Notes for

Set Theory/Math 488/Winter, 2000

copyright: M. Makkai, January, 2000

Contents:

Introduction

These notes* deal with mathematical theories concerning the foundations of mathematics. As a result of these theories, most of which have been elaborated in the last hundred years or so, our view of mathematics has reached a level of maturity where several previous conjectural ideas and beliefs have received exact formulation and confirmation. Also, these theories have revealed a previously unsuspected depth and richness, in the world of abstract ideas.

The main ingredient of the modern view of mathematics is that mathematics is a formalized axiomatic theory. It has to be added that several established philosophies of mathematics would categorically reject the assertion made in the last sentence. When we learn in detail what the content of this assertion is, it will, I hope, turn out that this content has a core that is acceptable to a wide variety of philosophies of mathematics. The assertion intends to express a statement of fact about the organization of mathematics and leaves open the possibility of differing views concerning the ultimate nature of mathematics.

What is interesting is that the concepts that underlie the foundations of mathematics themselves have a rich mathematical theory. This theory actually becomes indispensable for understanding, as opposed to merely observing the fact of, these foundations.

In this course, the emphasis will be on the mathematical theories of the ingredients of the foundations of mathematics

* The notes I am referring to include, beyond the present set-theory part, notes on further areas of logic. The latter are used in Math 591 & 592.

How mathematics is actually built on these foundations will be of secondary importance. One reason for this is the fact that one recognizes in the practice of present day mathematics a strict adherence to foundational principles. In main-stream mathematics, the concepts are set-theoretic, and proofs are 'almost' formalized proofs in the formal language of set theory. Therefore, it is not necessary to make a special effort to uncover the foundations, since they are essentially made explicit in the particular mathematical disciplines themselves. It is interesting to note that many active mathematicians would not be able to formulate precisely the structure of the language they are using, but the 'linguist of mathematics', the person trained in mathematical logic, immediately recognizes the presence of this structure in everyday mathematical discourse.

The first ingredient of the foundations of modern mathematics is set theory. The concept of set (collection, totality) may be the simplest existing abstract concept. Starting with the empty set, one can successively build more complicated sets, without using anything else than sets. In this framework, the whole of traditional mathematics can be reconstructed, and in fact, contemporary mathematics is being built directly within the framework of set theory.

In the first part of these notes, a very short introduction to set theory is given. Modern set theory is a rich and complex discipline, with important consequences concerning the foundations of mathematics. Our introduction cannot even begin to give an idea of the scope of this theory.

Set theory will be presented as an axiomatic theory. The idea of axiomatic theory is very old; Euclid's 'Elements' was intended as an axiomatic theory of geometry. The axiomatization of a body of knowledge has two parts. The first is the choice of the primitive notions of the theory. In Euclidean geometry, these include the notion of point, line, 'betweenness' (point B being on the line AC and between A and C), etc. Once the primitive notions are delineated, it is stipulated that all other notions (which may have been 'unanalyzed' in the pre-axiomatic theory) have to be defined purely in terms of the primitive ones; resulting in the defined notions of the theory. The second part of the axiomatization is the choice of the axioms of the theory: those statements regarding the primitive (or defined) notions that are accepted without proof. Once the axioms have been fixed, all statements, if claimed, have to be proved by pure logical reasoning, from the axioms as hypotheses.

In the second half of the last century, a thorough revision of Euclid's axiomatics was undertaken, culminating in David Hilbert's work, Foundations of Geometry. It turned out that Euclid omitted to state certain primitive notions as well as axioms, i.e., he used them only implicitly, and thereby violated the basic rules of axiomatics. The most important conclusion, however, that emerged from these investigations was that traditional axiomatic theories in general, and Euclidean Geometry in particular, all made tacit use of what we today call set theory. In other words, a new, more elementary and abstract level of mathematical hypotheses was revealed in what previously had been considered 'pure logic' and had remained unanalyzed.

As we said, in Part I we present set theory as a traditional axiomatic theory. We mean by this that we keep the language of this theory unanalyzed. Of course, we do not make the mistake to keep set theoretical matters implicit, as Euclid and others did for geometry; that would defeat the purpose of the axiomatization of set theory itself. It may be said that this stage in axiomatizing set theory is, in fact, a pedagogical exercise, since it has been deliberately formulated as an intermediate stage on a way to a higher level.

Besides giving an example of a traditional axiomatic theory, Part I has other aims. Primarily, it is an introduction to set theoretical matters that every mathematician should know about. Then also, it serves as the foundation and framework for everything else done in this course.

Part II deals with the main subject, first order logic. First order logic, or predicate logic, is a scheme for generating languages, one for each axiomatic theory. In other words, dealing with first order logic we consider an arbitrary specification of the primitive notions; any such specification then, with appropriate axioms added, forms a first order theory, formulated in first order logic.

Axiomatic set theory is a particular first order theory. Actually, we will see essentially different versions of axiomatic set theory, but now we return to the use of the singular mode in reference to these, meaning any particular one. Axiomatic set theory is formulated in first order logic. This means that its axioms are first order sentences, well formed formulas of first order logic, based on the primitives of set theory (in one version, the only primitive is 'x is an element of y', '$x \in y$'). It also means that there are explicit rules of inference (including logical axioms) that are allowed to be used exclusively in deriving theorems of the formal theory. The explicit specification of the rules of inference is the chief novelty of modern axiomatics with respect to traditional ones. The specification of a first order theory, with its logical basis made explicit, is the ultimate realization of the idea of axiomatization. With this specification given, there is an unambiguous and fully communicable criterion for correctness within the theory. In the case of axiomatic set theory, with mathematics developed within it via a system of defined notions, one obtains a fully explicit criterion of correctness of mathematical proof; to a large extent, this precise criterion is the one that operates in mathematical practice today.

The claim that mathematics is exact and rigorous has been made for a long time, but it was only recently that this claim could be considered established. Although axiomatic set theoretical foundations of mathematics is quite simple (in particular, first order logic is a very simple artificial language), it has not been easy to arrive at this formulation. In the first place, it was a highly non-trivial discovery of the 19th century that the large body of seemingly disparate mathematical notions (numbers, geometric surfaces and bodies, sequences, functions) can be 'relocated' as defined notions in set theory. In the second place, it has taken a considerable amount of search to realize that first order logic provides an adequate linguistic framework for set theory.

For stating a first order theory, the formal or grammatical aspects of first order logic, i.e., its syntax, are sufficient. To understand the syntax of first order logic, however, one has to study its semantics, i.e., the theory of meaning in first order logic. The chief point here is the question "why exactly the chosen logical rules of inference, and not others". Although this is not a priori plausible, it turns out that there is a satisfactory theory explaining the choice of the rules of inference. However, this theory is based on semantics, and semantics turns out to be based on set theory in an essential way. Thus, we have the situation that aspects of axiomatic set theory are being explicated by using set theory itself. This seems to be an unavoidable situation. No defeating circularity is introduced this way, however. The statement of axiomatic set theory does not need set theory; it only needs a primitive syntactic level of abstraction. Accepting set theory as essentially correct, then we have a theoretical insight into why that statement was chosen in the first place. This insight is not a prerequisite for being able to state the axiomatic theory, it is only a welcome theoretical development in the meta-theory of the axiomatic theory.

Part II of these notes deals with first order logic. Chapter I of Part II treats a fragment of first order logic, propositional logic. This chapter is a pedagogical preparation for the study of the full language, as well as a store of results used later.

Chapter II deals with the basics of the semantic study of first order logic; as we said, this study uses set theory in an essential way. At first, we do not consider any specific first order theories; logic is studied "in abstracto". At the end of the chapter, we are in the position of being able to give an argued statement of the axioms of first order axiomatic set theory, based on the "naive set theory" of Part I.

Chapter III starts with the introduction and justification of the rules of inference of first order logic. The main result is Gödel's completeness theorem that expresses the sufficiency of a certain body of rules of inference, in view of the semantics of first order logic. Appending the rules of inference to the system of axioms of set-theory, stated in the previous chapter, we arrive at the full statement of first order axiomatic set theory. Of course, many other first order axiomatic theories are open for investigation.

The rest of Chapter III is a further study of the semantics of first order logic. One important theorem, the so-called downward Löwenheim-Skolem theorem is given already in Chapter II. Another basic result, the so-called compactness theorem, is proved as a by-product of the proof of completeness. Chapter III ends with proving Lindström's theorem that is an interesting answer to the question "what is so special about first order logic?". It turns out that, under a suitable general idea of 'logic', any 'logic' that obeys the downward Löwenheim-Skolem and compactness theorems must be identical to, or at least contained in, first order logic.

Returning to axiomatic set theory, we may ask why we chose exactly those set-theoretic axioms that we did. This question has two aspects. The first is: are the axioms correct?, the second is: are the axioms sufficient?

These questions seem to be rather intractable at first sight. If we had precisely statable reasons why the axioms of set theory are correct, we would state these reasons themselves as 'axioms', possibly involving primitives other than set theoretical ones. Then, of course, we could ask the correctness question about the new system, and, barring an infinite regress, we would arrive at an unanalyzable situation. As a matter of fact, set theory itself seems unanalyzable: there do not seem to be any <u>more primitive</u> notions from which set theory could be deduced. Answering the 'sufficiency' question also seems to presuppose a precise pre-axiomatic grasp of what is <u>true</u> of sets; if we had such a grasp that could be precisely stated, we would build that into the axiomatics itself.

It is important to compare the situation of our two "impossible" questions with the analogous ones for pure first order logic, successfully dealt with in Part II. The first point to emphasize is that that analysis involved an <u>acceptance</u> of a certain amount of set theory. The second is that it may be simpler to attack the question of correctness in the context of <u>any and all</u> first order theories, than in the context of a specific axiomatic set theory. The point is that in the first situation the consideration of first order theories other and possibly simpler ones, than set theory, may (and, in fact, do) decide questions we raise.

However, the first question of correctness has a dramatic explicit version. This is the question of formal consistency: Is there a logical contradiction in the theory under consideration? At the beginning of Part I, we encounter the famous Russel paradox, showing that a seemingly fully justified axiomatic theory of sets contains, in fact, a logical contradiction. Now, with having the full formalization of first order logic, the question of formal consistency (freedom from logical contradictions) becomes a very precise question, completely independent from the intuitions and assumptions of the theory itself: in fact, it becomes a question of the syntax of the theory alone.

Ever since the discovery of the paradoxes of set theory, there had been an intense interest in the possibility of establishing the consistency of axiom systems underlying mathematics. Hilbert's program has the ultimate goal of proving, by a direct examination of the syntax, the consistency of first order axiomatic set theory. In 1931, Kurt Gödel made the dramatic discovery that amounts to a proof of impossibility of Hilbert's program. He showed that the consistency of set theory, easily formulated in the language of the theory, cannot be proved in the theory, assuming that the theory is consistent. Gödel's argument does not use much of the specifics of set theory; he shows the result for any explicitly axiomatized theory containing a certain minimum of finite combinatorial (number-theoretical) capability. Thus, the situation is that either the theory is inconsistent (in which case anything can be derived in it!), or it is consistent, but knowing this fact requires going beyond the capability of the theory itself. The theoretical possibility is there for going beyond the established principles of set theory in a way that does not lead to more abstract principles, merely to different ones, in order to prove the consistency of set theory. However, such a possibility seems unlikely to be realized.

As far as the question of sufficiency of the axioms is concerned, ~~it has a similarly sweeping general nega~~tive answer; in fact, the answer is contained in the above! Roughly speaking, Gödel's incompleteness theorem says that no explicitly axiomatized, formally cosistent theory having a minimal combinatorial capability can be complete: there are statements that are recognized to be intuitively true, but are unprovable in the given axiomatic system. In fact, the consistency of the system is such a statement! Technically, the incompleteness theorem is derived first; the (stronger) result on consistency is then proved by applying the ingredients of the incompleteness proof.

Gödel's theorems are the highlights of Part III. The general framework of this part is that of recursive function theory. The concepts of recursive function theory serve to give a precise meaning to the adjective "explicitely axiomatized" used in the description of Gödel's theorems. Recursive function theory has relevance to an experience broader than axiomatic systems; it is a theory of effective generability and computability in general. Chapter I is an introductory one; Chapter II contains the basic concepts and results of recursion theory. Chapter III is dominated by Gödel's theorems. Chapter IV is an account of more recent and technical aspects of recursion theory.

Modern mathematical logic has developed explosively in the last quarter of a century. These notes can only give a first glance at the subject! The following books, containing much of the modern theory, are recommended for further study:

J.R. Shoenfield, Mathematical Logic, Addison-Wesley, 1967.

K. Kunen, Set Theory, An Introduction to Independence Proofs, North-Holland, 1980.

C.C. Chang and H.J. Keisler, Model Theory, North-Holland, 1973.

T. Jech, Set Theory, Academic Press, 1978.

H. Rogers, Theory of Recursive Functions and Effective Computability, McGraw-Hill, 1967.

1. Sets and classes

The concept of set is one of the simplest abstract ideas available to us. It is not possible to define the concept of set; we can only circumscribe it, relating it to other concepts, and ultimately, we can axiomatize it, putting down a hopefully complete list of postulates concerning it.

A particular set is given by a specification of what things belong to it (and by inference therefore, what things do not). In symbols, "x belongs to the set A" is written "$x \in A$". Usually, one arrives at a particular set by a <u>rule</u> that singles out certain things to belong to the set to be defined, or by a <u>property</u> (<u>predicate</u>) that is (somehow) defined to be possessed by (to hold for) certain things, and not for others. E.g., if a property P is specified (so that we know unambigously, if any thing <u>has</u> P or not), the <u>extension</u> of P is <u>the</u> set such that any thing belongs to the set just in case the thing has P. The difference between the property P and its extension $[P]$ is that whereas P uniquely determines $[P]$, $[P]$ is an abstraction of P, and it well may be that for another, very different, property Q we have $[Q] = [P]$: this happens precisely if any thing has P if and only if it has Q.

The above discussion contains two basic ingredients. One is that

> two sets A, B are equal just in case, for every thing x, x belongs to A if and only if x belongs to B,

which is called the principle <u>(axiom) of extensionality</u>. The other is that sets can be <u>created</u> by properties, by passing from the property to its extension:

> given any property P, there is a (by extensionality, unique) set [P] such that for any thing x, x belongs to [P] if and only if x has the property P.

This last principle is called the principle of <u>unrestricted comprehension</u>. As far as we do not undertake analyzing the notion of 'property', we remain in what may be called naive set theory. But it should be said that properties are linguistic objects. We can 'enunciate' a property, making a linguistic utterance, and thereby creating (mysteriously) a 'physical' (metaphysical) object, its extension. It is the main purpose of this course* to come to a satisfactory clarification of the notion of 'property'. But, it is possible, and in fact, also necessary, to deal with sets before we have made that clarification.

Let us next turn to the greatest embarass ment of set theory, which seemingly destroys the possibility of a theory of sets along the lines sketched so far. We will survive this destruction, but only at the expense of having to deal with an essentially more complicated theory.

If any of our initial declarations are to make any sense, a set is a <u>thing</u> again, with the implication that it may, or may not,

---

* meaning also Math 591 & 592

belong to another set. Confirmed in this belief, now consider the following property of sets: the set A has the property P in question just in case A does not belong to itself; in symbols:

$$A \text{ has } P \Leftrightarrow A \notin A.$$

Consider the extension $[P]$ of the property just defined. Does $[P]$ belong to itself? Well:

$$A \in [P] \Longleftrightarrow A \text{ has } P \Longleftrightarrow A \notin A \quad \text{for all } A$$

$$\text{put } A := [P]: \quad [P] \in [P] \Leftrightarrow [P] \text{ has } P \Leftrightarrow [P] \notin [P].$$

by the definition of 'extension'

by the definition of P itself, for the set $A = [P]$

We obtain that $[P]$ belongs to itself just in case it does not. This is a logical contradiction: neither of the two cases "$[P]$ belongs to itself", "$[P]$ does not belong to itself" may hold, since by the equivalence demonstrated, from either one the opposite follows. Even before doing anything specific about logic, it is clear that this situation is not containable. This argument is called <u>Russel's paradox</u>.

The way out of this situation will be the partial renunciation of the principle of unrestricted comprehension. Our point of view will be that the trouble is that unrestricted comprehension creates certain 'collections' that are too 'large' to be sets.

Although the analogy is not perfect, we still may contemplate the following. Suppose some creatures have an intuitive conception of 'finite set', but not of 'arbitrary' sets in our sense, although they are not aware of this since they don't know about any other than finite sets; they call the things they know simply sets, or in other words, they believe that every set is finite. In their set theory, they have a principle that says, 'every set is finite'; this may be formulated e.g. by saying that for every set A there is a natural number n (a concept that has to be, and can be, axiomatized itself) and a one-to-one correspondence between the elements of A and the natural numbers less than n. Now, unrestricted comprehension is clearly untenable; e.g. the 'set' of all natural numbers will not satisfy the axiom of finiteness.

The analogy fails in the sense that the hypothetical creatures have a way out by recognizing that their belief in the principle "every set is finite" was mistaken, and still hold on to unrestricted comprehension. Still, I believe, the analogy points to the basic attitude we will assume. This is that the 'galaxy' of sets is only a part of the 'cosmos' of all kinds of 'totalities' and when one 'comprehends' a property of sets, say, by forming the extension of the property, one may step out of the 'galaxy'. This attitude has the important positive element of recognizing the validity of comprehension with the modification that the result is an entity not necessarily among the ones (sets) so far considered.

In this realm of abstraction there is so little possibility for analysis in the sense of dissecting the notions into more primitive ingredients that we have to be content with formulating our standpoint as clearly as possible and venture out to test its consequences. This is what we will soon do.

To repeat, our situation differs from that of the hypothetical creatures believing in finiteness in that, although our sets necessarily form only a part of things, still, we do not have a neat delineation of the part in question, as the creatures do by having their finiteness principle. We can make only fairly vague statements what distinguishes (delineates) sets from the rest of things. These vague statements will guide us in our choice of several "special principles of comprehension" that we need to have the universe of sets sufficiently rich for building mathematics. In the view that nowadays is called the Cantorian view of sets, the sets are 'completed totalities', completed in the sense of an idealized process of construction. There are, on the other hand, non-completed totalities such as the universe of all sets. The complete statement of the philosophy of Cantorian set theory is tantamount to an argued listing of the principles of set theory.

An important, and mathematically meaningful realization of the idea that sets are gradually constructed is the notion of pure set, developed in the next section. Simply put, pure sets are those that can be built by starting with nothing, using the 'construction principle': take any number of previously constructed pure sets; if their collection qualifies as a set, take that set to be an accepted pure set as well. If this sounds to be empty words, wait until the next section.

In our naive set theory, we take advantage of the 'positive' attitude towards the paradox as follows. We take as basic the concept of set; then we also consider as basic the notion of class, however, we restrict classes to collections all whose members are sets. Soon we will be able to limit our attention to sets all whose members are sets again (even though this looks pretty much impossible); at that point, every set itself will be a class too. A class that is not a set is a proper class. We assume the principle of extensionality for sets as well as classes (if every set is a class, it suffices to assume the principle for classes alone). The principle of comprehension that we adopt will be:

> (Axiom of Class Comprehension) Given any property P of sets, there is a (by extensionality, unique) class [P] such that for any set x, x belongs to [P] if and only if x has the property P.

The actual unfolding of set theory will consist in declaring about certain classes that they are, in fact, sets.

The vagueness of the notion of 'property' will remain with us for some time, however.

A piece of notation: for the extension of a property P we usually write: $\{x : x \text{ has } P\}$. This notation is mainly used when P is given by an expression stating an assertion on x so that having P is equivalent to the truth of the assertion. E.g., the 'paradoxical set' (but harmless as a class) is $\{x : x \notin x\}$. The role of stating the variable x in front can be seen when one defines a property by referring to other (fixed) entities, also denoted by variables. E.g.

$$\{x : y \in x\}$$

is the class of all sets that have the fixed y as an element. Of course, $\{x : x \in y\}$ is the same as y.

Let us record here the content of Russel's paradox in our new setting:

Theorem 1.1 There is a class which is not a set.

Proof: exercise.

It could be legitimately said that our severe restriction on classes, namely that every class that is admitted in the theory is a subclass of the class of all sets, is unnatural. It would be natural to allow classes of proper classes, etc. In that case, however, a concept of 'superclass' would have to be contemplated; the totality of all classes would have to be a superclass. Notice that we do not have to stop at this new level either. It seems to be an important question of the foundations of mathematics what can be said about the seemingly infinite hierarchy of the so obtained "set theories". It turns out that, in a very definite sense, the theories so obtained are getting strictly stronger and stronger. This fact, and other important ones, will be discussed in due course.

2. The universe of pure sets

The next question we have to tackle is: what are the things sets are made out of? The answer will be: sets themselves!

We have to convince ourselves that there are enough sets around whose members are themselves sets. But how can we start building such sets if we do not have any material yet for their members? Is there at least one such 'pure' set?

The answer is: yes, the empty set. The empty set, $\emptyset$, has no members; hence, in particular, it is vacuously true that all its members are sets. At this point, note that the 'absurd' property (e.g., $x \neq x$) defines the empty class $\emptyset$.

$$x \in \emptyset \Leftrightarrow x \neq x.$$

Our first set-existence principle is:

(Axiom of Null Set) The empty class is a set.

Now, other sets are possible: the set, denoted by $\{\emptyset\}$, whose only member is $\emptyset$ is a new one, or the set having the sole elements $\emptyset$ and $\{\emptyset\}$, denoted $\{\emptyset,\{\emptyset\}\}$, is another one. Here is a principle that codifies the process of construction employed here.

(Axiom of Pair set) For any sets x and y, the class $\{x,y\}$ is a set.

Here $\{x,y\}$ denotes the class whose sole elements are x and y, in other words

$$\{x,y\} = \{z : z = x \text{ or } z = y \text{ (or both)}\}.$$

In fact, it is clear that by repeated use of the pair-set principle, we can have 'infinitely many' sets - but, we do not get any infinite set without further set-construction principles.

Before going on to those, we want to describe the general idea of 'pure set'. The above examples of sets are all pure. Moreover, the set-construction principles we will adopt give pure sets when we apply them to pure sets. In short, set theory will take place in the realm of pure sets.

Let us emphasize that we presuppose the concepts of 'set' and 'class', in the way we said in the last section. In other words, the definition of 'pure set' will rely on the "unanalyzed" notions of 'set' and 'class'.

The basic principle is that if all members of a given set x are pure sets, then x itself is pure. This expresses the primitive idea that we accept unlimited set-formation as the basic 'construction' for sets: if we have accepted that certain things are pure sets, and we throw them into a box, thereby forming a class of sets, and then we recognize that the class so obtained is a set, then we accept this set as a pure set. Moreover, we want to be parsimonious about the concept of pure set in the sense that we allow something to be a pure set only if we are forced to do so by the principle just formulated. Notice that we are so forced in the case of the empty set $\emptyset$, and then, in turn, for the two other examples of sets we mentioned above.

The precise concept now is dictated to us as follows. The class of all pure sets, $\mathbb{V}$, should have the property that it is contained in any other class X that satisfies:

> whenever every member of a set x belongs to X, so does x.

Call such a class one <u>closed under set-formation</u>. Let us define a set to be <u>pure</u> if it belongs to <u>all</u> classes closed under set-formation. In somewhat formal terms:

> $x \in \mathbb{V} \Leftrightarrow$ for every class X, if (for all y, $y \subset X$ implies $y \in X$), then $x \in X$.

(Notation: $y \subset X$ means: every element of y belongs to X.) Notice that on the right hand side of the definition, a <u>property</u> of the variable set x is stated; if we allow it as legitimate, we may form $\mathbb{V}$, as the extension of that property. Since our attitude towards properties is naive, we do so without hesitation.

<u>Proposition</u> 2.1 $\mathbb{V}$ is closed under set-formation.

Before we give the proof, we reiterate that our position with respect to properties, as well as to manipulating them (inferring the presence of a property from the presence of another, which, by the way, is <u>logic</u>) is, at present, naive: we follow established mathematical practice.

<u>Proof</u>: Suppose x is a set such that every element of x is a pure set (belongs to $\mathbb{V}$). By definition then, if $y \in x$,

then $y \in X$ for every class X closed under set-formation. Fixing an arbitrary such class X, we then see that every member of x belongs to X, hence by the definition of X being closed under set-formation, $x \in X$. Since this turned out to be true for all such X, x belongs to all classes closed under set-formation, i.e., $x \in \mathbb{V}$. □

As a consequence of the last proposition, we may say that $\mathbb{V}$ is the _least_ class closed under set formation: it is closed under set-formation and it is contained in every such class.

Proposition 2.2 Any element of a pure set is a pure set.

Proof: Consider the class X of all those pure sets for which it is true that all their elements are pure sets. [Notice that the last sentence means the application of an instance of class comprehension.] Of course, $X \subset \mathbb{V}$. We claim that X is closed under set-formation. That will imply $\mathbb{V} \subset X$, hence $\mathbb{V} = X$, proving the proposition. To show our claim, let x be a set and assume that all elements of x belong to X. In particular then, all elements of x belong to $\mathbb{V}$, i.e., all elements of x are pure sets. x itself is a pure set (since $\mathbb{V}$ is closed under set-formation); thus we see that x satisfies the requirements of being a member of X: $x \in X$. This completes the verification of X being closed under set-formation, and finishes the proof. □

Proposition 2.3 The principle of extensionality holds when restricted to $\mathbb{V}$. In other words, if x and y are pure sets, and for all pure sets z, we have that $z \in x$ iff $z \in y$, then $x = y$.

<u>Proof</u>: <u>exercise</u>.

The definition of $\mathbb{V}$ gives rise to a method of proof, called the <u>principle of $\in$-induction</u>. Suppose we want to show that an assertion $A(x)$, concerning an arbitrary pure set x, is in fact true for all pure sets x. Then, the principle says, it suffices to show the following: whenever x is a set all of whose elements satisfy $A(y)$, then x itself satisfies $A(x)$. Indeed, this principle is correct, since we can form the class

$$X = \{x : A(x) \text{ is true}\};$$

the assumption then says that X is closed under set-formation, hence by definition of $\mathbb{V}$, $\mathbb{V} \subset X$; in other words, every pure set x satisfies $A(x)$.

Here are two simple applications of $\in$-induction.

<u>Proposition</u> 2.4 (i) For all $x \in \mathbb{V}$, $x \notin x$.

(ii) For pure sets x, y, it is impossible that both $y \in x$, $x \in y$ hold.

<u>Proof</u>: (i) Suppose the assertion is true for all elements y of x. If we had $x \in x$, then x would be an element of x, hence, in particular, the assertion in question would be true for x, i.e., $x \notin x$, contrary to the assumption $x \in x$. It follows that $x \notin x$.

(ii) We now have two variables in the assertion:

x and y. We form an assertion for only one of them, namely x, by considering the assertion $A(x)$:

"for all $y \in \mathbb{V}$, $y \in x$ and $x \in y$ cannot hold simultaneously".

We show $A(x)$ for all $x \in \mathbb{V}$ by $\in$-induction; this will obviously prove the assertion. Indeed, assume $A(x')$ holds for all $x' \in x$, we want to show $A(x)$ itself. To this end, let $y \in \mathbb{V}$, and assume that, contrary to $A(x)$, $y \in x$ and $x \in y$ both hold. But by the first assumption, $y = x'$ is an element of x; we have both $x \in x'$ and $x' \in x$, contradicting the assumption that for x' the assertion $A(x')$ holds. In other words, $A(x)$ indeed follows. By set-induction therefore, the proof is complete. □

<u>Corollary</u> 2.4' $\mathbb{V}$ is not a set.

<u>Proof</u>: <u>exercise</u>.

We now describe how $\mathbb{V}$ is built in stages starting from the empty set. The description will be incomplete at this time; it will be completed in Section 5.

* We write $\mathbb{V}_0$ for the empty set,

$$\mathbb{V}_0 = \emptyset;$$

let

$$\mathbb{V}_1 = \{\emptyset\},$$
$$\mathbb{V}_2 = \{\emptyset,\{\emptyset\}\},$$
$$\mathbb{V}_3 = \{\emptyset,\{\emptyset\},\{\{\emptyset\}\},\{\emptyset,\{\emptyset\}\}\};$$

<u>in general, define for any natural number n,</u> $\mathbb{V}_{n+1}$ to be the

* Here the "informal description" <u>begins</u>; it will <u>end</u> on p. 27 as indicated. We have <u>not yet</u> defined 'natural numbers', among others; here we use natural numbers informally.

set of all subsets of $\mathbb{V}_n$,

$$\mathbb{V}_{n+1} = \{x : x \subset \mathbb{V}_n\}.$$

In other words, the elements of $\mathbb{V}_{n+1}$ are obtained by "set-formation from the elements of $\mathbb{V}_n$"; every element of $\mathbb{V}_{n+1}$ is one whose elements are in $\mathbb{V}_n$; and in fact, conversely, $\mathbb{V}_{n+1}$ contains as elements all such sets. Now, we can form

$$\mathbb{V}_\omega = \mathbb{V}_0 \cup \mathbb{V}_1 \cup \cdots \cup \mathbb{V}_n \cup \cdots = \bigcup_n \mathbb{V}_n$$

the union of all the $\mathbb{V}_n$'s; in other words, something belongs to $\mathbb{V}_\omega$ just in case it belongs to at least one $\mathbb{V}_n$. With appropriate set-construction principles, $\mathbb{V}_\omega$ will turn out to be a set. In fact, $\mathbb{V}_\omega$ turns out to be a very important set, the set of <u>hereditarily finite sets</u>. It is an instructive <u>exercise</u> to show that $\mathbb{V}_\omega$ is the smallest class X closed under finite set-formation: whenever x is a finite set all whose elements belong to X, $x \in X$. [We have not yet 'defined' "finite"; still, with reasonably plausible assumptions, this exercise can be done. Later, it can be done formally.]

It is clear that $\mathbb{V}_0 \subset \mathbb{V}_1 \subset \cdots \subset \mathbb{V}_n \subset \cdots \mathbb{V}_\omega$; (or: the sequence $\mathbb{V}_0, \mathbb{V}_1, \cdots, \mathbb{V}_n, \cdots, \mathbb{V}_\omega$ is <u>increasing</u>). It is also clear that each $\mathbb{V}_n$ is a subclass (in fact, subset) of $\mathbb{V}$. This is proved most satisfactorily by induction on the natural number n. But then it follows that $\mathbb{V}_\omega$ itself is a subset of $\mathbb{V}$. In other words, the sequence $\mathbb{V}_0 \subset \mathbb{V}_1 \subset \cdots \subset \mathbb{V}_n \subset \cdots \mathbb{V}_\omega$ represents a "beginning of a construction of $\mathbb{V}$ from below".

In fact, this construction can be continued. We may consider the class of all sets obtained by set-formation from elements of $\mathbb{V}_\omega$, in other words, the class $P(\mathbb{V}_\omega)$ of all subsets of $\mathbb{V}_\omega$ (notation: $P(x)$ denotes the class of all subsets of x: $P(x) = \{y : y \subset x\}$). We'll have a set-construction principle that recognizes $P(x)$ to be a set for all sets x (axiom of power set). It is clear that $P(\mathbb{V}_\omega) \subset \mathbb{V}$: the reason is that $\mathbb{V}$ is closed under (arbitrary) set-formation. It is also easy to see that $\mathbb{V}_\omega \subset P(\mathbb{V}_\omega)$ (exercise). Now, we have extended our sequence "constructing $\mathbb{V}$ from below" by a further member $P(\mathbb{V}_\omega)$; the spirit of the construction dictates the notation $\mathbb{V}_{\omega+1} = P(\mathbb{V}_\omega)$ (although note that neither ω, nor $\omega+1$ makes independent sense so far). It is now left to the reader to define appropriately

$$\mathbb{V}_{\omega+2}, \ldots, \mathbb{V}_{\omega+n}, \ldots$$

$$\mathbb{V}_{\omega+\omega} \qquad (= \mathbb{V}_{\omega\cdot 2})$$

$$\mathbb{V}_{\omega\cdot 2+1}, \mathbb{V}_{\omega\cdot 2+2}, \ldots, \mathbb{V}_{\omega\cdot 2+n}, \ldots,$$

$$\mathbb{V}_{\omega\cdot 2+\omega} \qquad (= \mathbb{V}_{\omega\cdot 3})$$

$$\mathbb{V}_{\omega\cdot 4}$$

$$\vdots$$

$$\mathbb{V}_{\omega\cdot n}$$

$$\vdots$$

$$\mathbb{V}_{\omega\cdot\omega}$$

$$\vdots \; ?$$

Later, we will make precise the notion of what goes into the above sequence as indices, namely that of ordinal number. Once we have done that, we will show that the "transfinite sequence" $\mathbb{V}_\alpha$ with α ranging over all ordinals will in fact exhaust all of $\mathbb{V}$: $\mathbb{V}$ = the union of all $\mathbb{V}_\alpha$.*

From now on by 'set' we always mean a 'pure set', and by 'class' a 'pure class', i.e., one all whose elements are pure sets. We will adhere to the useful convention that capital letters mean classes, lower case ones sets. Of course, a set is automatically a class; so e.g., a definition involving classes (capitals) can automatically be applied to sets (lower case letters). Also notice that if we construct a class and then declare it a set (by adopting a "principle of set-construction"), the set obtained is automatically a pure set, since all of its elements are. Especially at later points, we will use capitals expressly to denote sets; in such cases, it will be stipulated that the capital in question is in fact (not just a class but) a set.

Restricting ourselves to pure sets is equivalent to adopting the following axiom.

(Axiom of Foundation; class formulation) Every set is pure.

In fact, we should pause to consider the assertion just made, even though it seems to be an evident one. The axiom of foundation is to be true under the interpretation of 'set' as 'pure set', and of 'class' as 'pure class': if we happen to use the axiom of foundation, together with the formal definition of 'pure set',

* End of "informal description"

we automatically assign (as agreed) the meaning 'pure set' ('pure class') to the terms 'set' and 'class' in that definition. Therefore, the first thing is to check that the meaning of 'pure set' does not change when we make the formal substitution of 'pure set' for 'set' and 'pure class' for 'class' in the right-hand-side of the definition of pure set. In other words, we want to see that the following is true: for every pure (*) set x,

> $x \in \mathbb{V} \Leftrightarrow$ for every pure class X, if (for all pure sets y, $y \subset X$ implies $y \in X$), then $x \in X$.

Once we have spelled this out, the verification becomes routine (verify), and now we can rest assured that in $\mathbb{V}$, indeed, the Axiom of Foundation is true.

(*) In fact, this is true for all sets x.

3. Further principles of set-construction

We have seen two, fairly primitive, principles of set-constructions: the axioms of null-set and pair-set. These, and all the others, are remnants of unrestricted comprehension. Each principle asserts that certain classes are in fact sets. We give four more in this section. The next section will contain the last remaining such principle.

The list to be given is fairly incomprehensible at first sight: why exactly those and not more or less? In fact, the list was obtained by searching for those instances of comprehension that seemed essential to mathematics and still seemed sound. It is a sad fact that we do not know of any convincing argument in favor of the list, other than the empirical fact that mathematics seems to need essentially these and no more. Even worse, we know that, for several reasons, the list cannot be shown to be the only possible one or even the best one. For an explanation of this latter fact, we have to wait until later.

> (Axiom of Subset; or: Axiom of Separation) Any subclass of a set is a set.

This principle is clearly in the spirit of the belief that "the only thing that may prevent a class from being a set is that it is too big". Therefore, if a class is contained in a set, and hence is no 'bigger' than a set, it may be safely assumed to be a set itself. Notice that the following form of the principle is an equivalent version (because of class-comprehension):

Given any property P of sets, and a set x, the class $\{y \in x : y \text{ has } P\}$ is a set.

In other words, we can form a new set out of a given one, x, by 'separating' those elements of x that satisfy a given property. Indeed, this is the most common way of forming sets in mathematics.

Given a class X, $\cup X$ denotes the union of its elements [recall that every element of X itself is a set!]. In symbols,

$$\cup X = \{z : \text{ there is } y \in X \text{ such that } z \in y\}$$

In short, $\cup X$ may be called the <u>union of</u> X.

(Axiom of Union) The union of any set is a set.

If we write $A \cup B$ for $\{x : x \in A \text{ or } x \in B \text{ (or both)}\}$ ($A \cup B$ is the <u>union</u> of A and B), we have

<u>Proposition</u> 3.1 $a \cup b$ is a set.

<u>Proof</u>: Note that $a \cup b = \cup\{a,b\}$. By the axiom of pair-set, $\{a,b\}$ is a set; by the union-axiom, $a \cup b$ is a set. $\square$

(Axiom of Power-set) The class of all subsets of a given set x is a set.

The class of all subsets of X is $\{y : y \subset X\}$ where $y \subset X$ means: every element of y belongs to X (so, even if $y = X$, we have $y \subset X$). The class of all subsets of X is denoted by $P(X)$. The principle asserts that this class is in fact a set, in case X is a set.

Before we formulate more principles, we discuss some derived notions, and attendant notation. Set theory is used to interpret basic mathematical notions. We start the set-theoretic interpretation of mathematics with the concepts related to 'relation' and 'function'.

The first thing is to introduce a concept of ordered pairs. We should assign a set, denoted (a,b), to any given sets a and b in a unique manner such that "from (a,b) we can recover a and b, in this order", or more formally

$$(a,b) = (a',b') \text{ implies that } a = a' \text{ and } b = b'.$$

There are several ways of defining the assignment

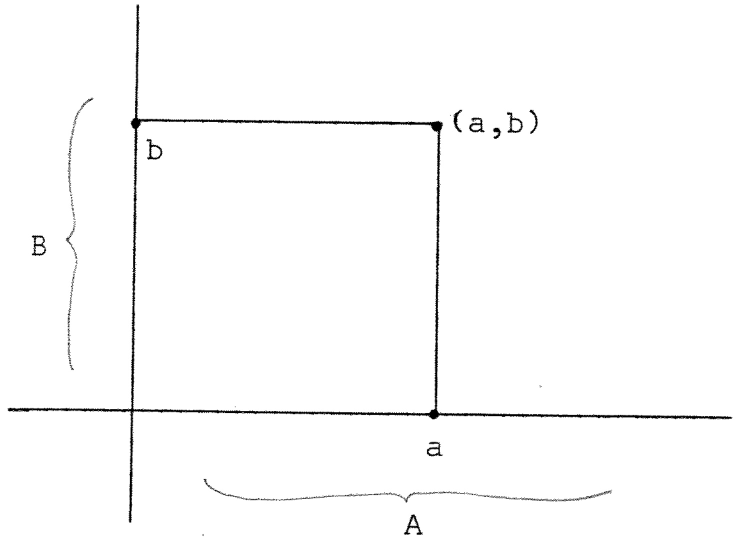
$$a,b \longmapsto (a,b)$$

with this property. One of the simplest is to put $(a,b) = \{\{a\},\{a,b\}\}$. Notice that $\cup(a,b) = \{a,b\}$, and if $a \neq b$, then a is distinguished among the elements of $\cup(a,b)$ by the fact that $\{a\} \in (a,b)$. It is easy to conclude that the required property of the ordered-pair operation indeed holds.

For classes A and B, $A \times B$ (the Cartesian product of A and B) denotes the class of all ordered pairs (a,b) with a from A, b from B: $A \times B = \{(a,b) : a \in A \text{ and } b \in B\}$

$$= \{x : x = (a,b) \text{ for some } a \in A \text{ and } b \in B\}.$$

We usually picture (resembling an important geometric example) the class $A \times B$ as points of the 'plane' with 'axes' A and B:



Proposition 3.2 $a \times b$ is a set.

Proof: Note that

$$a{\times}b \subset \mathcal{P}(\mathcal{P}(a \cup b))$$

(verify). By 3.1, $a \cup b$ is a set; by the power-set axiom $\mathcal{P}(a \cup b)$ is a set; and likewise, $\mathcal{P}(\mathcal{P}(a \cup b))$ is a set. Hence, by the subset axiom, $a{\times}b$ is a set. □

Extending our convention on the use of lower-case and capital letters, we will use capitals in naming special categories of classes, and lower case initials for indicating that we are talking about sets. E.g., we now introduce the notion of Relation; any Relation is a class; at the same time "relation" will automatically refer to a set which is, at the same time, a Relation.

A Relation is a class all of whose elements are ordered pairs. A is a Relation if and only if for every x in A there are a and b such that $x = (a,b)$. Dom(A) denotes $\{a$: there is an ordered pair in A whose first element is $a\}$, Range(A) denotes

$\{b\colon$ there is a such that $(a,b) \in A\}$. Thus, A is a Relation just in case $A \subset \mathrm{Dom}(A) \times \mathrm{Range}(A)$.

<u>Proposition</u> 3.3 Dom(x), Range(x) are sets (denoted dom(x), range(x), respectively).

<u>Proof</u>: By the axiom of subset, it suffices to exhibit a set y such that $\mathrm{Dom}\ x \subset y$, $\mathrm{Range}\ x \subset y$. But $y = \cup(\cup x)$ (union axiom!) qualifies (<u>verify</u>). □

Thus, we have the <u>corollary</u> that a Relation A is a set just in case both Dom A and Range A are sets (why?).

A <u>Function</u> is a Relation A such that (a,b), (a,b') both being in A implies that $b = b'$. In other words, for every $a \in \mathrm{Dom}\ A$ there is a <u>unique</u> b such that $(a,b) \in A$; this b is then denoted by $A(a)$. For a function f (i.e. a set which is a Function) and $a \in \mathrm{dom}(f)$ $(\mathrm{dom}(f) = \mathrm{Dom}(f)$ is a set!), we also write

$$f{'}a \quad \text{for} \quad f(a).$$

In fact, we have the particular Function <u>Application</u>:

$$\underline{\text{Application}} = \{((f,a),b)\colon\ f \text{ is a function, } a \in \mathrm{dom}(f) \text{ and } (a,b) \in f\}$$

for which we have

$$\underline{\text{Application}}((f,a)) = f{'}a$$

for $(f,a) \in \mathrm{Dom}(\underline{\text{Application}})$.

(Axiom of Replacement). If the Domain of a Function is a set, then its Range is a set as well.

Proposition 3.4 If the Domain of a Function is a set, the Function itself is a set.

Proof: immediate from replacement (exercise). □

This axiom is not frequently used in ordinary mathematics, although it is very important in set theory; we'll see applications of it later. The rationale behind the axiom is as follows. Suppose we have a Function F with $a = \mathrm{Dom}(F)$ a set. We imagine F to be a 'process' that 'constructs' $y = F(x)$ out of any x in a. Hence, F 'constructs' Range F as the class of all $F(x)$ with $x \in \mathrm{Dom}(F)$. Therefore, since $\mathrm{Dom}(x)$ is a 'completed totality', we may consider $\mathrm{Range}(F)$ to be one as well. Put it in another way, $\mathrm{Range}(F)$ is "no bigger" than $\mathrm{Dom}(F)$ (this phrase will acquire a precise meaning later), hence $\mathrm{Range}(f)$ can be accepted to be a set if $\mathrm{Dom}(F)$ is one. A warning: the above is not in any sense a proof of the axiom of replacement; if such a proof existed, we would not have called it an axiom!

We'll end this section by discussing a few common pieces of set-theoretical notation.

We write A^2 for $A\times A$, A^3 for $(A\times A)\times A$, etc. Also, for convenience, A^0 will be $\{\emptyset\}$ (!) and $A^1 = A$. The notation $F : A \longrightarrow B$ (similarly, $f : a \longrightarrow b$, etc.) means: F is a Function, $\mathrm{Dom}(F) = A$, $\mathrm{Range}(F) \subset B$.

With reference to a notation of the form $F : A \longrightarrow B$, F is called <u>onto</u> if Range $F = B$. It is called <u>one-to-one</u> if (a,b), (a',b) both being in F implies that $a = a'$, i.e. $F(a) = F(a')$ implies $a = a'$.

There is a notation for functions defined by 'rules' similar to extensions of properties. Suppose we have defined a 'rule' that assigns a certain uniquely determined set a^* to a; then the class

$$\{\langle a,a^*\rangle : a \in A\}$$

is a Function with Domain equal to A. An example is when a^* is a^2.* This same function is denoted as

$$[a \longmapsto a^* : a \in A]$$

or simply, $a \longmapsto a^*$, when A is understood.

If F is a Function with Domain A, then $\cup$Range F is, of course, the same as $\cup\{F(a) : a \in A\}$. We have the commonly used variation $\bigcup_{a\in A} F(a)$ for the same.

For any class A, $\cap A$ denotes the intersection of the elements of A: $\cap A = \{y : y \in x$ for all $x \in A\}$. Notice that $\cap\emptyset = V$ (= the class of all sets). Otherwise, however, we have that $\cap a$ is a set for any set a (<u>excercise</u>). We, of course, write $a \cap b$ for $\cap\{a,b\}$, $\bigcap_{a\in A} F(a)$ for $\cap$Range(F) (if $A = \mathrm{Dom}(F)$).

$A-B$ denotes $\{x : x \in A$ and $x \notin B\}$; if A is a set, $A-B$ is a set as well.

A further operation on sets is <u>exponentiation</u>. A^B denotes $\{f : f : B \longrightarrow A\}$, the class of all functions with domain B

---

* $a^2 = a \times a$ (Cartesian product)

and range included in A. If B is a proper class, then $A^B = \emptyset$ (why?); if both A and B are sets, then A^B is a set.

<u>Proposition</u> 3.5 a^b is a set.

<u>Proof</u>: Note that $a^b \subset P(a \times b)$. □

Because of a possible confusion with cardinal exponentiation to be defined later, we modify the notation A^B to ${}^B A$ from now on.

Let A be a set.* We can essentially identify $P(A)$ with a "power", namely ${}^A 2$; here $2 = \{0,1\}$ (and in fact $0 = \emptyset$, and $1 = \{\emptyset\}$, although the latter are not important for the time being; see later). Given any $X \subset A$, we can define char(X), the <u>characteristic function</u> of X, as the function with domain A for which $a \in X$ implies $(\text{char } X)(a) = 1$, and $a \notin X$ implies $(\text{char } X)(a) = 0$. Formally

$$\text{char } X = \{(a,0) : a \in A-X\} \cup \{(a,1) : a \in X\}.$$

It is easy to verify that the mapping

$$X \longmapsto \text{char } X$$

$$P(A) \longrightarrow {}^A 2$$

is one-to-one, and onto (<u>exercise</u>).

An important 'infinitary' operation on sets is the (<u>infinite</u>) <u>Cartesian product</u>. Suppose $\langle A_i \rangle_{i \in I}$ is a "family a sets"; i.e., we have the function

$$i \longmapsto A_i$$

with domain I; assume that I itself is a set. Then $\mathop{X}\limits_{i \in I} A_i$,

* Of course, there's nothing wrong with denoting a set by a capital letter. However, denoting a class by a lower-case

the Cartesian product of the family, is the class of all functions f whose domain is I and for which $f(i) \in A_i$ for all $i \in I$. E.g., $\underset{i\in\{0,1\}}{X} A_i$ is the same as the set of all $\langle a_0,a_1\rangle$ with $a_0 \in A_0$, $a_1 \in A_1$, where $\langle a_0,a_1\rangle$ denotes the function with domain $2 = \{0,1\}$ whose value at 0 is a_0, and that at 1 is a_1. It seems clear that $\underset{i\in\{0,1\}}{X} A_i$ is 'practically' the same as $A_0 \times A_1$, since $\langle a_0,a_1\rangle$ is 'practically' the same as (a_0,a_1).

<u>Proposition</u> 3.6 Under the stated conditions, $\underset{i\in I}{X} A_i$ is a set.

<u>Proof</u>: Notice that $\underset{i\in I}{X} A_i \subset {}^{I}(\bigcup_{i\in I} A_i)$. □

There are a few operations performed on functions that are often referred to. Given $F : A \longrightarrow B$, $G : B \longrightarrow C$, the <u>composite</u> of F and G is defined as the function H whose Domain is A, and for which $H(a) = G(F(a))$ for $a \in A$. The composite is denoted

$$G \circ F$$

(watch the order!); so we have

$$G\circ F = \{(a,c): \text{ there is } b \text{ such that } (a,b) \in F \text{ and } (b,c) \in G\}$$

which makes it possible to consider "$\circ$" to be defined on any classes F and G.

If $F : A \longrightarrow B$ is one-to-one and onto, it has an <u>inverse</u>: a function, denoted F^{-1}, such that $F^{-1} : B \longrightarrow A$, and

$$F\circ F^{-1} = Id_B$$

and

$$F^{-1} \circ F = \mathrm{Id}_A;$$

here Id_B is the <u>identify function</u> on B: $\mathrm{Dom}(\mathrm{Id}_B) = B$, $\mathrm{Id}_B(b) = b$, and similarly for Id_A. F^{-1} can be defined as $\{(b,a) : (a,b) \in F\}$, which again makes "raising to the power -1" an operation defined on all classes F. <u>Exercise</u>: verify the two above equalities under the conditions stated.

If $F : A \longrightarrow B$, and $X \subset A$, then the <u>image of</u> X <u>under</u> F, denoted $F[X]$, is the class

$$F[X] = \{F(a) : a \in X\}.$$

In the same situation, the <u>restriction of</u> F <u>to</u> X, $F \restriction X$, is $F \cap (X \times B)$; $\mathrm{Dom}(F \restriction X) = X$, and $(F \restriction X)(a) = F(a)$ for $a \in X$. Also, $F[X] = \mathrm{Range}(F \restriction X)$.

There are numerous identities and other relations between the operations on sets and classes we have introduced; e.g., $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$. I believe that there is no point in trying to list these. On the one hand, they are all very easy to verify, and, on the other, they are better motivated when they are stated for immediate use.

4. Natural numbers and ordinals

The first aim of this section is to give a set-theoretic reconstruction of the notion of natural number. Almost simultaneously with this, we will introduce a 'transfinite' number concept, that of ordinal number (simply: ordinal). At first, the notion of ordinal will not be much more than a formal extension of that of natural number, but it will eventually emerge that it is the notion of number that can be used to count arbitrary (even infinite) sets.

The most natural approach to natural numbers is the axiomatic one. After all, it is not that important what the numbers themselves actually are; we can identify them with series of strokes (primitive man's way of keeping count), or ones in the decimal notation, or even Roman numerals. One might say that what is important is what numbers are used for (counting), but actually, the right answer is that natural numbers form a structure which makes possible all the uses of numbers. This structure is described by the so-called Peano axioms. Here is Peano's 'characterization' of the system of natural numbers. We put it in the form of a definition.

Definition 4.1 (Peano Axioms) A Peano system $(\mathbb{N},0,S)$ consists of a set $\mathbb{N}$, a distinguished element (called 0) of $\mathbb{N}$, and a unary operation on $\mathbb{N}$, i.e. a function $S : \mathbb{N} \longrightarrow \mathbb{N}$, called "successor of", such that the following conditions are satisfied:

(i) 0 is not the successor of anything;

$Sn \neq 0$ for all $n \in \mathbb{N}$.

(ii) $Sn = Sm$ implies that $n = m$.

(iii) If a subset P of $\mathbb{N}$ contains 0, and is closed under successor ($n \in P$ implies $Sn \in P$), then $P = \mathbb{N}$.

A few remarks putting Peano's axioms in perspective. In more common language, Sn stands for $n+1$. The two axioms (i), (ii) are intended to ensure that the sequence

$$0, S0, SS0, SSS0, \ldots$$

is without repetition. Axiom (iii) is the same as the Principle of Induction. If one wants to prove that an assertion $A(n)$ is true for all natural numbers n, it suffices to show that $A(0)$ is true, and to show that, for any natural number n, $A(n)$ implies $A(Sn)$. In fact, the validity of the last sentence follows from (iii) by defining P to be the subset of N consisting of $n \in N$ for which $A(n)$ is true.

Proposition 4.1' If $(\mathbb{N}, 0, S)$ is a Peano system, then $S : \mathbb{N} \longrightarrow \mathbb{N}$ is one-to-one, and $\text{range}(S) = \mathbb{N} - \{0\}$.

Proof: exercise.

Of course, for $n \in N$, $n \neq 0$, the unique $m \in \mathbb{N}$ for which $Sm = n$ is denoted $n-1$.

We are not going to do any of the development of the axiomatic theory natural numbers; in fact, we later assume a reasonable amount of knowledge of the various operations and relations on natural numbers. We only want to point out here that the basic reason for the possibility of developing all the theory of natural numbers on the basis of Peano's axioms is the fact that these axioms essentially characterize the system of natural numbers.

Proposition 4.2 Given the Peano systems $(\mathbb{N},0,S)$, $(\mathbb{N}',0',S')$, there is an isomorphism between them, i.e., there is

$$h : \mathbb{N} \longrightarrow \mathbb{N}'$$

which is one-to-one and onto and which preserves 0 and S, in the sense that

$$h(0) = 0'$$

and $n = Sm$ implies $h(n) = S'(h(m))$.

In turn, the Proposition is a consequence of the Recursion Principle:

Proposition 4.3 Suppose $(\mathbb{N},0,S)$ is a Peano system. Then, for any class B, a unique function $f : \mathbb{N} \longrightarrow B$ can be defined by the requirements that

(1) $f(0)$ should be a given element, say a, of B;

and

(2) $f(Sn)$ should equal to, for any n, a given function of $f(n)$; $f(Sn) = g(f(n))$ for a given function $g : B \longrightarrow B$.

In other words, with given $a \in B$, $g : B \longrightarrow B$, there is a unique $f : \mathbb{N} \longrightarrow B$ satisfying

$$f(0) = a$$

$$f(Sn) = g(f(n)).$$

Exercise: Deduce 4.2 from 4.3.

Exercise: Show that, conversely to 4.1, if $(\mathbb{N}', 0', S')$ is isomorphic to the Peano system $(\mathbb{N}, 0, S)$, then it is a Peano system itself.

We will not take the time to prove 4.3 here; later we will prove a more general result (5.4). But we should point out that the Recursion Principle can be used to define operations like addition, multiplication, and many others. Also, the basic process of counting is closely related to recursion; this will be clarified later.

We wish to turn to the task of specifying a particular system $(N, 0, S)$ satisfying Peano's axioms. In particular, all natural numbers should be construed explicitly as pure sets. As a matter of fact, the construction is largely arbitrary, but we will find an attractive solution.

Imagine that we do have already a Peano system $(\mathbb{N}, 0, S)$. We now assign a pure set $\overline{n}$ to every 'natural number' n in $\mathbb{N}$, with which we wish to 'identify' n. We let $\overline{0} = \emptyset$ (= empty set). In fact, we pick for $\overline{n}$ a set with exactly n elements (although the last phrase has not yet been officially introduced in our set theory). Assuming that $\overline{k}$ has been defined for $k < n+1$, we put

$$\overline{n+1} = \{\overline{0}, \ldots, \overline{n}\}.$$

(This set has $n+1$ elements, assuming that the $\overline{k}$ are distinct for distinct k.) In this way, we have made sure that $\overline{n+1}$ is the set of all $\overline{k}$ for $k < n+1$. Since $\overline{n}$ is, similarly, $\{\overline{0},\dots,\overline{n-1}\}$, we actually have

$$\overline{n+1} = \overline{n} \cup \{\overline{n}\}.$$

Although above we used concepts related to natural numbers that we have not formally introduced, now we have a form of the definition of $\overline{n}$ that complies with the form of the Recursion Principle:

$$\overline{0} = \emptyset$$
$$\overline{Sn} = \overline{n} \cup \{\overline{n}\}.$$

(Question: what are f, B, a and g of the Recursion Principle?) We could easily show that $\overline{k} \neq \overline{\ell}$ for $k \neq \ell$. But then, if we put $\overline{\mathbb{N}} = \{\overline{n} : n \in \mathbb{N}\}$, the function

$$n \longmapsto \overline{n}$$

mapping $\mathbb{N}$ into $\overline{\mathbb{N}}$ is a one-to-one and onto map of $\mathbb{N}$ onto $\overline{\mathbb{N}}$. If, in addition, we define

$$\overline{S}x = x \cup \{x\} \tag{1}$$

for $x \in \overline{\mathbb{N}}$, we then have $(\overline{\mathbb{N}},\overline{0},\overline{S})$, a Peano system: it is isomorphic to $(\mathbb{N},0,S)$, and therefore, as is easily checked (see the exercise above), it is a Peano system itself. Finally, we note that the class $\overline{\mathbb{N}}$ is actually the smallest class containing $\overline{0}$ and closed under the operation $\overline{S}$ (exercise).

We have finally arrived at a pure-set-theoretical definition of natural number. In fact, we may, from a formal point of view, forget everything in this section so far, and start with the definition below. We drop the bars from above $\mathbb{N}$, 0 and S.

Definition 4.4 The class $\mathbb{N}$ of natural numbers is defined by

$x \in \mathbb{N} \Leftrightarrow x$ belongs to every class X that contains $\emptyset$ (called also zero, 0) and that is closed under the operation S (see (1), with S for $\overline{S}$): if $x \in X$, then $Sx \in X$.

It turns out that the set-construction principles so far are not sufficient to ensure that $\mathbb{N}$ is a set. We have one last such principle as a postulate:

(Axiom of Infinity) The class of natural numbers is a set.

Proposition 4.5 We have $0 \in \mathbb{N}$, and that $\mathbb{N}$ is closed under successors: if $n \in \mathbb{N}$, then $Sn \in \mathbb{N}$. Moreover, $\mathbb{N}$, 0 and S so defined form a Peano system. Also, any element of a natural number is again a natural number.

Proof: exercise (hint: we use 2.4(ii)).

It is clear from our preliminary discussion that the ordinary ordering relation $<$ on $\mathbb{N}$ should coincide with $\in$: $k < n \Leftrightarrow k \in n$. Namely, $\overline{n}$ having been defined so that $\overline{n}$ is the set of all $\overline{k}$ for $k < n$, we do have $k < n \Leftrightarrow \overline{k} \in \overline{n}$ in that context. Therefore, we define $<$ on $\mathbb{N}$ as $\in$: for $k,n \in \mathbb{N}$, we put

$$k < n \underset{df}{\Leftrightarrow} k \in n.$$

Of course, we now have to prove required properties of the relation $<$ so defined (by the way, formally:

$$< = \{\langle k,n\rangle : k,n \in \mathbb{N} \text{ and } k \in n\}).$$

Instead of doing this directly, we generalize the concept of natural number. For motivation, we turn to the discussion of the sequence $\langle \mathbb{V}_\alpha\rangle_\alpha$ at the end of Section 2, except that now we are interested only in the indices α of this sequence. We felt compelled to introduce the index ω as a 'stage of construction' coming right after all the natural numbers. Following the principle that "every natural number is the same as the set of the previous natural numbers", we try to identify ω with the set of all natural numbers, i.e. in fact $\omega = \mathbb{N}$. Continuing in the same spirit, we put $\omega+1 = \omega\cup\{\omega\} = S\omega$, etc. The indices so obtained are particular ordinal numbers. The spirit of the above is that "any ordinal should be equal to the set of all preceding ordinals". This entails, in particular, that every ordinal should be a <u>transitive set</u>; a transitive set x is one for which $z \in y$ and $y \in x$ imply that $z \in x$ (note that, for ordinals, $y \in x$ should mean 'y precedes x'; $z \in y$ means 'z precedes y'; so certainly then 'z precedes x').

<u>Definition</u> 4.6 The class of ordinals, $\mathbb{O}\mathrm{rd}$, is the smallest class X satisfying the following:

whenever x is transitive and $x \subset X$, then $x \in X$.

(In other words

$$\mathbb{O}\mathrm{rd} = \{x : x \in X \text{ for all } X \text{ such that for all transitive } y,\ y \subset X \text{ implying } y \in X\}.$$

Under the latter definition, it is easy to show that, indeed $\mathbb{O}\mathrm{rd}$ satisfies the first form of the definition.)

<u>Proposition</u> 4.7 (i) Every element of an ordinal is an ordinal.

(ii) Every ordinal is a transitive set.

<u>Proof</u>: <u>exercise</u>.

<u>Proposition</u> 4.8 0 is an ordinal, and if α is an ordinal, so is $S\alpha$. As a consequence, all natural numbers are ordinals.

<u>Proof</u>: <u>exercise</u>.

<u>Proposition</u> 4.9 (Trichotomy of $\in$ on $\mathbb{O}\mathrm{rd}$)

For any ordinals α and β, either $\alpha \in \beta$ or $\alpha = \beta$, or $\beta \in \alpha$.

<u>Proof</u>: We are going to use $\in$-induction; note that the assertion contains two variables. Let us abbreviate by $P\alpha\beta$ the assertion

"either $\alpha \in \beta$, or $\alpha = \beta$, or $\beta \in \alpha$".

$P\alpha\beta$ can be read: "α and β are comparable".

Let $Q\alpha$ stand for:

"for all ordinals β, $P\alpha\beta$".

We show $Q\alpha$ by ϵ-induction. Thus, we fix an ordinal α, and we assume that for all $\alpha' \in \alpha$ (any such α' is an ordinal!), we have $Q\alpha'$. It remains to show that $Q\alpha$ holds, i.e. that for all $\beta \in \mathrm{Ord}$, we have $P\alpha\beta$. Since the latter is an assertion concerning a fixed α, and all β, we regard $P\alpha\beta$ as an assertion <u>on</u> β and prove it by ϵ-induction <u>on</u> β. (Thus, we have an induction on β <u>inside</u> the induction on α.) Accordingly, we now fix another ordinal β as well and we assume that $P\alpha\beta'$ holds for all $\beta' \in \beta$. To summarize our assumptions: we have particular ordinals α and β; we have $Q\alpha'$ for all $\alpha' \in \alpha$, i.e.

$$\text{for all } \gamma \in \mathrm{Ord},\ P\alpha'\gamma \qquad (\alpha' \in \alpha) \tag{2}$$

(note the change of the variable β in the above to γ; β has now been reserved for the second fixed ordinal), and also,

$$P\alpha\beta' \qquad (\beta' \in \beta). \tag{3}$$

Our task is to establish $P\alpha\beta$ from the assumptions (2) and (3).

To complete the proof, we now distinguish some cases. First we make the assumption that $\alpha \not\subset \beta$ (<u>Case</u> 1). This means that there is $\alpha' \in \alpha$ such that $\alpha' \notin \beta$. By (2), for $\gamma = \beta$, α' and β are comparable. Since $\alpha' \notin \beta$, we must have either $\alpha' = \beta$, or $\beta \in \alpha'$. In the first case, by $\alpha' \in \alpha$ we have $\beta \in \alpha$, hence β and α are comparable as required. In the second case, we have $\beta \in \alpha'$ and $\alpha' \in \alpha$. Hence, since α is transitive, $\beta \in \alpha$ again, hence $P\alpha\beta$ holds too. This completes the proof in Case 1.

Case 2) has the assumption that $\beta \not\subset \alpha$. The proof that now $\alpha \in \beta$ is similar to the above, and it uses (3) instead of (2); the details are left to the reader.

To finish the proof, we notice that after Case 1 and Case 2, we have the only remaining possibility that both $\alpha \subset \beta$ and $\beta \subset \alpha$ hold. By extensionality, now $\alpha = \beta$; hence again, $P\alpha\beta$. □

For ordinals α and β, we write $\alpha < \beta$ for $\alpha \in \beta$. Notice that in the last proposition we established the so-called trichotomy property of the relation $<$ on $\mathbb{O}\mathrm{rd}$:

$$\text{either } \alpha < \beta, \text{ or } \alpha = \beta, \text{ or } \beta < \alpha;$$

and, in particular, the same for natural numbers. Also note that the transitivity of ordinals amounts to transitivity:

$$\alpha < \beta \text{ and } \beta < \gamma \text{ imply } \alpha < \gamma;$$

again, as a consequence, this holds for natural numbers as well.

Note also that irreflexivity

$$\alpha \nless \alpha$$

is a special case of 2.4(i).

The last three properties of the relation $<$ qualify it to be called a <u>linear ordering of</u> $\mathbb{O}\mathrm{rd}$.

Also note the following: if α, β are ordinals, then

$$\alpha \subset \beta \Leftrightarrow \alpha \leq \beta$$

where $\alpha \leq \beta$ abbreviates "$\alpha < \beta$ or $\alpha = \beta$" (<u>exercise</u>).

<u>Proposition</u> 4.10 If α is an ordinal, $S\alpha$ can be characterized as the smallest ordinal β such that $\alpha < \beta$

(i.e., $\alpha < S\alpha$, and whenever $\alpha < \beta$, we have $S\alpha \leq \beta$).

<u>Proof</u>: Suppose that $\alpha < \beta$. By trichotomy, $S\alpha$ and β are comparable; hence it suffices to show that $\beta < S\alpha$ is not possible. If $\beta \in S\alpha$ then either $\beta \in \alpha$ or $\beta = \alpha$. But both possibilities are incompatible with $\alpha \in \beta$ by 2.4. □

As we already mentioned, $\mathbb{N}$ itself is an ordinal; formally, this is because of the Axiom of Infinity and the last part of 4.5. In the context of ordinals, $\mathbb{N}$ is denoted by ω. The sequence of ordinals continues with $S\omega, SS\omega, \ldots$

Every natural number except 0 is of the form $S\alpha$ (see 4.1'). ω is the first example of a <u>limit ordinal</u>: one which is not 0 and not a successor either. Put in another way, a limit ordinal λ is one that is different from 0 and for which $\alpha < \gamma$ implies $S\alpha < \gamma$; this follows from 4.10.

Let X be any <u>set</u> of ordinals, and consider $\alpha =_{df} \cup\{S\beta : \beta \in X\}$. By 'replacement' and 'union', α is a set. Since every $S\beta$ $(\beta \in X)$ is transitive, it immediately follows that α is transitive. Since each member of each $S\beta$ is an ordinal, each member of α is an ordinal. It follows that α is an ordinal. For $\beta \in X$, $\beta < S\beta \leq \alpha$ (since $S\beta \subset \alpha$), thus $\beta < \alpha$. But also, α is the least ordinal α' such that $\beta < \alpha'$ for all $\beta \in X$ (we say that α is the <u>least strict upper bound</u> of X, and write $\alpha = \ell.s.u.b.(X)$): if $\beta < \alpha'$, then $S\beta \leq \alpha'$, hence $S\beta \subset \alpha'$; thus if $\beta < \alpha'$ for all $\beta \in X$, then $\alpha = \cup\{S\beta : \beta \in X\} \subset \alpha'$, i.e. $\alpha \leq \alpha'$. We have obtained

Proposition 4.11 For any set of ordinals X, $\ell.u.b.(X)$ S_1 exists and it is an ordinal. □

Proposition 4.12 Ord is not a set.

Proof: exercise.

Proposition 4.13 (Characterization of ordinals).

A pure set x is an ordinal if and only if it satisfies the following two conditions:

(i) x is transitive,

(ii) x is trichotome: for all $y \in x$ and $z \in x$, either $z \in y$ or $z = y$ or $y \in z$.

Proof: exercise.

Proposition 4.14 (Characterization of natural numbers).

A pure set x is a natural number if and only if it is an ordinal, and in addition, the following holds:

(iii) for all $y \in x \cup \{x\}$, either $y = 0$, or $y = Sz$ for some z.

Proof: exercise.