The Foundations of Mathematics; an Overview at the Close of the Second Millennium

Willliam S. Hatcher,
Université Laval, Québec, Steklov Institute of Mathematics,
St. Petersburg, Russia and Chair, Department of Ethics,
Landegg Academy, Switzerland

An intellectual tradition extending back at least to the time of Plato attributes various special qualities to mathematical knowledge. Within this tradition, mathematics may be viewed, for example, as more exact, certain, objective, universal, abstract, formal, or useful than other kinds of knowledge. The study of the foundations of mathematics examines the origin or genesis of mathematical ideas and methods, as well as the structure and organization of the mathematical corpus, primarily with the goal of understanding, explaining or justifying the perceived special qualities of mathematical knowledge.

Thus, in foundations, mathematics itself becomes the object of study. So conceived, foundational studies are partly philosophical and partly technical (mathematical).

Although the study of mathematical foundations has always been a recognizable part of mathematics and philosophy, it is only since about 1900 that foundational study has emerged as a relatively independent discipline with its own methods, techniques and goals. These modern developments have witnessed a sharpening of the philosophical issues relating to mathematics as well as a refinement of the techniques and methods used in foundational study. The present article will therefore concentrate on the modern period, but give sufficient historical background to allow for an adequate understanding of modern developments. Accordingly, the article has the following structure:

THE BASIC ISSUES

The question of what it means to know mathematically is clearly a part of the larger philosophical question of what it means to know generally. Foundational study therefore seeks to understand both the intrinsic nature of mathematics and the status and role of mathematics within the overall scientific and philosophical enterprise.

The abstract nature of mathematics.

Mathematics is traditionally conceived as the science of space (geometry) and of quantity (arithmetic). Empirical observation is thus an obvious source of many mathematical ideas (put one apple together with another, and we have in fact two apples). However, mathematics itself proceeds by the contemplation and study of abstract (nonphysical), ideal entities, i.e., entities that have no exact counterpart in observable, physical reality (e.g., the infinite and perfectly regular lines and planes of Euclid's geometry). A basic concern of foundational study is to determine the nature of these entities and the extent to which it is legitimate to attribute objective existence to them.

Three classic schools of foundational study arise from three possible answers to this question. *Platonism* or *realism* holds that mathematical entities have objective existence on a par with other such mind-independent entities as stones and stars. In this view, mathematical knowledge is the knowledge of these mathematical entities, a knowledge that is discovered but not invented by the human mind. *Intuitionism* or *constructivism* holds that the ideal objects of mathematics exist only within the human mind, arising as mental constructions based on our observation of physical approximations to the ideal. Mathematical knowledge is therefore viewed as partly discovered (through observation) and partly invented by creative intellectual acts. *Formalism* or *nominalism* holds that the objects of mathematics have no existence whatsoever but are only helpful mental fictions that enable us to generate certain useful, though purely conventional, rules for the manipulation of symbols in various contexts. We may have used these rules implicitly and intuitively before formulating them explicitly. Their explicit formulation is the *formalization* of mathematics.

Everyone agrees that mathematical activity involves all three processes: the contemplation of abstractions, the generation of mental constructions, and the explicit formulation of rules for symbolic manipulation. The philosophical differences arise when we consider the question of the relative status of these activities and the extent to which they comprise all or part of mathematics.

The special status of mathematical knowledge.

Each of the three basic philosophies of foundations has its own characteristic explanation of the perceived special qualities of mathematical knowledge. For example, Platonists would ascribe mathematical certainty and exactness to the stability of mathematical objects, which are held to be nonphysical, absolute, and unchanging. In contrast, empirical knowledge is less certain or exact because the physical world is in a state of continual flux. Intuitionists would tend to attribute mathematical certainty and exactness to the degree of control we exert over mathematical objects: since they are explicit constructions of our minds, we can manipulate them freely and know them certainly. In this view, mathematics is exact because it contains only what we deliberately put into it. Formalists would hold that mathematical certainty and exactness derive from the formal and explicit character of mathematical rules. For the pure formalist, the goal is always complete formalization, which achieves total objectivity in

that formal rules can, in principle, be executed by a machine that is utterly devoid of human subjectivity and its vagueries.

Logic and mathematics.

Basic to foundational study is the central role that (deductive) logic has played in the development, organization and articulation of mathematics. Some, for example the French structuralist school of N. Bourbaki, consider the most distinctive feature of mathematics to be the extent and manner that mathematics uses deductive logic. Others (e.g., Bertrand Russell) have gone further and proposed a mathematical *logicism,* which holds that mathematics is (or reduces to) logic.

In opposition to this are mathematicians (e.g., René Thom) who have insisted that (geometrical) intuition is the most fundamental aspect of mathematical activity. But even those mathematicians who give great value to intuition recognize that logic is an inextricable aspect of mathematical activity. Indeed, logical methods have often succeeded in validating mathematical principles that appeared quite unnatural intuitively.

Computability theory or *algorithmics* is a limited but extremely useful expression of deductive logic in mathematics. An *algorithm* (the term is derived from the name of the 9th century Arabic mathematician Al-Khwarizmi) is a finite set of instructions that can be carried out mechanically in a finite number of discrete steps. Modern computability theory is based on a precise, mathematical formulation of the notion of an algorithm, first developed by the English mathematician A. Turing in the 1930's.

The programmed instructions that drive electronic computing devices are examples of algorithms, and the increasing sophistication and availability of these computers has considerably enhanced the status of computability theory, making it one of the most active branches of contemporary mathematics. Pure formalists (and some constructivists) would hold that abstract mathematics in general, and deductive logic in particular, are truly useful only when they yield algorithms.

In any case, the special role that deductive logic plays in mathematics has given rise to a particular method that may be said to characterize mathematics.

THE AXIOMATIC METHOD AND ITS ORIGINS

Logic and reason have played a major role not only in mathematics but also in science and in discursive philosophy. However, there are important differences between the way mathematics uses logical techniques and the way other disciplines use them.

The empirical (or natural) sciences make equal use of both deductive logic, which is a movement of thought from general to particular (analysis), and inductive logic, which is a movement of thought from particular to general (synthesis), proceeding by an alternation of inductive and deductive moves. Induction is used to establish certain general principles that then form the basis for chains of deduction. It is usually rare to encounter extremely long chains of deductive reasoning in the empirical sciences, especially in the early stages of their development.

Long deductive chains are also rare in traditional philosophy. This is because the more *a priori* method of philosophy (particularly metaphysics) tends to multiply the philosophical assumptions, thereby reducing the necessity of engaging in long chains of pure deduction. The philosopher more naturally dedicates his efforts to finding and justifying the underlying principles of his subject than to the technical task of generating extensive and complex deductive chains (though this has changed somewhat in the modern period, primarily due to the influence of mathematics on philosophy).

However, it is characteristic of mathematics that deduction takes priority over induction, that extremely long deductive chains are used, and that the initial assumptions are reduced to the barest minimum rather than multiplied. Thus, it is principally to mathematics that we owe the *axiomatic method,* which consists in organizing a large body of knowledge by explicitly deducing every single proposition from a few explicitly designated assumptions. The assumed propositions are called *axioms* and the deduced (or derived) propostions *theorems.* According to the axiomatic method, inductive logic is relegated to a purely informal use. It may serve to suggest to the mathematician that a certain proposition is true and therefore potentially deducible from the axioms, but the proposition will be accepted as justified only when an explicit deduction of it has been given (or shown to exist), never on the basis of informal, inductive reasoning alone. Moreover, the axiomatic method deliberately seeks an economy of thought, and does not countenance the easy multiplication of assumptions in the manner of philosophy. In particular, if it is discovered that a given axiom p can be deduced from the other axioms, then the proposition p becomes a proved theorem and is deleted from the set of axioms.

Thus, an *axiomatic system* S has at least the five constituents (L, P, Ax, R, Th), where L is an explicitly formulated language, P is a collection of propositions (statements) of L, Ax is the collection of assumed propositions, R the deductive rules, and Th the derived propositions. A deductive chain (or proof) in S is an ordered list $p_1, p_2, \ldots, p_n$ of propositions such that each proposition in the list (each *line* of the proof) is either an axiom or else derived from previous propositions of the list according to the rules R. A theorem t is precisely a proposition of L for which there exists a deductive chain (derivation) $p_1, p_2, \ldots, p_n = t$, ending with t. Notice that every axiom p is a theorem, for which p is a one-line proof of itself.

Because the rules and principles R of deductive logic preserve propositional truth, all theorems Th of an axiom system S are true if the axioms Ax of S are true. Thus, when once the axioms of a system S have been verified to be true, the verification that a proposition p of L has a valid proof in S also constitutes a verification that p is

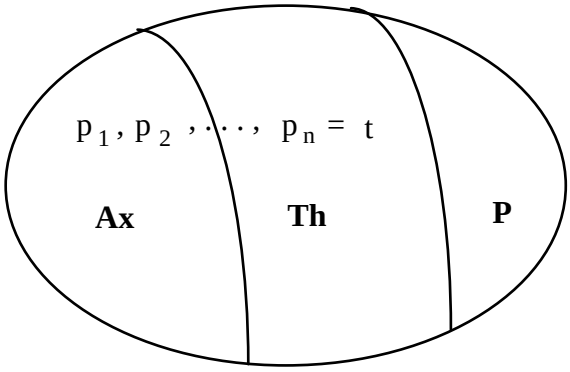
true. (However, a proposition p of L can be true without having a valid deduction from the particular axioms of S.)

To validate or verify a proof $p_1, p_2, \ldots, p_n$ in S means to go through the proof step by step and ascertain that each line p_i does indeed follow from the previous lines $p_1, p_2, \ldots, p_{i-1}$ according to the deductive rules and principles R. Usually the rules R of logic are formulated in such a way that checking any given step in a deductive chain is a simple matter. Thus, finding a valid proof is a highly creative and possibly quite difficult task (and one for which there are, in general, no rules) but verifying that a purported proof is indeed valid is, in principle, an easy, rule-based task (though possibly tedious and time-consuming).

Typically, a well-developed axiomatic system is extremely complex, containing thousands of known and validated theorems. But any question of the truth of any of these theorems reduces to the question of the truth of the axioms on which they ultimately depend. Thus, the organization of knowledge into an axiomatic system puts the burden of truth on the axioms of the system, rather than distributing the truth burden throughout the body of knowledge as with natural science and philosophy.

This feature of the axiomatic method is one of the main justifications for the immense effort and time mathematicians invest in the search for deductive proofs. Another reason is the following: since the truth of the axioms implies the truth of the theorems, the falsity of even one theorem implies the falsity of at least one of the axioms. Of course, when the assumed propositions of a system S are sufficiently simple to be obviously true, there is no problem. But, the more powerful the axioms, the more complex and abstract they tend to be, often making them neither obviously true nor obviously false. When this is the case, then one way of detecting their falsity is to exhibit an obviously valid proof of an obviously false theorem. In this way, extensive deduction serves both as an enrichment of a true axiom system and as a protection against a false one.

Constituents of an axiomatic system S.



The collection P of all propositions of the language L contains the subcollection Th of all theorems. Each theorem t is deduced from the further subcollection Ax of all axioms by a finite number of applications $p_1, p_2, \ldots, p_n = t$ of the rules R.

Since the rules R preserve truth, the theorems Th are all true if all of the axioms Ax are true. Thus, if any theorem t is false, then at least one of the axioms is also false, and the set Ax of axioms is then inconsistent, i.e., has no model (see below).

One frequently studies different systems S with the same language and deductive rules but with different sets of axioms. In this case, we refer to the logical rules as the *underlying logic* of these systems.

Although the axiomatic method has been used and applied outside of mathematics, it originated within mathematics and has been carried to a much higher development in mathematics than elsewhere. It can thus be seen as one of the most distinguishing features of mathematics. The axiomatic method also accounts for some of the perceived special qualities of mathematical knowledge such as certainty and exactness.

Euclid's Elements.

Historically, geometry was the first mathematical discipline to be successfully organized as an axiomatic system. Accomplished by Euclid in the fourth century B. C., this achievement was and is astonishing for a number of reasons. To begin with, one would have thought that arithmetic, with its rule-like behaviour, would be a more likely

candidate for early axiomatization. Yet arithmetic was successfully axiomatized only by Richard Dedekind in the late l9th century (see below).

Geometry codifies our intuition of spatial relationships, an intuition that does not appear to lend itself easily to the formulation of exact rules or relationships. Even to conceive of the possibility of axiomatizing geometry was bold enough, but to have accomplished the task in one lifetime is breathtaking. Moreover, the system of Euclid's *Elements* is so sophisticated that no substantial improvement was made in it until the work of M. Pasch and D. Hilbert in the late 19th century. Thus, Euclid's work became the unsurpassed canon of the axiomatic method for over two-thousand years.

Multiple interpretation, consistency and abstact axiomatics.

Even though Euclid's geometry remained essentially unmodified until the modern period, there was nonetheless a certain evolution in the conception of the axiomatic method itself. This evolution was due to the gradual realization that a given set of axioms could be valid under several different interpretations, i.e., that axioms, if appropriately interpreted, could be equally true of different realities. To see this, consider the following true proposition of Euclidian plane geometry:

(A) Any line is completely determined by any two non-identical points that lie on the line.

The proposition A has the following general form: "Any object x of type Ln is completely determined by any two objects y and z of type Pt that are not in the relation Id and that bear the relation Ot to x." Now, let us reinterpret the types and relations Ln, Pt, Id and Ot as follows: Ln is now the category (type) 'point', Pt is now the type 'line', Id is now the relation 'parallel' and Ot is the relationship 'passes through'. Under this interpretation, proposition A now become proposition B: "Any point x is completely determined by any two lines y and z that are non parallel and that pass through the point x." The proposition B is also true in Euclidean plane geometry though the meaning of A and B are quite different. Yet A and B "have the same form" in that we have reinterpreted only the specific types and relations. We say that we have reinterpreted the *non-logical* (or *specific*) parts of A and that we have maintained or preserved its *logical* (or *general*) parts (e.g., such terms as 'any', 'object', 'type', 'completely determined', 'not' or 'two'.)

Let us consider yet another interpretation: Pt means 'human zygote', Ln means 'human gamete', Id means 'same sex', and Ot means 'sexually generated'. Under this interpretation, A now means C: "Any human zygote x is completely determined by any two human gametes y and z of the opposite sex that have sexually generated x."

Again, C is true, but of a completely different reality, having nothing to do with geometry. Thus, under the appropriate interpretation of its non-logical parts, the same statement A is true of three different realities. However, these three realities can be distinguished by other properties. For example, there is not a unique pair of distinct points that determine a given line (or a unique pair of non-parallel lines that determine a given point), but every human zygote is determined by only one pair of gametes. However, points and lines do not share all the same properties either. For example, a plane can be passed by any three points, but a plane cannot be passed through any three lines.

We have given examples of reinterpretations that preserve truth, but clearly many reinterpretations will be false. For example, under the first interpretation just

reinterpret the category Ln as 'point', (keeping the other types and relations unchanged) and the resulting statement is false. Indeed, arbitrary reinterpretations are more likely to be false than true. We now summarize the points illustrated by this example:

1. The specific terms (types and relations) of a statement can be reinterpreted without changing the (syntactical) form or structure of the statement with respect to its purely logical parts.

2. The statement may be true under some of these reinterpretations and false under others.

Any interpretation of the specific terms of a set Ax of axioms which makes all of the axioms true is said to be a *model* of the axioms. A set of axioms is *consistent* if it has at least one model. It is *universally valid* if true under all possible interpretations of its specific terms.

Now, the rules R of deductive logic are *formal* in that they preserve the syntactical form (logical structure) of propositions, and in such a way that all of the theorems Th of an axiomatic system S are true in any model of the axioms Ax of S. Moreover, for *first-order* systems (systems based on a wide class of so-called first-order languages), the rules of deductive logic are *complete* in the precise sense that a formal logical contradiction (a proposition of the form
"p and not-p") can be deduced from any inconsistent set of axioms (i.e., any set of axioms that has no model). The converse also holds, since any statement of the form "p and not-p" is false under every interpretation of p (i.e., every interpretation that preserves the meaning of the logical terms 'and' and 'not'). Thus, if a formal contradiction t is provable as a theorem of a system S, then the axioms of S are false under any interpretation and hence have no model. In other words: the axioms Ax of a first-order system S are inconsistent (have no model) if and only if a formal contradiction "p and not-p" is provable as a theorem of S.

It follows from the above that, for first-order systems, the consistency of a given set of propositions does not depend on the meaning of the specific terms in the propositions; it depends only on the form or structure of the propositions with regard to their logical parts. Moreover, formal deduction alone can be relied upon to detect inconsistency.

According to the logical rules of most systems (and first-order systems in particular), any proposition q can be deduced from a formal contradiction. Thus, a further property of an inconsistent system S is that every proposition of S is a theorem of S. In other words, anything whatsoever can be proved in an inconsistent system.

These results on logic, which have only been achieved in the twentieth century, have allowed for a much more general form of the axiomatic method, called the *abstract* (or *formal*) axiomatic method. This method consists in leaving the specific terms of an axiomatic system uninterpreted (or "undefined") from the beginning. The only requirement is that the axioms be consistent, and thus true of some reality, but we no longer require that this reality be specified. In the formal axiomatic method, the role previously played by truth is now played by consistency, and consistency depends only on the syntactical form of the axioms with respect to their logical parts, not on their specific content (meaning) under a given interpretation.

Thus, the same abstract axiomatic system S may be used to study concrete, existing realities and also abstract, logically possible realities. The abstract axiomatic method thereby gives enormous power and flexibility to mathematics, allowing us to apply (by reinterpretation) the same system S (the same body of mathematical results) in many different contexts. For example, it may turn out that the mathematical theory, say,

of languages, of genetics and of machine computation is essentially the same. In such a case, each theorem t of our system S will have an interpretation as a true proposition in each of the respective models.

The abstract axiomatic method also allows us to sharpen our understanding of the nature and status of mathematics within the overall scientific enterprise. At one extreme, the natural sciences study concrete realities that actually exist (or that may exist under given physical conditions). At the other extreme, pure logic studies all possibly existing realities, whether abstract or concrete, actually existing or not. Mathematics lies between these extremes. It studies both concrete and abstract realities. However, mathematics (it is now generally agreed) is not pure logic. Mathematics is not interested in all possibly existing structures, because many of these structures are *useless* to us (for many different reason, e.g., triviality or gratuitous complexity). Thus, we may sum up the relationship between mathematics and logic by saying that logic has general content but no specific content, while mathematics has both general (logical) content and specific content (e.g., truths about spatial relationships or numerical calculations).

The criterion of usefulness gives a certain pragmatic, normative aspect to mathematical activity. Mathematics tries to solve real problems by providing useful theories of these problems in appropriate axiomatic systems. The mathematician does not, therefore, waste time exploring systems that are arbitrary or that seem to hold no promise of solving the problems at hand. Such systems may be explored by philosophy, and perhaps with no initial motivation beyond idle curiosity, but the mathematician will examine them only when he has some reason to believe they will be useful in solving mathematical problems.

Finally, the foundational study of mathematics explores those structures or realities that are relevant to solving *foundational problems*, i.e., those problems relating to the genesis and nature of mathematics itself. The systems used to study foundational problems are called *foundational systems.* These are abstract axiomatic systems of great power and generality, which contain many branches of mathematics as subsystems. We want now to examine the major foundational systems that have appeared in the modern period.

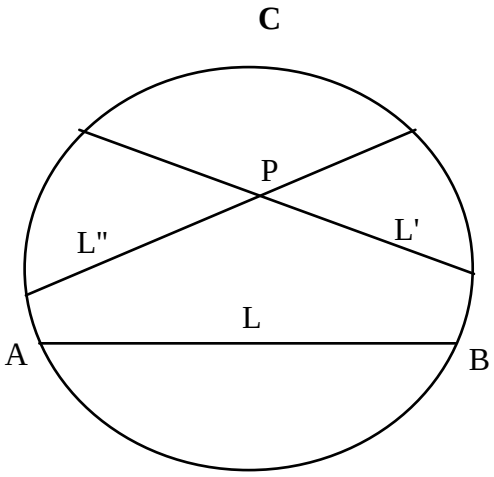
The historical origin of multiple interpretation and thus of the abstract axiomatic method was the appearance of non-Euclidean geometries in the l9th century.

1. Euclidean geometry.

In Euclidean plane geometry, [1] any two distinct points determine a unique line and [2] any two different, intersecting lines determine a unique point. *Hyperbolic* and *parabolic* geometry maintain these properties, but differ from Euclidean geometry with respect to a third property: According to Euclid's axioms, [3] one and only one line L' that does not intersect a given line L can be drawn through any given point P exterior to L.

2. Hyperbolic geometry.

In hyperbolic plane geometry, the notions "point", "line" and "plane" are reinterpreted as follows: The plane is no longer infinite in extent, but consists of the interior of a fixed circle C. Thus, the points of hyperbolic geometry are those Euclidean points that lie inside the circle C. Hyperbolic lines are the chords AB of C, excluding the endpoints A and B (which lie on the circle and thus outside the plane). Under this interpretation, there are many lines L' and L" that may pass by a point P, exterior to a given line L, without intersecting L. But this system satisfies [1] and [2] and, indeed, all the other axioms of Euclidean geometry except [3].



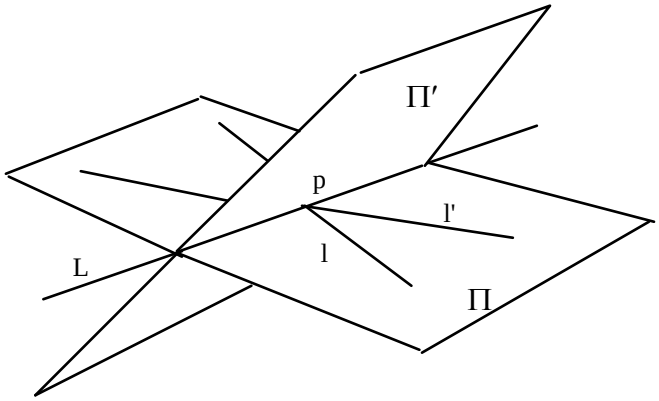
Hyperbolic geometry was first discovered (independently) by the Hungarian J. Bolyai and the Russian N. I. Lobachevsky (c. 1825).

2. Parabolic geometry.

In parabolic geometry, some particular point p in Euclidean three-dimensional space is chosen. The parabolic plane is then the set of all Euclidean lines through p, and a parabolic line is a Euclidean plane containing the point p. Since the intersection of any two distinct Euclidean planes is a line, the intersection of any two distinct parabolic lines Π and Π' will be a parabolic point L (i.e., a Euclidean line through p). Also, any

two distinct parabolic points l and l' will determine a unique parabolic line Π (i.e., a Euclidean plane Π containing p). Parabolic geometry thus satisfies [1] and [2].

However, in parabolic geometry, any two parabolic lines Π and Π′ will intersect (since both must contain the point p). Thus, in parabolic geometry, no line whatever can be drawn that does not intersect a given line Π. Hence [3] fails in parabolic geometry but in a different way than in hyperbolic geometry. Also, as with hyperbolic geometry, parabolic geometry satisfies all of the other axioms of Euclidean plane geometry except [3].



Parabolic geometry originated with B. Riemann in 1848, and was part of a comprehensive treatment of geometry that included parabolic, hyperbolic and Euclidean geometry as special cases.

FOUNDATIONAL SYSTEMS

It has taken several thousand years for the axiomatic method to evolve and develop into its current form, which has become the primary technique of modern foundational study. However, parallel to the evolution of the methods of mathematics was a similar evolution of its content. We need to sketch the latter before undertaking a direct discussion of the principal modern foundational systems.

The emergence of modern mathematical analysis.

Though the Egyptians, the Sumerians, the Babylonians and the Chinese all developed various rudimentary mathematical systems dealing either with geometry or arithmetic, the efflorescence of ancient Greece (roughly the period 500-300 B.C. of the Athenian city-state) represents without doubt the summit of premodern mathematical (and philosophical) development. The Greeks saw all of mathematics through the prism of geometry. Numbers represented geometrically defined quantities such as lengths, areas or volumes, and the manipulation of numbers was primarily through geometrical constructions.

The manipulation of numbers as pure quantities derives not from the Greeks but from the Indic and Arabic cultures. The name — algebra — eventually given to the discipline that codifies and axiomatizes the rules for these manipulations derives from the title of the book *Al-Jabr* published in the 9th century by the Muslim mathematician Al-Khwarizmi.

Modern mathematics owes to the Indic-Arabic cultures at least two outstanding contributions. The first is the highly flexible number system based on positional or place value, which uses only a finite number of ciphers (digits) to give a unique numeral (name) for each member of the infinite set **N** of *natural numbers* (the *nonnegative integers* 0, 1, 2, . . .). Each numeral is itself a finite list of ciphers in which the value of each cipher c is multiplied, according to its position in the list, by an appropriate power of a fixed base b. Because the base b was most often given the value ten, the system of positional value has become popularly known as the *decimal* system, but b can in fact be given any fixed value greater than one.

The system of place value leads to highly efficient algorithms (see above) for the four arithmetic operations of addition, multiplication, subtraction and division (and for other operations as well); whereas, systems not based on place value (e.g., the well known system of Roman numerals) do not seem to allow for algorithms of comparable power and efficiency. Indeed, so sophisticated and flexible is the system of Arabic numerals that no significant modification of it has been necessary even for modern computers, which are constructed on a binary system of place value and whose programmed computations are based directly on the algorithms generated by the Arabic system.

The Greeks studied the system **P** of *positive integers* 1, 2, 3, . . . , which is just the system **N** without 0. However, enumeration by positional value makes unavoidable and essential use of 0. Thus, although the Greeks did succeed in devising some effective algorithms (e.g., the well-known algorithm devised by Euclid to find the greatest common divisor of two given positive integers), they never elaborated a number system as such, and their approach to arithmetic remained heavily geometrical: the sequence 1, 2, 3 . . . of positive integers was viewed as the end-to-end repetition, along a fixed axis, of a unit line segment of fixed length. Since a line segment cannot have zero length, it is easy to understand why the Greeks omitted 0 from their number system and thereby failed to devise a system of positional value.

The second major Arabic contribution to mathematics was the development and codification of rules and techniques for manipulating numbers together with *variables* or *indeterminates*, i.e., non-numerical symbols x_i standing for arbitrary numbers (called the *values* of the variables). An expression $P(x_1, \ldots, x_n)$ involving both numerals (called the *constants* or the *coefficients* of the expression) and variables $x_1, \ldots, x_n$, which are added, multiplied or subtracted, is now called a *polynomial*. Many practical (and even theoretical) problems of mathematics can be resolved by finding those values $a_1, \ldots, a_n$ of the respective variables $x_1, \ldots, x_n$ that reduce to zero the expression $P(a_1, \ldots, a_n)$, obtained from
$P(x_1, \ldots, x_n)$ by replacing each variable x_i by the corresponding value a_i and applying the operations of the expression to these substituted values. In this case, we say that the list of values
$a_1, \ldots, a_n$ are *zeroes* or *roots* of the polynomial and we write:
$P(a_1, \ldots, a_n) = 0$. We also say that the list $a_1, \ldots, a_n$ constitutes a *solution* of the polynomial equation $P(x_1, \ldots, x_n) = 0$. (By subtraction, any equation P = Q between two polynomials is logically equivalent to the simpler form P - Q = 0, since the difference P - Q of two polynomials is also a polynomial).

Even though the coefficients of a polynomial are natural numbers, the solutions do not have to be natural numbers. For example, the solution to the equation $3x-2 = 0$ is 2/3, which is a rational fraction (ratio of two integers). Similarly, solutions to the equations $x + 2 = 0$, $x^2 - 2 = 0$, and $x^2 + 1 = 0$ include, respectively, the negative integer -2, the irrational number $\sqrt{2}$, and the imaginary number $i = \sqrt{-1}$. (Of course, the search for zeroes of a polynomial can be deliberately restricted to integer solutions. In that case, we speak of a *diophantine equation.*)

Thus, the search for solutions to polynomial equations with natural number coefficients leads naturally to a wider class of numbers, the algebraic numbers. In modern terms, the *algebraic numbers* are precisely those numbers that can occur as roots of some polynomial P(x), in a single variable x and with natural number coefficients. In a somewhat restrictive sense of the term, *algebra* can be considered the discipline that has, as its principal object of study, the algebraic numbers.

Just as the natural numbers themselves can be codified by the Arabic numeral system of place value, so the algebraic numbers can be codified or named by the polynomials of which they are roots. Indeed, a numeral in the Arabic system is precisely a "one-variable polynomial" in which the base b is substituted for the variable x.

When algebraic numbers are added, multiplied, subtracted or divided (except by zero) the result is always an algebraic number. We say that the algebraic numbers are *closed* under each of the four basic operations of arithmetic. In modern mathematics, any system closed under these operations is called a *field.* Though the algebraic numbers constitute a particularly natural field, there are in fact many fields. For example, the *rational numbers* **Q** (those numbers expressible as a ratio of two integers) and the *real numbers* **R** (those not involving the imaginary number i) also constitute a field. Non-algebraic reals are *transcendental* and non-rational reals are *irrational.* Since all rationals are algebraic, all transcendentals are irrational.

The degree to which geometry and algebra evolved independently of each other is initially surprising, but is probably due to the fundamental difference in the nature of the respective intuitions that generate them. Geometric intuition is synthetic and continuous, perceiving configurations as completed wholes endowed with various global regularities. Algebraic intuition is analytic and discrete, perceiving global

structures as built up by the accretion of distinct quanta, gradually extending local regularities. The beginning of modern mathematics was the fusion of these two intuitions into a single discipline: *algebraic geometry.* This was accomplished in the 17th century by the consummate French philosopher and mathematician, René Descartes.

The Greeks had already observed that geometrical figures, though continuous, can nonetheless be considered as sets of points. For example, a circle of radius l is the *locus* (set) of precisely those points having a fixed distance of one unit from a given fixed point (the center of the circle). Building on this notion, Descartes' method consisted in establishing a correspondence between points and numbers and thus between sets of numbers and geometrical figures (loci). Along a given axis (Euclidean line), each point corresponds to one, distinct real number (algebraic or transcendental). Since a line is one-dimensional, the field of real numbers has a dimension of one (a single number, which corresponds to a point, has dimension zero). The two-dimensional Euclidean plane can be algebraically encoded as pairs of numbers, by taking two axes perpendicular to each other. The generalization to three, and thus to arbitrary finite dimensions is obvious.

In this way, the set of zeroes of a polynomial $P(x_1, \ldots, x_n)$ in n distinct variables determines a set of points — thus a geometrical figure — in n-dimensional Euclidean (Cartesian) space. For example, the set of pairs (x,y) of numbers satisfying the polynomial equation
$x^2 + y^2 - 1 = 0$ corresponds precisely to a circle of radius one, centered at the origin (the intersection of the axes in the plane).

We generalize by allowing real numbers not only as values of variables but also as coefficients in polynomials. Now, for example, every line in the plane has one and only one equation of the form
ax +by+c = 0, where the coefficients a,b and c are real numbers.

Descartes' fusion of geometry and algebra was not only elegant, but extremely fertile for both disciplines. On the one hand, geometrical intuition could be applied to what were previously purely abstract expressions, while on the other hand, the algorithmic and discrete methods of algebra allowed for a quantizing of space.

Within one generation after Descartes' fundamental advance, the unified science of *mathematical analysis* was born though the simultaneous and independent discovery of the calculus by Newton in England and Leibniz on the continent. In trying to understand the transition from algebraic geometry to analysis, observe that the algebra inherited from the Arabs consisted of finitary operations on finite quantities. Newton's approach to the calculus was to add an infinitary operation — *the limit.* Thus, the infinite sum
$1 + 1/2 + 1/4 + \ldots + 1/2^n + \ldots$ can be defined to have the limit 2, even though this value is never actually attained directly as a finite sum of values. Leibniz' approach was to maintain finitary operations but to introduce infinitely small (infinitesimal) and infinitely large quantities, extending the real number field **R** to a hypperreal field **R***. The complete equivalence of these two approaches was definitively established only in 1960 by Abraham Robinson, and constitutes one of the major results of modern foundational studies.

The power of the calculus is that it allows for an analytic study of dynamic processes. Such processes are represented by *functions,* i.e., operations f that associate exactly one value f(x) to each argument (object) x chosen from a given set X. We say that f(x) represents the application of the function f to its argument x. Functions are

often symbolized as f:X $\Rightarrow$ Y, where f is the operation, X is the given set of arguments, and Y is the set of possible values of f. Such a function could represent, for example, temperatures f(t) = r recorded at different moments of time t or the area A = f(d) of a circle having diameter d.

Also, the action of two functions f:X $\Rightarrow$ Y and g:W $\Rightarrow$ X can be combined by the operation of functional composition ö to form a new function h:W $\Rightarrow$ Y, h = göf, whose action on any argument w in the set W is defined by göf(w) = g(f(w)). Thus, functional composition (the left hand side in the above equation) is defined in terms of functional application (the right hand term of the equation). Where f:X $\Rightarrow$ Y, we sometimes use applicative notation assigned to the whole set X, f(X), to represent the set (contained in Y) of all values f(x) of arguments x in X. f(X) is the *image* of X under the functin f.

The two central notions of the calculus are the *derivative*, which represents the instantaneous rate of change of a function at a given point, and the *integral*, which allows for an exact calculation of the portion of a space (e.g., an area or volume) determined (bounded) by a given function. Modern foundational studies have now shown that "set" (deriving from the loci of Euclidean geometry) and "function" (deriving from the calculus) are the two most fundamental notions of mathematics.

The arithmetization of analysis.

In the development of analysis during the years immediately following Newton and Leibniz, geometrical ideas tended to predominate over purely algebraic notions. This was due in part to the retrospective realization that Newton's limit operation had already been successfully used in special cases by the Greek mathematician Archimedes (third century B.C.), whose "method of exhaustion" had led him to calculate correctly certain geometrical limits. However, the analytic work of L. Euler, K. Gauss, A. Cauchy, B. Riemann, and others led to a shift towards the predominance of algebraic and arithmetic ideas. In the late 19th century, this tendency culminated in the so-called arithmetization of analysis, due principally to K. Weierstrass, G. Cantor and R. Dedekind.

Weierstrass developed the theory of *real functions*, i.e., functions f:**R** $\Rightarrow$ **R**, and, among other things, furnished the first example of a real function that is everywhere continuous (the geometrical graph of the function has no breaks) but nowhere differentiable (the instantaneous rate of change of the function does not exist at any point). In an impressive series of papers, Cantor elaborated the theory of infinite sets, including transfinite cardinal and ordinal numbers. In particular, he showed that there were different orders or levels of infinity (see below) the lowest level being that of the natural numbers **N**, which he called *denumerable infinity*. However, it was Dedekind who brought the work of Weierstrass and Cantor to completion by giving abstract, axiomatic characterizations of each of the major number systems in his landmark work *Was Sind und Was Sollen die Zahlen?*, 1888, thereby establishing definitively that mathematical analysis was logically independent of geometry. Of course, geometrical ideas were and are always present and available via Descartes' correspondence between geometry and algebra, but Dedekind's work showed that, though convenient and intuitively useful, such ideas were in no wise logically necessary to the development of analysis. We begin with a sketch of Dedekind's construction and characterization of the natural numbers **N**.

Dedekind and the axiomatization of arithmetic.

Already in the 17th century, Galileo had noticed that infinite sets like the natural numbers admit functions into themselves that are *injective* (one-to-one) without being *surjective* (onto). For example, the function 2n: **N** ⇒ **N**, which associates with each natural number n its double 2n, constitutes a *bijective* (one-to-one and onto) correspondence between the set **N** and the proper subset of all even numbers 2**N**, where 0 «⇒ 0, 1 «⇒ 2, 3 «⇒ 6, etc. However, it is easy to prove by mathematical induction that no finite set has this property. More precisely, every injection f:X ⇒ X of a finite set X into itself must also be surjective, i.e., f(X) = X (every element x of X is of the form x = f(y) for some y in X), and thus bijective. Dedekind saw that this provided an intrinsic definition of an infinite set. Accordingly, a set X is said to be *(Dedekind-)infinite* if there exists some injective function f:X ⇒ X that is not surjective.

Dedekind's treatment of the natural numbers begins by observing that the system of natural numbers (**N**, 0, ß), where
ß: **N** ⇒ **N** is the *successor function* defined by ß(n) = n+1, is Dedekind-infinite because ß is injective but not surjective. Indeed, ß(n) = ß(m) only if n = m, and 0 is not a successor of any natural number (and is, in fact, the only non-successor). By an ingenious construction, Dedekind showed that any Dedekind-infinite set contains the system (**N**, 0, ß) as a subsystem. He furthermore showed that the system (**N**, 0, ß) is completely determined (axiomatically characterized) by three axioms now known as the *Peano axioms* (though G. Peano published them only in 1889 and later acknowledged that he had taken them from Dedekind's 1888 work). We use Peano's symbol '‰' to represent the relationship between an object x and a set X of which it is a member. Thus, where **N** is a nonempty set, 0‰**N** and ß: **N** ⇒ **N**, the Peano axioms are as follows: (1) 0»ß(**N**); 0 is not a successor; (2) ß(n)=ß(m) implies n = m; ß is injective; (3) If X is an inductive subset of **N** then X = **N** (a subset X of **N**, symbolized Xó **N**, is inductive if 0‰X and if X is invariant under ß, ß(X)ó X). In other words, **N** is the smallest inductive set.

Using the technique of *recursive definition* Dedekind also showed how all the usual operations of arithmetic could be uniquely defined in terms of 0 and ß. For example, addition + is the unique binary operation on **N** that satisfies the two recursion equations (i) n + 0 = 0 and (ii) n + ß(m) = ß(n + m). Similarly, multiplication × is the unique binary operation satisfying the two recursion equations (i') n × 0 = 0 and (ii') n × ß(m) = (n × m) + n.

Definition by recursion equations provides not only a complete logical definition of the operation in question, it also gives an algorithm for computing its values. For example, by repeated application of equations (i) and (ii), we can determine that 3 + 5 = 8. Moreover, all of the usual laws of these operations (e.g., the commutative law of addition, n + m = m + n) can be logically deduced from the recursive definitions and the three Peano axioms. Finally, the relation ' >' of 'greater than' between natural numbers a and b can be defined in terms of addition: b > a if and only if a + n = b from some nonzero n‰**N**.

Thus, in one sweep, Dedekind gave both an axiomatic characterization of the natural numbers and a method of constructing them, if given the existence of any Dedekind-infinite set. He showed further that any set of recursion equations, such as those for addition and multiplication above, are a special case of a single recursion scheme, called *simple recursion*, which can be stated as follows: Given any nonempty set S, any designated element a‰S, and any function f:S ⇒ S, then there exists one and only one function h: **N** ⇒ S such that h(0) = a and h(ß(n)) = f(h(n)) (in the box below,

we give a schematic form to the statement of this result). Moreover, the proof of the simple recursion scheme uses only the three Peano axioms (and each is necessary to the proof).

In order to understand the import of this result, we introduce some modern terminology. A function f:S $\Rightarrow$ S, from a nonempty set S into itself, is now called a *(discrete) dynamical system*. Beginning with any element a‰S as an initial value, we can *iterate* the application of the function f applied to a, obtaining the sequence of values a, f(a), f(f(a)), . . . , f(. . .(f(f(a))). . .), . . . , called the *orbit* of a under f. The study of the orbits of elements of a dynamical system has become a major tool in modern analysis. If we introduce the notation $f^n(a)$ for the nth iterate of f applied to a, then the orbit of a is precisely the sequence a = $f^0(a)$, $f^1(a)$, $f^2(a)$, . . ., $f^n(a)$ = h(n), . . ., where h:**N** $\Rightarrow$ S is the unique function whose existence is guaranteed by Dedekind's scheme of simple recursion.

In other words, the natural numbers **N**. together with the successor function ß, constitute a universal dynamical system generated by the single element 0: The axiom of induction tells us that the set **N** is precisely the orbit of 0 under ß and the theorem of simple recursion tells us that any orbit $f^n(a)$ of any element a in any dynamical system f:S $\Rightarrow$ S is the "image" of the orbit **N** of 0 by a unique function h:**N** $\Rightarrow$ S, h(n) = $f^n(a)$ for all n‰**N**.

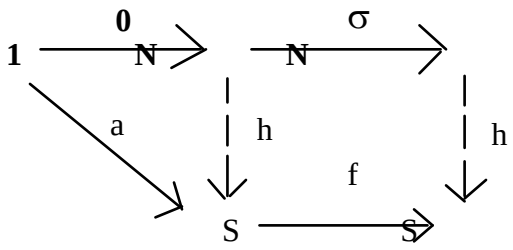
It is now known that any such "universal system" is unique up to isomorphism, meaning that the structure of the system is unique, though the particular set of objects on which we define the structure may vary. Since the universality of such systems resides in their structure we will henceforth speak of universal structures.

The general theorem on the existence (and uniqueness) of universal structures was formulated and proved only in the mid twentieth century — essentially by the Bourbaki group of French mathematicians and the American P. Freyd — but Dedekind did explicitly establish the existence and uniqueness theorem for the case of the universal structure (**N**, 0, ß). (The work of the American G. Birkhoff in the 1930's also contributed to the achievement of the general theorem on universal structures.)

In proving the scheme of simple recursion using only the Peano axioms, Dedekind therefore proved that any model of the Peano axioms is a universal dynamical system. The converse result, namely that any universal dynamical system satisfies the Peano axioms, is also true (without any prior assumption of the existence of an infinite set) but was only explicitly proved in the early 1960's by F. W. Lawvere. Putting this latter result together with those of Dedekind, we now have what might be called the "fundamental theorem" or "core theorem" of modern foundations:

The existence of an infinite set, the existence of a (necessarily unique) model of the Peano axioms, and the existence of a (necessarily unique) universal dynamical system are all logically equivalent (the existence of any one of these entities implies the existence of the others).

The Natural Numbers as a Universal Dynamical System.



The given functions σ:**N** $\Rightarrow$ **N** and f:S $\Rightarrow$ S are represented by solid arrows, while the function h whose unique existence is derived from these data is represented by the broken-bodied arrows. The resulting diagram is said to be commutative, meaning that every way of composing one arrow with another yields an equality. Thus, where 'ö' represents the operation of functional composition (see above), the equalities hö0 = a and höσ = föh hold. (Here, the element 0 of **N** is represented as the target of a function from a one-element set **1** to **N**, and the element a of S is represented in a similar way.) These equalities are just the two recursion equations of the scheme of simple recursion: h(0) = a and h(σ(n)) = f(h(n)), for all n in **N**.

We say that h is a morphism from the structure (**N**, 0, σ) to the similar structure (S, a, f) (see below). The existence and uniqueness of the morphism h from (**N**, 0, σ) to any similar structure (S, a, f) is the mark the universality of the structure (**N**, 0, σ).

This fundamental result has many consequences. It means, for example, that it is not necessary to answer the age-old philosophical question "what is a natural number" in order to do mathematics. One only has to establish the existence of some system satisfying the Peano axioms. However, it also means that classical mathematics is based squarely on the existence of infinite sets.

Once the axiomatization of the system **N** of natural numbers was accomplished, an axiomatic characterization of the other major number systems followed with relative ease, as Dedekind himself showed. Using modern terminology these characterizations are as follows.

A *ring* is an algebraic system that it is closed under addition, subtraction, and multiplication (but not necessarily division). A ring is *ordered* if it has a relation of 'greater than', symbolized by ' >', with 1 > 0, b > a if and only if b-a > 0, and such that the positive elements (those greater than 0) are closed under addition and multiplication. An ordered ring is *well-ordered* if every nonempty subset of its positive elements has a (necessarily unique) smallest element. Then, the (positive and negative) integers **Z** form the unique (up to isomorphism) well-ordered ring.

Recall that a field (see above) is a ring that is also closed under division by nonzero elements. Then, the system **Q** of rational numbers is the unique (up to isomorphism) smallest ordered field.

An ordered field is *complete* if every nonempty set X of its elements that is *bounded above* (there is some element of the field that is greater than or equal to every element in X) has a smallest upper bound. Then, the system **R** of real numbers is the unique (up to isomorphism) complete, ordered field.

Using ideas of P. Erdös, L. Gillman, M. Henrikson, and H. J. Keisler, W. S. Hatcher succeeded (in 1980) in giving an algebraic axiomatic characterization of the so-called minimal ultrapower model of Leibniz' hyperreal number field **R*** (however Hatcher's characterization does assume Cantor's *continuum hypothesis* CH, discussed below).

Frege's system and the paradoxes.

Dedekind's axiomatic characterizations of the major number systems accomplished for analysis what Euclid had accomplished for geometry some two-thousand years earlier. Yet, Dedekind's results raised new questions even as they answered old ones: (1) Is the assumption of the existence of an infinite set — so necessary to Dedekind's constructions — logically justified? Can we actually prove that an infinite set exists and, if so, on what basis? (2) Dedekind's constructions involve Cantor's infinitary set operations (such as infinite set unions and intersections). Are these legitimate? Can the general rules for operating with arbitrary sets be codified and axiomatized? (3) Surely there must be some limits to the ever increasing generality and abstraction of mathematics. What are these limits? How indeed can we be certain that mathematics (in particular, infinite mathematics) is logically consistent? (4) Can we in fact axiomatize all of mathematics in one system?

In his *Grundgesetze der Arithmetik*, published in 1893, G. Frege attempted to provide a positive answer to all of these questions by presenting a formal axiomatic system for the whole of mathematics. In modern terminology, the essentials of Frege's system are as follows. Formula and set (collection of objects satisfying a formula) are the basic notions of mathematics. The basic relationship of all mathematics is the relationship '‰' between an object x and a set y of which it is a member. Thus, the basic formulas of mathematics have the form 'x‰y', where x and y are variables. The general formulas are built up from basic formulas by the logical propositional connectives 'or', 'and', 'not', 'if . . . then - - -', '. . . if and only if - - -', and the quantifiers 'there exists' and 'for all'. The language is thus a first-order language, and the logical rules are those of first-order logic (see above).

The specific principles (axioms) of Frege's system are extensionality, which asserts that sets with the same elements are equal, and comprehension, which asserts that every formula F(x) determines a set w whose elements are exactly those objects satisfying the formula, i.e., x‰w if and only if F(x), for all x. The set w is often symbolized as {x : F(x)}, "the set of all x such that F(x)." In that case, the comprehension principle can be symbolically rendered as x‰{x : F(x)} $\Leftrightarrow$ F(x), where the double arrow symbolizes 'if and only if'. Both the extensionality and comprehension principles were (and are) extensively used in mathematics and regarded as intuitively natural.

Based only on these principles, Frege constructs a system (**N**, 0, ß) satisfying the Peano axioms. Frege first defines the empty set Ö as the set determined by any contradictory formula

(e.g., x‰x & x»x). Then, 0 is defined as the set {Ö} whose only element is Ö (0 is thus the set of all empty sets), 1 is the set of all singletons {x}, 2 the set of all doubletons {x,y}, etc. The successor ß(x) of a natural number (in fact of any set) is defined by comprehension as the set of all sets y such that y*‰ x, where y* is the result of removing any single element from y. An inductive set is then defined as a set having 0 as an element and also the successor of any element it contains. **N** is now definable as the intersection of all inductive sets, and the Peano axioms are provable, thereby establishing that **N** is an infinite set.

Frege's method and approach were ingenious but, as Bertrand Russell discovered in 1902, Frege's system is logically contradictory. To deduce formally a contradiction in this system, we need take only the formula x»x and apply the comprehension scheme to obtain a set w whose elements are exactly those which satisfy this formula. Thus, x‰w if and only if x»x. But since this holds for all sets x, it holds in particular for w, giving the contradiction w‰w if and only if w»w. This contradiction is known as 'Russell's paradox' and it shows that {x : x»x}, "the set of all sets x not elements of themselves," is a contradictory notion.

Russell's paradox sent a shock wave through the mathematical and philosophical community of the day. In particular, the failure of Frege's system showed that there are indeed limits to mathematical/logical generality and that the intuitive naturalness of a principle like comprehension is not a sufficient guarantee against logical error. The basic problem of foundations thus became that of finding a coherent method of distinguishing between legitimate and illegitimate uses of the comprehension principle. To that end, two new systems were propounded in 1908, one by E. Zermelo and the other by Russell himself.

Type theory and set theory.

Russell's theory of types considers that all sets are built up from individuals, which are defined as abstract, simple entities (i.e., entities devoid of any complexity). Individuals are declared to be of type 0, sets of individuals of type 1, sets of sets of individuals of type 2, and, generally, entities of type n+1 are sets whose elements are all of type n. The language of type theory is more restrictive than the language of Frege's system, because now basic formulas have the form x^n‰y^{n+1}, where x^n and y^{n+1} are variables of types n and n+1 respectively. Other formulas are built up from basic formulas by the usual propositional connectives and the quantifiers applied to typed variables. We say that the formulas of type theory are stratified.

The logical rules of type theory are appropriate generalizations of the rules for first-order logic. (In fact, type theory can be given a first-order formulation.) The specific axioms of type theory are just the two principles of extensionality and comprehension, formulated within the language of type theory. Russell's paradox is avoided because the troublesome formula x»x is not stratified and thus not a part of the language. Frege's construction of the system (**N**, 0, ß) can be carried through in type theory because any set x and its successor ß(x) are of the same type. Moreover, the first and third Peano axioms can be proved. However, because of the type restrictions on the formulas of the language, the theorem of infinity can no longer be proved and must now be added as an explicit axiom. In fact, the form Russell chose for the axiom of infinity was precisely the second Peano axiom, which asserts that the successor function ß is injective, and thus implies (in conjunction with the first Peano axiom) that **N** is infinite.

Russell's system of type theory (which we have presented here in a somewhat simplified form) was clearly an attempt to salvage as much as possible of Frege's

approach to foundations. However, in its final form type theory confronts us with a dilemma. The system is clearly (and provably) consistent without the axiom of infinity, but in that form it does not provide an adequate basis for infinite mathematics. A flexible and natural foundation for mathematics does result when we add an axiom of infinity, but now the infinity postulate acquires a somewhat*ad hoc* character losing thereby some of its logical justification. In particular, we have not derived the natural numbers from any more fundamental intuition, because the postulation of infinity is already equivalent to the existence of the natural numbers according to Dedekind's core result.

A liberalized version of type theory, called New Foundations, was devised by the American logician W. V. Quine in 1937. It requires that the formulas F(x) of the comprehension axiom be stratified but otherwise allows general (non-stratified) formulas in the language. It was shown in the 1950's by E. Specker and N. Goodman that the principle of infinity is provable in Quine's system. In itself this is a positive result, but a number of anomalies and bizarre features of New Foundations continue to appear, making it perhaps the most controversial of all foundational systems. The current consensus seems to be that the system is most probably free from formal contradiction but too unnatural to be considered a satisfactory foundation for mathematics. It allows for the generation of analysis and for most of the central principles of classical mathematics but also generates a number of principles and properties that do not occur in the usual practice of mathematics.

E. Zermelo's axiomatic theory of sets also appeared in 1908. It was initially presented as a straightforward codification of the most useful and generally accepted instances of the comprehension principle (essentially, G. Cantor's intuitively-defined set operations). The original system had some defects and limitations that were eventually corrected and removed by A. Fraenkel, T. Skolem, J. von Neumann, A. Mostowski, A. Morse, and J. Kelley. However, in the system's most definitive version, the dominant ideas are those of von Neumann. We will present the system in its most complete form as a *class/set* theory.

In class/set theory, mathematical entities are considered to constitute a hierarchy, but a much more flexible one than in Russell's type theory. At the bottom of the hierarchy are simple objects, called atoms, which have no elements whatever, but which can themselves be elements of composite entities, called classes.

The language of class/set theory is almost the same as Frege's: basic formulas have the form x‰y or At(x) ('x is an atom'), and general formulas are built up using the propositional connectives and quantifiers. Thus, the underlying logic of class/set theory is first-order logic. A class x, symbolized Cl(x), is any non-atom: Cl(x) $\Leftrightarrow$ $\neg$At(x), where '$\neg$' is the symbol for 'not'. Frege's principle of extensionality holds for classes but not for atoms. Thus, there can be many atoms, but exactly one no-element class, the null class Ö, classes being determined by their elements. Moreover, every atom occurs as an element of at least one class. Thus, formally, an atom can occur on the left of the '‰' sign but never on the right.

The classes are themselves divided into two further categories called sets and proper classes (nonsets). A set is a class (thus, not an atom) that occurs as an element of at least one other class, whereas a proper class is one that, while having elements, is never itself an element. Thus, formally, sets can occur on either side of the '‰' sign, while proper classes can occur only on the right of '‰'. If we define the predicate 'Sat(x)' to mean 'x is an atom or a set', then we have the principle: 'x‰y only if Sat(x)&Cl(y)', i.e., 'only atoms and sets are members and only classes have members'.

The distinction between atoms and classes is already contained in Zermelo's 1908 paper, but the distinction between sets and classes originates with von Neumann in 1925.

In class/set theory, the criterion that distinguishes proper classes from sets is size: any proper class is bigger than every set. Thus, proper classes are collections that are too large to be considered as a single entity and thus as a component (element) of another entity.

There is a suggestive analogy between the ontology of class/set theory and the ontology of modern physics, in which the atoms of class/set theory correspond to elementary particles, sets correspond to macro-objects, and proper classes correspond to macro-physical systems (e.g., galaxies) composed of many, possibly disparate, objects. It is not clear whether von Neumann (who made major contributions to theoretical physics) ever had such an analogy in mind.

Besides positing the extensionality principle, which holds for classes but not for atoms, class/set theory posits a comprehension scheme that holds for classes but not for sets: Given any formula F(x) of the language of class/set theory, there exists the class w of all sets or atoms x such that x‰w if and only if F(x). Letting w = {x : F(x)} as in the above, we thus have x‰{x : F(x)} ⇔ F(x)&Sat(x). An attempt to deduce Russell's paradox now only yields the result that {x : x»x} is a proper class (i.e., not a set).

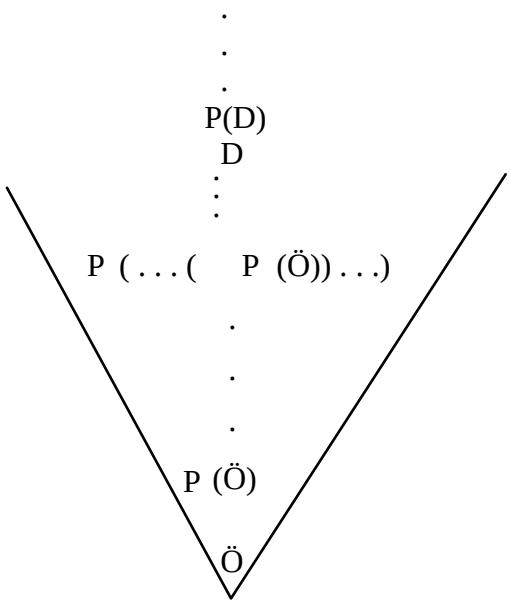
The remaining axioms of class/set theory affirm that the null class is a set, posit the existence of an infinite set, and establish several basic ways of generating new sets from existing sets. The axiom of pairing allows the formation of the doubleton set {x,y} whose elements are any two given sets x and y. The axiom of separation says that the intersection (common part) x”y of a set x and a class y is always a set. The operation of (infinite) union is defined with respect to a given function f: x ⇒ y, where x is a set. In this case, union affirms that the class {z : z‰f(k) for some k‰x} is a set. Furthermore, the image f(x) = {z : z = f(k) for some k‰x} is declared to be a set. Combining union with pairing immediately yields the fact that any finite class is a set. The powerset P(x) of any set x is the class of all subsets of x. Finally, there is a restriction axiom which asserts, in effect, that all sets are built up from atoms and the empty set Ö by (an unlimited number of) iterations of the powerset and union operations.

A theorem of Cantor establishes that the powerset P(x) of any set x has greater cardinality (more elements) than x. Thus, by iterating powerset, we can create sets of greater and greater cardinality. In particular, starting with the infinite set whose existence is postulated by the axiom of infinity, we can create sets of increasingly greater infinite cardinality. This yields Cantor's hierarchy of transfinite numbers.

The union operation allows us to create many different sets by forming all possible collections of the sets that exist. As we iterate powerset and union, we therefore progressively create bigger sets and more sets. The universe (proper class) V = {x : x = x} of all sets and atoms therefore grows both upward and outward as we iterate powerset and union. This is often illustrated "geometrically" as follows.

The universe V of all sets of class/set theory.

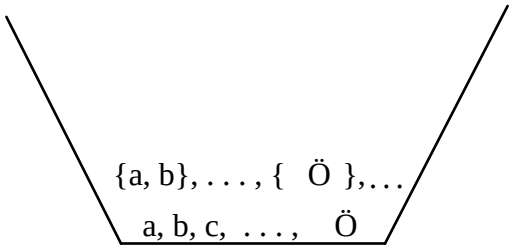
The shape of the universe of sets depends in part on how many atoms are assumed to exist. In the case where no atoms at all are postulated, we have a theory of pure sets in which the only no-element entity is the null set.



In the universe of pure sets, all sets are built up from the null set by alternated iterations of powerset and union. Thus, after taking the union D of all the sets Ö, P(Ö), . . . , P(. . . (P(Ö)) . . .), . . ., we will continue with iterations of the powerset applied to D, thereby generating the hierarchy D, P(D), . . . , P(. . . (P(D)) . . .), . . ., which will, again, be united into a set D', etc., *ad infinitum.* The universe V of sets is the proper class of all sets generated in this manner. Powerset generates increasingly bigger sets (augmenting the "height" of the universe) whereas union generates more and more sets (augmenting thereby the "width" of V).

The universe V of sets obtained in this manner is said to be *well-founded*, meaning that V has a bottom (a beginning point) but no top (no highest or final point). The well-foundedness feature of V is due to von Neumann's axiom of restriction, and establishes a strong analogy between the universe of sets and the physical universe, which, according to modern quantum theory, has irreducible, smallest elements (elementary particles) but no apparent limit to the ultimate complexity of systems that can be formed by combinations of elementary particles.

If we assume the existence of atoms, then we "widen the base" of the universe without changing its fundamental structure. This gives the following shape to V:



The atoms a,b, c, . . . and the null set Ö represent the lowest level (the "bottom") of the universe. The next level will be sets having as elements only those objects of the lowest level, and so on with alternated iterations of powerset and union as before.

The universe of pure sets is sufficient as a foundation for pure mathematics, but it is sometimes quite convenient to have atoms in the universe.

Without the axiom of restriction, the universe V could be thought of as both infinitely descending and infinitely ascending. Such *anti-founded* universes have been studied by the logicians P. Aczel, J. Barwise, and J. Etchemendy.

The lavish existence assumptions of class/set theory certainly provide a rich theory in which to carry on mathematics. For example, since the existence of an infinite set is explicitly postulated, one can directly implement Dedekind's original construction of the natural numbers and prove the Peano axioms. Or, following von Neumann, one can use the operations of set theory to construct a privleged model (**N**, 0, ß) of the Peano axioms. 0 is the null set Ö and the successor ß(x) of any set x is the *self-adjunction* of x, ß(x) = x ’{x} (the union of x with {x}), which amounts to adjoining the entity (set) x to the collection x. Thus defined, ß(x) is always different from x since, as a consequence of the axiom of restriction, x‰x never holds in class/set theory.

Except for the axioms of extensionality and restriction, each of the other axioms of class/set theory is a particular case of Frege's comprehension scheme. However, for the full development of mathematics, a further axiom is needed, the *axiom of choice*, which states that an infinite choice is always possible, even when no formula or rule for making the choice is given. In 1953, E. Specker showed that the negation of the axiom of choice is deducible in Quine's system of New Foundations (see above). Since the axioms of New Foundations are only extensionality and comprehension for stratified formulas, Specker's result shows that Frege's system would have been inadequate as a foundation for all of mathematics even if it had been consistent.

Note that there is an obvious model of type theory within class/set theory. The individuals (entities of type 0) of type theory are interpreted as the von Neumann natural numbers **N.** Entities of type 1 are then elements of P(**N**), entities of type 2 elements of P(P(**N**)), and so on. Thus, the axioms of type theory are all true when appropriately interpreted in the hierarchy **N**, P(**N**), . . . , P(. . . (P(**N**)) . . .), . . . , of class/set theory.

Foundational aspects of class/set theory.

Ever since Cantor's development of the theory of infinite sets and transfinite numbers, set theory has been regarded as a legitimate mathematical theory in its own right. Evaluating class/set theory as a foundation for mathematics is a more difficult matter. On one hand, the *ad hoc* postulation of infinity makes it difficult, if not

impossible, to consider class/set theory as a logical justification for infinite mathematics. Indeed, as a foundation, the whole of class/set theory has a certain *ad hoc* character, since it is essentially a straightforward codification of those principles that are perceived by most mathematicians as necessary for mathematics.

On the other hand, class/set theory does not appear to generate any principles that are shocking or unacceptable to mathematicians; whereas other systems in which a theorem of infinity is provable (e.g., Quine's system of New Foundations discussed above) do generate various unacceptable results along with the acceptable ones. The experience of several generations of mathematicians with set theory has restored a certain confidence in infinite mathematics, and the principles of class/set theory have come to be viewed generally (but not universally) as intuitively natural. However, various other foundational limitations of class/set theory have appeared.

In 1931, K. Gödel employed an ingenious argument which, as slightly improved by J. B. Rosser, establishes that any consistent, recursively axiomatic system S adequate for Peano arithmetic (and thus, in particular, any foundational system S) must contain an infinity of *undecidable* propositions p, i.e., propositions in the language L of S such that neither p nor its negation not-p is a theorem of S. (A *recursively axiomatic* system is one whose set of axioms can be determined by an algorithm (see above). All known foundational systems have this property.) Since either p or not-p must be true under any interpretation of a system, Gödel's result means that if class/set theory is consistent, then there are infinitely many true statements in the language of the theory that cannot be proved from its axioms (and that this will continue to hold even if any finite number of new axioms be added to the system).

This result sent a second wave of shock through the mathematical community, just as the recovery from the initial shock of Russell's paradox seemed complete. However, the undecidable propositions actually exhibited by Gödel's construction appeared rather contrived and artificial, thus minimizing the initial concern about "lost truths" of mathematics.

Nevertheless, from the beginning of Cantorian set theory in the late l9th century, there were several fundamental propositions that had resisted all efforts either of proof or of disproof. Perhaps the most significant was Cantor's continuum hypothesis. According to Cantor's theorem (see above) the cardinality of the natural numbers **N** is strictly less than the cardinality of its powerset P(**N**). It was also established that the cardinality of the real numbers **R** (sometimes called the continuum) was the same as P(**N**). The question then arose as to whether or not there are sets whose cardinality is strictly intermediate between the cardinalities of **N** and **R**. Cantor himself hypothesized that there were none and called this proposition the continuum hypothesis, abbreviated CH. In 1963, P. Cohen proved that CH was an undecidable proposition of class/set theory (provided that the latter is consistent). Cohen also proved that the axiom of choice was independent of the other axioms of class/set theory.

The independence of the choice axiom can be viewed as nothing more than a proof of the efficiency of the axiomatization of set theory, because there is general agreement that choice is an intuitively valid principle of mathematics. But there is no comparable consensus regarding CH (other than the observation that the assumption of CH simplifies considerably the theory of cardinal numbers). Subsequent to Cohen's result, and using his method of *forcing,* a number of other undecidable propositions of class/set theory have been discovered. This plethora of independence results shows conclusively that the truths of mathematics are underdetermined by the axioms of class/set theory, provided these latter are consistent.

The necessity for the *ad hoc* postulation of infinity in both type theory and set theory, and the consequent failure to derive infinite mathematics from any more fundamental logical intuition, raised the question of the consistency of these theories in a particularly sharp way. That a contradiction has not in fact been derived in these systems is not in itself a guarantee of consistency. For example, it would have been logically possible for the contradiction in Frege's system to have remained undiscovered for many years.

In the 1920's, the German mathematician D. Hilbert conceived an ingenious approach to this problem. Hilbert proposed a "program" consisting of a detailed mathematical analysis of the logical structure of the formal system S of class/set theory with the intention of establishing a rigorous mathematical proof that S cannot produce a contradiction. Initially such a program appears circular, because it proposes to use mathematics to prove the consistency of mathematics. Hilbert met this objection by proposing to use only finitary mathematics to prove the consistency of infinitary mathematics, pointing out that even though the system of class/set theory contains principles of infinitary mathematics, the formal system itself is a concrete, finitary mathematical object whose language, propositions, axioms and rules are explicitly and constructively defined.

However, Hilbert's program foundered on a second fundamental result of K. Gödel, which establishes that if a foundational system S is consistent, then its consistency can be proved only in a stronger system. Indeed, Gödel's second result shows explicitly how to deduce a formal contradiction "p and not-p" within any foundational system S, once given a proof within S of the consistency of S. Since infinitary mathematics contains finitary mathematics as a subsystem, one cannot therefore use the latter to prove the consistency of the former, if the former is consistent (recall that anything is provable in a contradictory system).

Our model of type theory within class/set theory (see above), is a proof of the consistency of type theory within class/set theory. Thus, according to Gödel's second result, class/set theory is strictly stronger than type theory, provided type theory is consistent. It likewise follows from Gödel's result that we cannot prove the consistency of class/set theory within type theory, provided, again, that type theory is consistent.

Thus, the fact that class/set theory has so far produced no intuitively unacceptable or contradictory propositions is, in the final analysis, the only guarantee of its integrity and coherence.

Constructivism and predicativity.

From the beginning of the modern period of foundational study, a certain number of mathematicians have articulated a *constructivist* conception of mathematics. Though partly based on a negative and skeptical view of infinite mathematics, constructivism is seen by its proponents as a positive and vigorous philosophy of mathematics. Nevertheless, all forms of constructivism propose some kind of restriction on infinite mathematics.

L. Kronecker, H. Poincaré, L. Brouwer, H. Weyl, and more recently E. Bishop, are among those who have propounded constructivist conceptions of mathematics. However, the leading champion of constructivism was undoubtedly Brouwer, to whom we owe the most comprehensive and thorough exposition of constructivist ideas and principles.

Brouwer held that all attempts to derive mathematics, and the theory of the natural numbers in particular, from some more fundamental or more general logical principle or intuition were mistaken. For him, the intuition of the succession of positive integers, æ, ææ, æææ, . . . , conceived as the repetition of an abstract unit æ, was the ultimate and primary intuition upon which to found mathematics. Although the sequence of natural numbers is *potentially infinite* because unending, Brouwer held it illegitimate to consider these numbers as constituting a completed whole or set **N**, to which further operations, such as the formation of a powerset P(**N**), could be applied. Indeed, Brouwer rejected all of Cantor's infinitary set operations. Thus, Brouwer would accept that the Peano axioms (see above) are clearly true of the natural numbers, but would reject Dedekind's proof of this fact, based as it is on certain principles of infintary mathematics.

Brouwer also rejected some general principles of first-order logic itself, more particularly the classical law of *excluded middle* which asserts that any proposition of the form 'p or not-p' is universally valid (true under any interpretation of p). For Brouwer, to assert the truth of a proposition of the form 'p or q' is to give an explicit proof either of p or of q. More generally, Brouwer identified the notion of mathematical truth with the notion of constructive provability. Philosophically, Brouwer's constructivism is thoroughly intuitionistic (see above) and represents a total rejection of mathematical Platonism.

An oversimplified but nonetheless useful approximation of Brouwer's vision of mathematical reality can be obtained from class/set theory by deleting the axioms of infinity and choice and removing the principle of excluded middle from the underlying logic. In such a system, there will still be an infinity of sets, e.g., the sets Ö, P(Ö), . . . , P(. . . (P(Ö)) . . .), . . . , but no infinite set. Thus, the individual natural numbers exist but not the set **N** whose elements consist of the natural numbers (and nothing else). The other number systems are treated in a similar manner, but considerable portions of classical mathematics are sacrificed in the process. For example, one deals only with the field of *constructive real numbers* — those that can be explicitly approximated by a constructive sequence of rational numbers — and not the classical complete ordered field of Dedekind.

However, because of constructivism's rejection of the Platonic conception of a mathematical universe of stable, nonphysical and ideal objects, it is more accurate to regard constructive mathematics as a "reality in process of being determined" rather than as a predetermined reality. For example, some laws of Platonic mathematics may fail practically in constructive manipulations of extremely large numbers, even with the use of a high-powered, modern electronic computing system. Consider, for instance, the equation (E) (a+b)-c = (a-c)+b, which holds between any (Platonic) integers a,b, and c. When a and b are extremely large, the left hand side of this equation may be practically undefined because the addition algorithm applied to a and b will generate a number too large to be represented in the system, thereby causing it to "overflow." But the calculation may become manageable once a is diminished by c, thus allowing the right hand side to be computed. Hence, from a purely constructive point of view, the set of values of a,b, and c for which the equation (E) holds is not completely determined but evolves and changes as we build more powerful computers or devise more efficient algorithms.

Though Brouwer and others have been quite vigorous in defense of the constructivist-intuitionist vision of mathematics, their school of thought has gained relatively few adherents. For the majority of mathematicians, constructivism is not

infrequently perceived as an attempt to impose an unreasonably restrictive philosophy on mathematical practice rather than to resolve genuine foundational issues. The question can be put quite simply and squarely: if infinite mathematics (including the logical law of excluded middle) is in fact without contradiction, then why should we arbitrarily restrict ourselves to a weaker, emasculated form of mathematics just to satisfy certain essentially philosophical dictates?

However, independent of philosophical issues, it is now generally recognized that constructivism has made positive and genuine contributions to mathematics. For example, a non-constructive proof of the existence of a certain limit may give us no idea of what the value of the limit actually is; whereas a constructive proof of the same result may furnish an explicit means of calculating an approximate value of the limit. Thus, even though constructive proofs are often more complicated than non-constructive ones, the extra effort involved in finding a constructive proof is frequently rewarded by extra information about the mathematical object in question.

Because of this extra information usually contained in constructive results, they are regarded as the most genuinely useful part of mathematics by some mathematicans who do not otherwise adhere to a constructivst philosophy of mathematics. This raises the question as to whether all constructive results can be obtained by constructive methods, or whether nonconstructive mathematics is an unavoidable necessity for certain constructive results. Work on this question has provided examples of both extremes: results first obtained by nonconstructive methods but later obtained constructively, and seemingly constructive results for which no known constructive proof has yet been discovered. This suggests that the constructive and nonconstructive aspects of mathematics are delicately intertwined and perhaps cannot be strictly separated in any way that isolates and preserves just the constructive part as an undivided whole.

Constructive proof vs. nonconstructive proof.

It is easy to give examples of rational powers of rational numbers that are themselves irrational. For example $2^{1/2} = \sqrt{2}$, which is irrational. The more difficult question arises: are there irrational powers of irrational numbers that are rational? A positive answer to this question can be given nonconstructively as follows: Consider the number $n = (\sqrt{2})^{\sqrt{2}}$, which is the irrational number $\sqrt{2}$ raised to itself and thus to an irrational power. The number n is either rational or not (principle of excluded middle). If, on the one hand, n is rational, then it is a positive example of an irrational power of an irrational number that is rational. If, on the other hand, n is irrational, then $n^{\sqrt{2}}$ is, again, an irrational power of an irrational number, but $n^{\sqrt{2}} = ((\sqrt{2})^{\sqrt{2}})^{\sqrt{2}} = (\sqrt{2})^2 = 2$, which is rational. Thus, in either case, we have an example of an irrational power of an irrational number that is itself rational.

The nonconstructive nature of this proof is reflected in the use of the principle of excluded middle, and the nonconstructive nature of the result is that we have proved the statement 'either n or $n^{\sqrt{2}}$ is an irrational power of an irrational number that is rational' without determining which of these two alternatives is in fact the case. One way of giving a constructive proof of this result would be to establish, by a direct argument, either that n is irrational or that n is rational. Such a proof would provide a positive answer to our initial question but also contain the extra information about the rationality or irrationality of n.

There are some mathematicians who find Brouwer's constructivism too radical but who also have difficulty swallowing all of the infinitary principles of class/set theory. As a result, a number of intermediate or "soft constructivist" proposals have been advanced over the years. For example, some propose abandoning just the axiom of choice or else replacing the full choice axiom with various weaker versions (e.g., principles of *denumerable choice* or *relative choice*).

Another proposal, first put forward by H. Poincaré, would accept the principle of excluded middle, but ban *impredicative definitions* in which an object m is a member of a set M but is defined only with reference to M. Predicative mathematics would accept the axiom of infinity — and thus the existence of such infinite objects as the completed set **N** of natural numbers — but would restrict the use made of these objects by disallowing the application of impredicative definitions to them.

It appears from his writings that Poincaré himself may have supposed that, when once granted the existence of **N**, the rest of infinite mathematics could be constructed in a strictly predicative manner. However, this turns out not to be the case. The formation of the powerset P(X) of a given set X is predicative, but Cantor's proof that P(X) has higher cardinality than X is impredicative. Thus, the sequence of sets **N**, P(**N**), . . . , P(. . . (P(**N**)) . . .), . . . , exists predicatively, but the proof that these sets constitute infinities of progressively higher order is impredicative. Moreover, arbitrary unions are impredicative, and Dedekind's construction of the real numbers **R** from the rationals **Q** makes unavoidable use of impredicative infinite unions. Thus, not even the real numbers **R** are predicative over the natural numbers **N**. Hence, in the last analysis,

predicativity appears also as an unnatural, philosophical, non-mathematical limitation on mathematical practice.

Combinatory logic and category theory.

Beginning in the 1920's, and continuing for the next decade, the Russian logician M. Schönfinkel and the American logicians H. B. Curry, A. Church, S. C. Kleene and J. B. Rosser developed an approach to foundations based on 'function' and 'application of a function to its argument' rather than 'set' and 'a set is an element of another set'. The functions of *combinatory logic* are conceived as operators f that are universally defined and thus applicable to everything, including themselves. Thus, in combinatory logic, f(f) (the value of the operator f for the argument f) is meaningful; whereas in class/set theory, self-application of functions is excluded by von Neumann's axiom of restriction.

However, in 1935, Kleene and Rosser derived a contradiction in combinatory logic similar to Russell's paradox in Frege's system. A revised and weaker version of the system was proved consistent by Church and Rosser, subsequent to which work on combinatory logic was carried on almost exclusively by Curry and his students. However, the Curry school never succeeded in reconstructing a consistent system of combinatory logic sufficiently strong to serve as a foundation for infinite mathematics.

In the early 1960's, there emerged another foundational system based on a generalized notion of function called a 'morphism' or 'map'. *Category theory* was different from and more flexible than combinatory logic in several ways. To begin with, the basic relationship between morphisms is composition 'ö' rather than application. Moreover, the morphisms of category theory are only locally and not universally composable. Indeed, each morphism h is accompanied by an explicit domain A and an explicit codomain B, composition between morphisms being defined only when domains and codomains correspond appropriately. Thus, where $h:A \Rightarrow B$ symbolizes a morphism with domain A and codomain B and $g:C \Rightarrow D$ a morphism with domain C and codomain D, then the composite göh will be defined precisely when B = C, i.e., when the codomain of h is the domain of g.

The analogy between morphisms and functions is obvious. Indeed, if we think of domains and codomains as 'sets endowed with a similar structure' then morphisms can be thought of as 'functions that preserve certain structural features from domain to codomain'. A (concrete) *category* is then a collection of sets endowed with similar structure, together with a collection, closed under composition, of structure-preserving functions between these sets. (An *abstract category* has arbitrary objects for domains and codomains and arbitrary relations as morphisms, provided these data satisfy a few fundamental axioms such as compositional closure and the associativity of composition.)

In the initial period of its development, starting with the work of S. Eilenberg and S. MacLane in 1945, the motivation for category theory was more geometric and algebraic than analytic. It was primarily in the 1963 doctoral dissertation of F. W. Lawvere, a student of Eilenberg, that the foundational potential of category theory became apparent. The key notion is that of universal structure (see above), already implicit in Dedekind's characterization of the natural numbers. Lawvere noticed that the system (**N**, 0, ß) of natural numbers could be completely described by the structure-preserving functions $h:\mathbf{N} \Rightarrow S$, where S is endowed with a structure (S, a, f) similar to the structure (**N**, 0, ß), and that this description involves only the notion of the composition of functions (see box above). Lawvere then set out to identify and

characterize universal mathematics in a similar way that Brouwer had sought to characterize constructive mathematics.

Ingeniously applying the general notion of universal structure formulated and developed by the French school of algebraic geometers, Lawvere succeeded in showing that the heart of classical mathematics, including analysis, was indeed composed of universal systems. For example, Cantor's powerset operation (see above) was reformulated by Lawvere in purely universal terms, as were many other basic notions of set theory. However, in its most general form, Cantor's union operation does not seem to be universal.

Thus, *topos theory* (which was the name Lawvere gave to the ultimate form of his foundational system) includes the Cantorian hierachy of infinite sets **N**, P(**N**), . . . , P(. . . (P(**N**)) . . .), . . . , and is much more comprehensive than Brouwer's constructive mathematics, but still does not include all of classical analysis. Interestingly (and surprisingly), the so-called internal logic of toposes turns out to be intuitionistic (the principle of excluded middle does not hold). Thus, looking at mathematics through the prism of universality, intuitionistic logic appears natural rather than arbitrary or forced. This has led some constructivists to embrace topos theory as the golden mean between the perceived excesses of class/set theory on one hand and the ravages of Brouwerian constructivism on the other.

In the last analysis, the foundational power of topos theory turns out to be roughly equivalent to Russell's type theory with an axiom of infinity (which, through the prism of universality, appears more natural and justified) but without the axiom of choice or the principle of excluded middle. However, G. Osius has shown how to add further axioms to topos theory in order to obtain a system equivalent to full class/set theory. An interesting result, due to R. Diaconescu, shows that the addition of the axiom of choice to topos theory immediately implies the principle of excluded middle (and thus that the internal logic is no longer intuitionistic).

In the 1970's, the American logician Dana Scott obtained a set-theoretic model of Curry's combinatory logic by interpreting the functions of combinatory logic as morphisms in a certain category (the category of complete lattices). Subsequent work, in particular by J. Lambek, has shown that combinatory logic is a special case of category theory in which, among other things, all of the morphisms have the same domain and codomain.

Thus, ultimately, the function-theoretic and set-theoretic versions of mathematical reality coincide, even though they represent rather different ways of looking at that reality.

The current situation: comparative and pluralistic foundations.

The experience of over a hundred years of modern foundational study has been exciting, frustrating, surprising, rewarding and rather less conclusive than most mathematicians would have liked. On the one hand, it now seems incontrovertible, in the light of Gödel's undecidability theorems and the plethora of independence results in class/set theory, that the initial goal of establishing a unitary, global foundation for the whole of mathematics is unrealistic. On the other hand, the fact that no contradictions, and indeed no unacceptable principles, have been forthcoming from class/set theory increases our confidence in the coherence and integrity of infinite mathematics.

Moreover, the availability in recent years of extraordinarily powerful electronic computing devices has allowed for an extensive exploration of computer-generated approximations of certain mathematical configurations previously inaccessible to any

practical verification. For example, even though the theory of *fractal geometry* (the geometry of spatial forms which, like some organs of the human body, combine global regularity with local irregularity) is simple and straightforward, it generates extremely complex configurations that can only be effectively represented by computer graphics. It is difficult to imagine that the mathematics of fractals would have developed to the same extent without the availability of computers.

The accumulation of various computer experiences has not yet revealed any significant or fundamental error in mathematical theory. Rather it has shown a remarkable harmony between mathematical theory and mathematical practice. This is, of course, an empirical and relative rather than logical and absolute verification of mathematical theory, but significant nonetheless.

A reverse influence of mathematical computation on theoretical mathematics is also emerging. For example, a recent innovation in and refinement of first-order logic, the *linear logic* of J.-Y. Girard, combines a logic of computation and a logic of proof within a single, unified system. Similarly, the negative solution, by the Russian logician Y. Matiyasevich, of Hilbert's tenth problem (concerning the existence of an algorithm for the solution of diophantine equations) has had an equally significant impact on both computational and theoretical mathematics. Such results show that even in the most established and fundamental parts of mathematics, much remains to be explored and discovered.

The failure to achieve a global and unitary foundation for all of mathematics has led some mathematicians to proclaim the "loss of certainty" or the "loss of truth" in mathematics. However, most mathematicians would regard these as exaggerated reactions and misguided interpretations of the current state of foundational study.

It is more balanced and more realistic to consider that mathematics exists as a body of truths about relationships between abstract entities and structures. These abstract relationships are reflected or instantiated, in various ways and at different levels, in the concrete structures of the physical world. We have no way of acceding directly to this body of truths and so we approach it from below by inductive generalization, based on analytical observation of empirical reality, and from above by creative conceptualization, based on our synthesized experience of reality as a whole. It is reasonable to assume that there may be any number of consistent and fruitful foundational systems that will generate a significant portion of this body of truths, but no system that will generate all of these truths and nothing else.

Foundational study can thus be viewed as an ongoing, flexible and pluralistic enterprise of elaborating foundational systems that are then carefully studied, compared and refined. Our experience has shown that this process invariably leads to new insights into the nature of mathematical truth and the structure of mathematical reality.

Bibliography. An excellent reference work that includes reprints of significant articles and materials from all of the major schools of foundational development from 1879 to 1931 is JEAN VAN HEIJENOORT (ed.), *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931* (1967). Probably the single most comprehensive, comparative study of nonconstructive foundations, including a treatment of category theory and Gödel's theorems, is WILLIAM S. HATCHER, *The Logical Foundations of Mathematics* (1982). For a more extensive treatment of topos theory than found in this latter reference, see R. GOLDBLATT, *Topoi, The Categorial Analysis of Logic* (second revised edition, 1984). For intuitionism and constructivism, see A.S. TROELSTRA, *Principles of Intuitionism* (1969). An excellent treatment of the major

logical systems related to foundational study is S. C. KLEENE, *Introduction to Metamathematics* (1952, reprinted 1971). Finally, RICHARD DEDEKIND, *Essays on the Theory of Numbers* (1901, reprinted 1963) is still quite accessible, except for some transparently outmoded terminology.