining such a numeral.

Unlike an elementary assertion and its negation, an existential assertion and its sharpened negation are not assertions about the only two possible results of *one and the same decision*, but they correspond to two distinct epistemological possibilities, namely on the one hand the finding of a numeral with a given property, and on the other the discernment of a general law about numerals.

It is not logically obvious that one of these two possibilities must come up. From the finite point of view then we cannot use the alternative, that there is either a numeral $\mathfrak{n}$ such that $\mathfrak{A}(\mathfrak{n})$ or the holding of $\mathfrak{A}(\mathfrak{n})$ for a numeral is precluded.

The case of a universal judgment of the form "for every numeral $\mathfrak{n}$, $\mathfrak{A}(\mathfrak{n})$ holds" with respect to finite negation is similar to that of existential judgements. The negation of the validity of such a judgment does not have a direct finite sense; however, if it is sharpened as the assertion that the logical validity of $\mathfrak{A}(\mathfrak{n})$ can be refuted by a counterexample, then this sharpened negation no longer constitutes the contradictory opposite of a universal judgment; for again it is not logically obvious that either a universal judgment or the sharpened negation must hold, i.e., that either $\mathfrak{A}(\mathfrak{n})$ holds for every given numeral $\mathfrak{n}$ or that a numeral can be given for which $\mathfrak{A}(\mathfrak{n})$ does not hold.

It must be added that the finding of a counterexample is not the only possibility for refuting a universal judgment. Pursuing the consequences of a universal judgment can lead to a contradiction in other ways. This circumstance, however, does not eliminate the difficulty but only increases the complication. Namely, neither the alternative is logically obvious that a universal judgment about numerals must either hold or lead to a contradiction in its consequences and therefore be refutable, nor it is obvious that such a judgment if refutable is refutable through a counterexample.

The complicated situation that we find here with respect to the negation of judgments from the finite point of view corresponds to BROUWER's thesis that the law of the excluded middle does not hold for infinite totalities. This invalidity exists indeed from the finite point of view insofar as we are unable to find a negation with finite content which satisfies the law of the excluded middle for existential and universal judgments.

These considerations may suffice as an indication of the finite point of view. If we look at arithmetic in its customary treatment, checking whether it corresponds to this methodological point of view, we realize that this is not the case; arithmetic inferences and concept formations often go beyond the limits of the finite way of reflection in many ways.

The inferences of number theory already go beyond the finite point of view; for here existential assertions about integers—in ordinary mathematics we speak of "integers" (more exactly "positive integers" or "numbers" for short) instead of "numerals"—are permitted, regardless of the possibility of actually determining the number in question; and also use is made of the alternative that an assertion about integers either holds for all integers or there is a number for which it does not hold.

This alternative, the "tertium non datur" for integers is implicitly applied in "least number principle" which says: "If an assertion about integers holds for at least one number, there is a least number for which it holds."

In its *elementary* applications the least number principle has a finite character. Indeed, if $\mathfrak{A}(n)$ is the respective assertion about a number n and $\mathfrak{m}$ is a definite number for which $\mathfrak{A}(\mathfrak{m})$ holds, then one may go through the numbers from 1 to $\mathfrak{m}$; then one must once come to a number $\mathfrak{k}$ for which $\mathfrak{A}(\mathfrak{k})$ is correct since $\mathfrak{m}$ is such a number at the latest. The number $\mathfrak{k}$ is then the least number with the property $\mathfrak{A}$.

But these considerations depend on two presuppositions which are not always fulfilled in non-elementary applications of the least number principle. In the first place it is presupposed that $\mathfrak{A}$ holds of a number in the sense that a number $\mathfrak{m}$ with the property $\mathfrak{A}(\mathfrak{m})$ is actually given, while in applications the existence of a number with the property $\mathfrak{A}$ is often only derived using the "tertium non datur" so that we do not get to the actual determination of such a number. The second presupposition is that it can be decided for any number $\mathfrak{k}$ in the row from 1 to $\mathfrak{m}$ whether or not $\mathfrak{A}(\mathfrak{k})$ holds; of course, it is possible to decide this for elementary assertions $\mathfrak{A}(n)$, whilst for a non-elementary expression $\mathfrak{A}(n)$ the question whether it holds for a given number $\mathfrak{k}$ may constitute an unsolved problem.

For example, let $\psi(a)$ be a function defined by a sequence of recursions and substitutions, and so admissible in finite number theory; and let $\mathfrak{A}(n)$ stand for the assertion that there is a number a for which $\psi(a) = n$. Then, for a given number $\mathfrak{k}$, the question whether $\mathfrak{A}(\mathfrak{k})$ holds is not in general (i.e., when the function ψ is not especially simple) decidable by direct inspection; rather it has the character of a mathematical problem. For the recursions which enter into the definition of ψ give the values of the function only *for given arguments*; but the question whether there is a number a for which $\psi(a)$ has the value $\mathfrak{k}$ involves the whole value-range of the function ψ.

In any case then where these presuppositions of the finite founding of the least number principle are not fulfilled, the founding of the principle requires reference to the "tertium non datur" for integers.[3]

[3]We will later present the proof of the principle of least number in the framework of the formalism. See § 6, pp. 284–285.

We give some examples of number-theoretic alternatives which result from the tertium non datur for integers, but which are not provable in a finite way, given our present state of knowledge:

"Either every even number > 2 is representable as the sum of two prime numbers or there is an even number > 2 and not representable as the sum of two prime numbers."

"Either every integer of the form $2^{(2^{\mathfrak{k}})} + 1$ with $\mathfrak{k} > 4$ is divisible into two factors > 1, or there is a prime number of the form $2^{(2^{\mathfrak{k}})} + 1$ with $\mathfrak{k} > 4$."

"Either every sufficiently great integer is representable as the sum of less than 8 cubes, or for every integer n there is an integer m greater than n which is not representable as the sum of less than 8 cubes."

"Either there are arbitrarily great prime numbers p with the property that $p + 2$ is also a prime number, or there is a greatest prime number with this property."

"Either for every integer $n > 2$ and arbitrary positive integers a, b, c the inequality $a^n + b^n \neq c^n$ holds, or there is a least such integer $n > 2$ for which the equation $a^n + b^n = c^n$ has a solution in the positive integers."

This kind of example from number theory is appropriate for making clear the simplest forms of non-finite argumentation. However, we will not really feel the need to go beyond the finite point of view in number theory; for there is hardly any proof using number-theoretic means in which the non-finite kinds of inference that happen to be made can not be circumvented with rather simple modifications.

It is quite different in the case of analysis (infinitesimal calculus); here non-finite kinds of concept formation and proof really belong to the methods of the theory.

We briefly recall the fundamental concept of analysis, the concept of a real number. The real numbers are defined either as a strictly increasing sequence of rational numbers

$$r_1 < r_2 < r_3 < \dots,$$

which are all less than a given bound ("fundamental sequence"), or as an infinite decimal fraction, or binary fraction, or as a partition of the rational numbers into two classes, every member of the first class being smaller than every member of the second ("DEDEKIND cut").

In doing so the view is fundamental that rational numbers form an enclosed totality which can be considered as a *domain of individuals*. In analysis the totality of possible sequences of rational numbers or of possible partitions of all rational numbers is also conceived of as a domain of individuals.

However, it is sufficient to take as a basis the totality of integers instead of the totality of rational numbers and the totality of all partitions of integers instead of that of all rational numbers. For in fact every positive rational number is given by a pair of numbers m, n and every rational number whatever can be represented as the difference between two positive rational numbers, i.e., as a pair of pairs of numbers $(m, n; p, q)$. Also every binary fraction of the form

$$0.a_1a_2a_3\dots$$

with $a_1, a_2, a_3, \dots$ are all either $= 0$ or $= 1$ can be interpreted as a partition of all integers, namely the partition into those numbers k for which $a_k = 0$ and into those for which $a_k = 1$. In this way there is a one-one correspondence between the partitions of the positive integers and the binary fractions of the above form; and on the other hand every real number can be represented as the sum of a positive integer and a binary fraction of this form.

It is possible to consider *sets* of integers instead of partitions; for every set of integers determines the partition of the numbers that belong to the set and those that do not; and equally every set of integers is completely determined by such a partition.—The same remark holds for the Dedekind cut; it likewise can be represented by a *set* of rational numbers, namely the set containing the smaller rational numbers. Such a set is characterized by the following properties: 1. it contains at least one and not every rational number; 2. together with a rational number it contains all smaller and at least one bigger rational number.

By these transformations, however, the existential presupposition which we had to take as the basis of analysis is weakened in an unessential way only. It is still required to construe the manifold of the integers and also that of the sets of integers as a fixed domain of individuals; the "tertium non datur" is taken to hold for this domain, and with reference to which an assertion of the existence of an integer or set of integers with a property $\mathfrak{E}$ is taken to be meaningful independently of its possible interpretation as a partial judgment. So even though the infinitely large and the infinitely small in any genuine sense are excluded by this theory of real numbers and remain only as a mode of expression, still the *infinite as a totality* is retained. One can even say that here for the first time the idea of infinite totalities is introduced and validated in the rigorous foundation of analysis.

In order to to convince ourselves really that the presupposition of the totality of the domain of integers or rational numbers and moreover of the domain of sets (partitions) of integers or rational numbers has an essential application in the founding of analysis we only need to introduce some of the fundamental concept formations and thoughts.

If the reals are defined as a sequence of increasing rational numbers

$$r_1 < r_2 < r_3 < \dots ,$$

the concept of equality for real numbers is already non-finite. For whether or not two such sequences of rational numbers define the same real number depends upon whether or not for every number in one of the sequences there is a larger in the other and vice-versa. But we do not have a general procedure for deciding this question.

If, however, we begin with the definition of the real number via a DEDEKIND cut, we have to prove that every bounded sequence of increasing rational numbers gives rise to a cut representing the upper bound of the sequence. One gets this cut by partitioning the rational numbers into those which are less than at least one number of the sequence and those which are not. That is to say: a rational number r is said to be in the first or second class according to whether there is among the numbers of the sequence a number $> r$ or whether all numbers in the sequence are $\leqq r$. This again is no finite distinction.

The case is similar if real numbers are defined via infinite decimal or binary fractions. Again it must be shown that a bounded sequence of rational numbers

$$r_1 < r_2 < r_3 < \dots ,$$

determines a decimal or binary fraction. For simplicity let us suppose we are dealing with a sequence of positive proper fractions:

$$0 < b_1 < b_2 < \ldots < 1 \ ,$$

and we wish to determine the binary fraction

$$0.a_1 a_2 a_3 \ldots$$

which represents the upper bound of the sequence of fractions. This is done as follows:

a_1 is $= 0$ or $= 1$ depending on whether or not for all fractions $b_n < \frac{1}{2}$ holds;

$a_{m+1} = 0$ or $= 1$ depending on whether or not all fractions b_n are less than

$$\frac{a_1}{2} + \frac{a_2}{4} + \ldots + \frac{a_m}{2^m} + \frac{1}{2^{m+1}} \ .$$

In each of these cases one considers the alternatives according to whether all rational numbers in a given sequence

$$r_1, r_2, r_3, \ldots$$

satisfy a certain inequality or whether there is at least one exception to this inequality. Such an alternative makes use of the "tertium non datur" for integers; for it is presupposed that either for every integer n the rational number r_n satisfies the inequality in question, or there is at least one integer n such that r_n fails to satisfy it.

However, this use of the *totality of integers* as a domain of individuals is not sufficient for analysis; we need in addition the *totality of real numbers* as domain of individuals. As we saw, this totality is essentially equivalent to that of the sets of integers.

The need for domain of individuals of real numbers is already necessary in connection with the proof of the theorem of the upper bound of a bounded set of real numbers. In order to prove the existence of the upper bound of a bounded set of real numbers, e.g., reals in the interval between 0 and 1, defined on the basis of DEDEKIND cuts, one considers the partition of the rational numbers into those which are and those which are not exceeded by a real number in the set. Thus one counts a rational number r as being in the first class if and only if there is a real number $a > r$ in the set.

Now one has to realize that in analysis a set is in general given to us only by a defining property; that is to say, the set is introduced as the totality of those real numbers which satisfy a certain condition $\mathfrak{B}$. Therefore the question whether there is a real number $a > r$ in a set under consideration amounts to the question whether there is a real number greater than r and at the same time satisfying a certain condition $\mathfrak{B}$. In this formulation it becomes clear that we take the totality of real numbers as a domain of individuals as a basis.[4]

It should also be remarked that the process just described for obtaining an upper bound amounts to forming a *union set*. In fact every real number is defined by a partition of the rational numbers into larger and smaller ones or by the set of the smaller rational numbers. The given set of real numbers is therefore represented as a set $\mathfrak{M}$ of sets of rational numbers. And the

[4]WEYL has pointed to the state of affairs given here in a particular explicit way in his monograph "Das Kontinuum" (Leipzig 1918).

upper bound of the set $\mathfrak{M}$ is formed from the set of those rational numbers which belong to at least one of the sets in $\mathfrak{M}$. The totality of these rational numbers is, however, exactly the union set of $\mathfrak{M}$.

Defining the real numbers by means of a fundamental sequence or a binary fraction instead of using DEDEKIND's definition does not make it possible to circumvent the the use of the domain of individuals of real numbers. By this the process will rather become even more complicated because an additional recursive procedure enters. It may briefly be indicated what is involved in the case of the definition of real numbers via binary fractions. We are then concerned with a set of binary fractions

$$0.a_1a_2a_3\dots$$

which is again determined by a certain criterion $\mathfrak{B}$; and the upper bound is represented by a binary fraction

$$0.b_1b_2\dots$$

defined in the following way:

$b_1 = 0$, if 0 stands in the first binary position in all binary fractions satisfying the condition $\mathfrak{B}$; otherwise $b_1 = 1$;

$b_{n+1} = 0$, if 0 stands in the $(n+1)$st position in all binary fractions satisfying the condition $\mathfrak{B}$ and having the first n binary numerals coincident or with $b_1, b_2, \dots, b_n$ respectively; otherwise $b_{n+1} = 1$.

Here the totality of the real numbers occurs as the totality of all binary fractions; and we make use of the assumption that the "tertium non datur" holds for infinite sequences formed of zeros and ones.—

Now this presupposition of the totality of all real numbers or all binary fractions, however, is not sufficient. This can be seen in the following simple case: Let a be the upper bound of a set of real numbers. We want to show that there is a sequence of real numbers *from the set* which converges toward a. To do this we make the following inferences:

It follows from the property of an upper bound that for every integer n there is a number c_n in the set such that

$$a - \frac{1}{n} < c_n \leqq a \ ,$$

and so

$$|a - c_n| < \frac{1}{n} \ .$$

The numbers c_n constitute therefore a sequence which converges toward a, and they all belong to the set under consideration.

When we argue in this way our manner of expression hides a fundamental point in the proof. For when we use the notation c_n we presuppose that for each number n among those real numbers c belonging to the set under consideration and satisfying the inequality

$$a - \frac{1}{n} < c \leqq a,$$

a certain one is distinguished.

There is a presupposition involved here. All we can immediately infer is only this: for every number n there is a subset $\mathfrak{M}_n$ of our set under consideration which consists of those numbers satisfying the above inequality; and for every n this subset has at least one element. Now what is assumed is that in each of these sets

$$\mathfrak{M}_1, \mathfrak{M}_2, \mathfrak{M}_3, \ldots$$

we can distinguish an element c_1, in $\mathfrak{M}_1$, c_2 in $\mathfrak{M}_2$, $\ldots c_n$ in $\mathfrak{M}_n$ and thereby get a determinate infinite sequence of real numbers.

We have here a special case of the *principle of choice*; its general formulation is the following: "If for every thing x of type $\mathfrak{G}_1$ there is at least one thing y of type $\mathfrak{G}_2$ which stands to x in the relation $\mathfrak{B}(x, y)$, then there is a function φ which correlates each thing x of type $\mathfrak{G}_1$ unambiguously with one thing $\varphi(x)$ of type $\mathfrak{G}_2$ which stands to x in the relation $\mathfrak{B}(x, \varphi(x))$."

In the case at hand the type $\mathfrak{G}_1$ is that of the positive integers, $\mathfrak{G}_2$ that of the real numbers; the relation $\mathfrak{B}(x, y)$ is the inequality

$$a - \frac{1}{x} < y \leqq a \ ,$$

and the function φ, the existence of which is derived from the axiom of choice, correlates the real number c_x with its number x.

Zermelo was the first to recognize the principle of choice as a special presupposition and to formulate it set-theoretically; its use involves a further overstepping of the finite point of view, which goes even beyond the application of the "tertium non datur." The above consideration of methodological examples teaches us that the foundations of the infinitesimal calculus, as they have been given since the discovery of rigorous methods, does not involve a reduction to *finite* number-theoretic thought. The *arithmetizing* of the theory of magnitudes carried out here is in so far *not a complete one* as it involves certain systematic and fundamental conceptions that do not belong to the domain of intuitive arithmetic thinking. The insight that the rigorous foundation of analysis has brought us is that these few fundamental assumptions already suffice to build up the theory of magnitudes as a theory of sets of numbers.

Large areas of mathematics, such as function theory, differential geometry, and topology (analysis situs) are governed by the methods of analysis. General *set theory*, the methods of which have penetrated the newer abstract algebra and topology, makes the most extensive use of non-finite assumptions going well beyond the presuppositions of analysis.

Arithmetic in its usual treatment is by no means an expression of the finite point of view, but depends essentially upon additional principles of inference. We see us therefore confronted with the task of justifying the application of those principles which transcend finite thinking by means of a consistency proof, if we want to keep arithmetic in its current form while acknowledging the demands of the finite point of view with respect to evidence. If such a proof of the consistency of customary ways of inference in arithmetic would be successful, we would also have the guarantee that the results of these ways of inference could never be refuted by a finite determination or finite reflection; for finite methods are included in ordinary arithmetic, and a finite refutation of an assertion proved by ordinary means of arithmetic would therefore indicate a contradiction within ordinary arithmetic.

We return then to the problem raised in § 1. It remains to answer the question from which the considerations of this chapter began: whether instead of using the formalization of inferences to prove the impossibility of a contradiction arising in arithmetic, we couldn't more easily found all of arithmetic directly without additional assumptions and make that proof of impossibility superfluous.

The answer to this is for one part positive, for the other negative. The investigations of Kronecker and Brouwer have shown what is involved in the possibility of a direct finite founding arithmetic to an extent sufficient for practical applications.

Kronecker was the first to insist on the requirements of the finite point of view; he intended to eliminate completely non-finite modes of inference from mathematics. He reached his aim in the theory of algebraic numbers and number fields.[5] Sticking to the finite point of view is possible in this case in such a way that nothing essential of theorems or methods of proof has to be given up.

Kronecker's presentation of the problem was completely rejected for a long time. In more recent times Brouwer set himself the task of founding arithmetic independently of the law of the excluded middle and developed considerable parts of analysis and set theory in terms of this program.[6] Of course, in using this procedure essential theorems have to be given up and considerable complications in the forming of concepts have to be accepted.

The methodological standpoint of "Intuitionism," which Brouwer makes fundamental constitutes, a certain *extension of the finite position* insofar as Brouwer permits the introduction of the assumption that an inference or a proof is given even though the intuitive nature of the inference or proof is not determined. For example, from Brouwer's point of view assertions of the following forms are permissible: "if on the assumption that $\mathfrak{A}$, the sentence $\mathfrak{B}$ holds, then $\mathfrak{C}$ holds too" or also "the assumption that $\mathfrak{A}$ is refutable leads to a contradiction" or in Brouwer's words: "the absurdity of $\mathfrak{A}$ is absurd."

[5]Kronecker did not systematically publish the results of these investigations, but only presented them in lecture courses.

[6]A detailed list of Brouwer's publications in this area is found in A. Fraenkel's "Einleitung in die Mengenlehre," *third* edition. Berlin: Julius Springer 1928.

An extended version of the finite point of view of this kind turns out to be necessary if one is going to go beyond a certain elementary domain using finite considerations; from an epistemological point of view this version amounts to adding considerations of a general logical character to the intuitive insights. We will be led to the requirement of this extension at an advanced stage of our considerations.

Although the above-mentioned investigations indicate a way by which one can get by in mathematics without using non-finite ways of inference, a proof of the consistency of the ordinary methods of arithmetic is by no means superfluous. For the avoidance of non-finite methods of inference does not work in the sense of completely replacing these methods with other considerations; it rather succeeds in analysis and related areas of mathematics only at the cost of an essential loss in systematization and proof-technique.

A mathematician, however, cannot be expected to accept such a loss without compelling reasons. The methods of analysis have been tested to a greater extent than hardly any other scientific presupposition, and they most brilliantly proved a success. If we criticize these methods from the point of view of evidence, then we face the task of tracing the reason of their applicability, just as we do everywhere in mathematics where a successful procedure is applied on the basis of conceptions which, in terms of evidence, leave much to be desired.

Insofar as we accept the finite point of view we cannot escape the problem of obtaining a clear understanding about the applicability of non-finite methods, and this understanding, insofar as our trust in these methods is not misleading, can only consist in our gaining certainty that these ordinary arithmetical methods can never lead to a provably false result, that is to say, that the results of their application are compatible with each other and with any fact evident from the finite point of view.

This problem is, however, the very same as that of the proof of consistency of our ordinary arithmetic.

For dealing with this problem we have already considered in § 1 the method of formalizing logical inference as developed in formal logic.[7] In any case this method satisfies the condition of making the demanded task of a consistency proof a *finite* problem—provided that the complete formalization of ordinary arithmetic succeeds. For if ordinary arithmetic is formalized, i. e., if its presuppositions and ways of inference are translated into initial formulas and rules of deduction, an arithmetical proof presents itself as a succession of intuitively surveyable processes each of which belongs to a stock of relevant actions specified in advance. So we have in principle the same methodological state of affairs as in elementary number theory; and to the same extent as it is successful there to make impossibility proofs in a finite sense, e. g., for the fact that there are no two numerals $\mathfrak{m}$, $\mathfrak{n}$ with

$$\mathfrak{m} \cdot \mathfrak{m} = 2 \cdot \mathfrak{n} \cdot \mathfrak{n} \ ,$$

it is a finite problem as well to show that there is in finite arithmetic no two proofs such that the end-formula of the first is identical with the negation of the end-formula of the second.

However, we are still far away from a solution of this problem. But in pursuing this goal many rewarding results have already been gained; and

[7]Cf. p. 18.

in this way a new field of research has been opened by making use of the formalization of logical inference for a systematic *proof theory* which deals with the question of the significance of logical ways of inference in systematic generality, a question which was posed and solved in traditional logic only in a very special form. By its methods of investigation, the problems of the foundations of mathematics are directly connected with logical problems.

This proof theory, also called "*metamathematics*," will be developed in the following. We start with the formalization of inference which we will present in the beginning independently of its application to proof theory.