Extremal axioms

Jerzy Pogonowski

Extremal axioms
Logical, mathematical and cognitive aspects

Poznań 2019

REVIEWER
Prof. dr hab. Jan Woleński

First edition

COVER DESIGN
Robert Domurat

COVER PHOTO
Przemysław Filipowiak

ENGLISH SUPERVISION
Jonathan Weber

EDITORS
Jerzy Pogonowski, Michał Staniszewski

Publication supported by the National Science Center research grant 2015/17/B/HS1/02232

ISBN 978-83-64902-78-9
ISBN 978-83-7589-084-6

SOCIAL SCIENCE AND HUMANITIES PUBLISHERS
ADAM MICKIEWICZ UNIVERSITY IN POZNAŃ
60-568 Poznań, ul. Szamarzewskiego 89c
www.wnsh.amu.edu.pl, wnsh@amu.edu.pl, tel. (61) 829 22 54

WYDAWNICTWO FUNDACJI HUMANIORA
60-682 Poznań, ul. Biegańskiego 30A
www.funhum.home.amu.edu.pl, drozd@amu.edu.pl, tel. 519 340 555

PRINTED BY: Drukarnia Scriptor Gniezno

Contents

Preface

The term *extremal axiom* was introduced by Carnap and Bachmann in their article *Über Extremalaxiome* (Carnap and Bachmann 1936). Axioms of this sort ascribe either maximal or minimal property to models of a theory. Examples considered by Carnap and Bachmann included: the completeness axiom in Hilbert's system of geometry, the induction axiom in arithmetic and Fraenkel's axiom of restriction in set theory. The completeness axiom (replaced later by the continuity axiom) expresses the condition of maximality of the geometric universe: that universe cannot be expanded without violating the other axioms of the system. The axiom of induction is an axiom of minimality: it expresses the idea that standard natural numbers form a minimal set satisfying the other axioms of the system. Fraenkel's axiom of restriction says that the only existing sets are those whose existence can be proved from the axioms of set theory.

There are further axioms which can be considered extremal in the above sense. Gödel's axiom of constructibility and Suszko's axiom of canonicity are minimal axioms in set theory; as in the case of Fraenkel's axiom, they also express the idea that the universe of sets should be as narrow as possible. On the other hand, axioms of the existence of large cardinal numbers are maximal axioms; they express the idea that the universe of sets should be as rich as possible.

Extremal axioms are related to the notion of an intended model. The intended model of a mathematical theory is a structure that has usually been investigated for a long time and about which we have collected essential knowledge supported also by suitable mathematical intuitions. As examples of such structures one can take for example, natural, rational or real numbers, the universe of Euclidean geometry and perhaps also the universe of all sets considered in Cantorian set theory. Given such a structure one tries to build a theory of it, ultimately an axiomatic one. It may happen that one can prove that a theory in question characterizes the intended model in a unique way (up to isomorphism or elementary equivalence). On the other hand, such theories as group theory or general topology do not have one intended model; they are thought of as theories concerning a wide class of structures.

We feel obliged to explain the reasons for collecting the material that follows within a single monograph. There are numerous publications devoted to particular problems mentioned above. However, a synthetic approach to them seems to be rare. Besides the original paper by Carnap and Bachmann, there exist only relatively few works devoted to the extremal axioms in general. Those worth mentioning include: Awodey and Reck 2002a, 2002b, Hintikka 1986, 1991, Schiemer 2010a, 2010b, 2012, 2013, Schiemer and Reck 2013, Schiemer, Zach and Reck 2015, Tarski 1940.

In our opinion, the following topics are relevant with respect to the issue of extremal axioms:

1. Revolutionary changes in mathematics of the 19th century. This concerns above all:

    (a) The rise of modern algebra understood as the investigation of arbitrary structures (domains with operations and relations) rather than – as before – looking for solutions of algebraic equations.

    (b) Discovery of systems of geometry different from the Euclidean geometry known from Euclid's *Elements* and hitherto considered the *true* system of geometry. The proposal of thinking about geometries as determined by invariants of transformations.

2. The rise of mathematical logic. The codification of the languages of logic (type theory, first-order logic, second-order logic, etc.) made it possible to talk about mathematical structures in a precise way.

3. Attempts at axiomatic characterization of fundamental types of mathematical systems. The axiomatic method previously (before the 19th century) present only in Euclid's system of geometry has become widespread in other mathematical domains.

4. Attempts at *unique* characterization of chosen mathematical structures. In these cases, where mathematical research was focused on the properties of specific a priori intended models, the question

arose of a possibility of a unique (with respect to structural or semantic properties) characterization of such models.

5. The emergence of metalogic. About one hundred years ago investigations, and the existence of several systems of logic was admitted. As a consequence, questions naturally arose concerning the comparison of these systems, their general properties, and so on.

6. Limitative results in logic and the foundations of mathematics. Metalogical reflection very soon brought important results showing the possibilities and limitations of particular logical systems, above all concerning the famous incompleteness results. It became evident that certain methodological ideals can not be achieved simultaneously – for instance, "good" deductive properties and a great "expressive power" are in conflict, in a precisely defined sense.

7. Philosophical reflection on mathematical intuition as a major factor in the context of discovery.

Our main research goal is modest; we admit that the material covered by the book is to a large extent known to specialists. However, we believe that presenting the origins of extremal axioms and the development of research on them could be of some value to the reader who is interested in mathematical cognition. We make some use of the material contained in: Pogonowski 2011, 2016, 2017, 2018a, 2018b.

The book consists of three parts focusing in turn on logical, mathematical and cognitive aspects of extremal axioms.

Part I: Logical aspects. We begin here with a discussion concerning the relations between mathematical theories and their models. Then we present remarks about the origin of extremal axioms as well as the origin of certain metalogical properties, notably those of categoricity and completeness. We recall some famous limitative theorems which show possibilities and limitations in the unique characterization of models. The notion of the expressive power of logic is useful in this respect. The final chapter of this part gives examples of important results from model theory related to the properties of categoricity and completeness.

Part II: Mathematical aspects. Here we discuss the role of particular extremal axioms and selected properties of theories based on extremal axioms. First, we present results related to the continuity axiom in geometry, algebra and analysis. Then we recall the role played by the induction axiom in arithmetic. Finally, we discuss two types of extremal axioms in set theory, namely the axioms of restriction and the axioms of the existence of large cardinal numbers.

Part III: Cognitive aspects. The main topic of this part is mathematical intuition. This cognitive ability should be conceived of as a major factor responsible for characterization of intended models. We discuss selected philosophical standpoints concerning this notion as well as related opinions among professional mathematicians. To the well-known contexts of discovery and justification we add a new one, namely the context of transmission, which embraces activities related to learning and teaching mathematics, as well as the popularization of mathematics.

The book is addressed mainly to cognitive scientists interested in mathematical cognition. We are aware that logicians and mathematicians are well familiar with extremal axioms and their consequences in mathematics itself. Cognitive scientists, in turn, could be interested in the ways of characterization of models of mathematical theories or, more generally, in cognitive access to fragments of mathematical reality described by the theories in question.

The mathematical and logical terminology used in the book is standard (however, there are some differences in the notations used by the authors whose works are discussed). We do not explicate the elementary formal concepts, but we provide definitions of those advanced ones which are important for the main subject. It is assumed that the readers are acquainted with the fundamentals of formal logic and the rudiments of set theory (including the notions of ordinal and cardinal numbers) as well as elementary algebra (including such notions as isomorphism, group, and field) and a little of general topology. The presentation is in rather simple English due to the fact that the author is not a native speaker of the language.

The compilation of this monograph was supported by the National Scientific Center research grant 2015/17/B/HS1/02232 *Extremal axioms: logical, mathematical and cognitive aspects*. The research was conducted in the years 2015–2019 at the Department of Logic and Cognitive Science, Faculty of Psychology and Cognitive Science of the Adam Mickiewicz University in Poznań.

Part I

Logical aspects

Chapter 1

Mathematical theories and their models

If one wants to discuss the role of extremal axioms in investigations concerning the intended models of mathematical theories, then one should start with preliminary remarks about such theories and models and make clear what the primary goal was in the formulation of such axioms. This, in turn, implies that one has to discuss the methods of comparison of models and such typical results as theorems on representation and classification of mathematical objects. Finally, the investigation of intended models presupposes decisions as to which mathematical objects and structures are considered standard.

A few words about terminology are in order. The term ‘mathematical theory’ is used in several senses, notably the following ones:

1. A theory is a set of axioms formulated in a fixed formal language in terms of certain primitive notions. This set is often finite or recursively enumerable. This is the traditional mathematical notion of a ‘mathematical theory’.

2. A theory is a set of sentences closed under some (deductive or semantic) consequence operator. This is a contemporary standard logical notion of a ‘theory’.

3. A theory is the set of all sentences true in a particular mathematical structure, say M. In this case we talk about the ‘theory of M’.

There are still other senses in which one uses the term ‘mathematical theory’, but those listed above are the fundamental ones.

Mathematical theories have recently been formulated in formal languages equipped with some system of logic (deductive consequence) and semantics (model theory). When one talks about mathematical research conducted before the development of formal languages, formal logic and semantics, then one often proposes suitable reconstructions of the earlier reasonings, although this should be done with great care.

We assume that the reader is familiar with the notions of formal language, a deductive consequence relation in such a language and such semantic notions as model, satisfaction and semantic consequence. If $\vdash$ is a deductive consequence in a language L, φ is a sentence of L and Φ is a set of sentences of L, then $\Phi \vdash \varphi$ reads 'φ is deducible from Φ'. If a structure M satisfies a sentence φ, then we write $M \models \varphi$. If $M \models \Phi$, then we say that M is a model of Φ. We say that Φ semantically implies φ (in symbols: $\Phi \models \varphi$), if all models of Φ satisfy φ. Two important properties related to a deductive consequence relation $\vdash$ and a semantic consequence relation $\models$ (for a given language L) are the following:

1. The relation $\vdash$ is *sound* with respect to the relation $\models$, if for all sentences φ and sets of sentences Φ: if $\Phi \vdash \varphi$, then $\Phi \models \varphi$.

2. The relation $\vdash$ is *complete* with respect to the relation $\models$, if for all sentences φ and sets of sentences Φ: if $\Phi \models \varphi$, then $\Phi \vdash \varphi$.

These properties concern the system of logic under consideration. As the reader surely remembers, first-order logic is sound and complete. Second order-logic with the standard set theoretic semantics is not complete.

1.1 Theories in polymathematics and monomathematics

The development of mathematics is a continuous process, though there are moments (or, better, periods) in its history which are – in a sense – revolutionary in character. There are many factors causing such events, including, accumulation of knowledge, surprising results, the emergence of antinomies and paradoxes, change in research perspective, and even deliberatively proposed programs. Well-known and extensively discussed critical moments of this kind are, among others (we do not include most recent events):

1. The discovery of incommensurable magnitudes by the Pythagorean school.
2. The invention of analytic geometry by Descartes.
3. The introduction of calculus by Newton and Leibniz.
4. The rise of abstract algebra in the 19th century.
5. The discovery of non-Euclidean geometries in the 19th century.
6. Arithmetization of analysis in the 19th century.
7. The emergence of set theory in the 19th century.
8. The discovery of incompleteness phenomena in the 20th century.

Each of the above deserves more attention. However, our main task is connected with extremal axioms and not with the history of mathematics in general. Therefore we shall limit ourselves to the events directly connected with the said axioms. In this respect one should notice the following:

1. The formulation of extremal axioms was possible only after the development of the axiomatic method (second half of the 19th century and first decades of the 20th century). Beforehand the axiomatic method has been used explicitly only in the system of geometry, going back to Euclid's *Elements*. In the period mentioned several axiom systems were proposed: for natural, rational and real numbers, for many systems of geometry, for algebraic structures, etc.
2. The borderline in the history of mathematics before and after 19th century is clearly visible. One can justly claim that the roots of modern mathematics stem from the 19th century in which the *structural revolution* took place. By this revolution we mean accepting a new perspective according to which mathematics – for instance algebra – is a study of several *structures* rather than a collection of methods of solving equations. The same concerns geometry, where investigations embraced projective, affine, multidimensional and non-Euclidean systems, among others.

Contemporary mathematics has several hundred branches. They were developed over the millennia, starting from the first reflections about magnitudes (later: numbers) and representations of space. The traditional classification of mathematical investigations into arithmetical and geometrical is, of course, obsolete. Currently there is an alphanumerical classification schema in use, called the *Mathematics Subject Classification (MSC)* which has multiply levels (the current version is MSC2010 and consists of three levels). The first level embraces 64 mathematical disciplines. They are sometimes grouped by common area names that are not part of the MSC in the following way:

1. General/foundations [Study of foundations of mathematics and logic].
2. Discrete mathematics/algebra [Study of structure of mathematical abstractions].
3. Analysis [Study of change and quantity].
4. Geometry and topology [Study of space].
5. Applied mathematics/other [Study of applications of mathematical abstractions].

There exist other divisions of mathematics, used for classification purposes – such as the Library of Congress Classification or the Dewey Decimal Classification. Mathematics taught at school has its own division, for obvious reasons. Mathematics at the university level may be presented as the study of several types of structures: arithmetical, algebraic, topological, differential, measure-theoretic, and so on. Here we are interested in one special division of mathematical investigations which could be called, after a proposal presented in Tennant 2000, *monomathematics* versus *polymathematics*:

1. *Monomathematics.* This includes theories whose main goal is to characterize a chosen mathematical structure.
2. *Polymathematics.* This includes theories which characterize whole classes of mathematical structures.

It is understood that this classification applies to theories based on a system of axioms chosen appropriately to serve the aims of the theory in question. Examples of theories belonging to monomathematics are: arithmetic (of natural, rational, real and complex numbers), Euclidean geometry, and set theory (in its initial stage). In these cases theories were supposed to provide a unique characterization of the corresponding structures, i.e., respectively: domains of numbers of the mentioned sorts, spaces including points, straight lines and planes, and the universe of all sets. The methodological ideal was to propose axiom systems which described these systems categorically: in a unique way, up to isomorphism.

Examples of theories belonging to polymathematics are much more numerous. They include: theories of algebraic structures (groups, rings, fields, modules, vector spaces, Boolean or Heyting algebras, and lattices, etc.), theories of topological spaces (either general or more specific, e.g. metric, compact, or connected, etc.). In these cases the mentioned theories have many different interpretations.

One may say that polymathematics began in the 19th century with the change of perspective in algebra, the discovery of numerous systems of geometry, investigations into hypercomplex numbers, investigations into multidimensional spaces and manifolds. The distinction between polymathematics and monomathematics also gained sense in the second half of the 19th century, when axiom systems for number systems, algebraic structures and systems of geometry were proposed.

The first important results of monomathematics are the axiomatic categorical characterizations of natural numbers by Peano and Dedekind and the categorical characterization of Euclidean geometry in Hilbert's *Grundlagen der Geometrie*. As we will discuss later in greater detail, such categorical descriptions are possible in a rather strong metatheory – one which essentially makes use of second-order language. The impossibility of categorical descriptions (of sufficiently rich mathematical theories) in the first-order language was proven in the first half of the 20th century.

To sum up, extremal axioms are typical of the investigations conducted in monomathematics. Some fixed mathematical structure is given, we have at our disposal mathematical knowledge about it, and then we look for conditions which can characterize the structure in a unique way.

Such a characterization problem was not present in mathematics before the 19th century. Euclid's system of geometry was considered *the correct* geometry. Number systems were not characterized axiomatically and set theory has not even emerged. Observe that the *naive set theory* created by Cantor and developed by several authors before 1908 treated sets as forming a fixed universe of all *true* sets. Even the first axiomatic setting of set theory, proposed by Zermelo in 1908, did not suggest the possibility of talking about different interpretations of the concepts of set and the membership relation. Its modifications, by Skolem, Fraenkel and von Neumann, focused on the form of particular axioms mainly in order to eliminate "unwanted" sets (cf. the axiom of regularity) and in order to make certain operations on sets possible (cf. the axiom schema of replacement). The twenties and thirties of the 20th century saw the opening discussion on possible interpretations of set theory (von Neumann, Fraenkel, Mostowski, Zermelo), mainly in the context of independence of particular axioms. In his second axiomatization of set theory from 1930, Zermelo discusses the necessary and sufficient conditions for the existence of an isomorphism between his *normal domains*, which can be considered an attempt to characterize these interpretations categorically. The beginning of investigations into models of set theory is the famous work Gödel 1940, where the constructible universe is used for the proof of consistency of the continuum hypothesis (relative to the axioms of ZF set theory). In Cohen 1966 the independence of the continuum hypothesis (from the axioms of ZF set theory) was proven. The plethora of independence results obtained later shows that if the Zermelo-Fraenkel set theory is consistent, then it really has a huge class of drastically different interpretations and at the same time evokes motivation for the discovery of new axioms which could characterize the universe of all sets more uniquely than the standard Zermelo-Fraenkel theory. We will discuss this topic in some detail in chapter 7 of this book.

1.2 Types of models and their comparison

1.2.1 Intended models

We noted earlier that in monomathematics one deals with a chosen mathematical structure which is supposed to be characterized in a unique way. Such a structure may be justly called the *intended model* of the theory in question. Thus, the intended model of a theory is a structure given in advance, which is initially characterized by results concerning its properties obtained subsequently. Only after the accumulation of knowledge about the structure under investigation does its theory become more stable, and it finally becomes an axiomatic theory. Consider the following examples:

1. *Natural numbers.* Probably the first mathematical structure under consideration. Involved in the process of counting (as ordinal numbers) and fixing the numerical size of collections of objects (as

cardinal numbers). They form a structure of an ordered set with the first element and without the last element in which each element has an immediate successor and (except for the first element) an immediate predecessor. They are equipped with arithmetical operations of addition and multiplication (and restricted subtraction). One can also distinguish among them prime numbers (those divisible only by 1 and itself). Millennia of mathematical knowledge about them served as a basis for their axiomatic descriptions in the 19th century, proposed by Grassmann, Peano, Dedekind, Frege, among others.

2. *Arithmetic continuum.* Objects which are now called real numbers have been investigated since antiquity. Their presence in mathematics was always ubiquitous. They were represented in many ways, for example as non-terminating continued fractions or in decimal expansions. The development of algebra made it possible to treat irrational numbers as algebraic objects. The distinction between real and imaginary roots of polynomials was made by Descartes. The proof that some real numbers are transcendental was only obtained in the 19th century. The early works on Calculus used real numbers but without a solid logical background. The first precise definitions of real numbers were proposed by several authors, most notably by Hilbert, Cantor and Dedekind in the 19th century.

3. *Sets.* Georg Cantor originated set theory in the second half of the 19th century. Initially he was interested in specific sets of real numbers because of his investigations into convergence of trigonometric series. There had of course already been talk about collections of elements even much earlier in mathematics, but the *naive set theory* developed by Cantor and others was devoted to the general concept of set and to the transfinite hierarchy of infinities. Before the beginning of metalogical investigations in set theory, there would seem to have been a belief in the existence of one universe of all sets, being the intended model of this preliminary version of the theory.

In each of these cases we now know what the possibilities are of a unique characterization of intended models – which we describe in more detail in chapters 5–7 of this book.

1.2.2 How do we compare models?

A few words concerning the uniqueness of characterization of models are in order. In general, two mathematical structures may be compared with respect either to the way they are built (i.e. regarding the network of relationships between their elements) or with respect to their properties expressible in the language of the corresponding theory. Thus, in the first case we take into account the *internal* structure of models, while in the second we take into account the *semantic* properties of models. The two relations connected with these cases are, respectively:

1. *Isomorphism.* Two structures are isomorphic if there exists a one-one correspondence between their domains (i.e. a bijection) which preserves all the relations from the signatures of these structures. If, for example, $\mathbf{A} = (A, R, f)$ and $\mathbf{B} = (B, S, g)$, where A and B are sets, R and S are binary relations (on A and B, respectively) and f and g are functions $f : A \to A$, $g : B \to B$, then $\mathbf{A}$ and $\mathbf{B}$ are isomorphic if there exists a bijection $h : A \to B$ such that for all x and y in A:

   (a) $R(x, y)$ holds, if and only if $S(h(x), h(y))$ holds;

   (b) $h(f(x)) = g(h(x))$.

   In a similar way one defines this notion for arbitrary structures, with many relations and functions (including 0-argument functions, i.e. constant elements). If one considers not algebraic structures but for instance topological spaces, then the morphisms of appropriate type between them are *homeomorphisms*, for differential structures the appropriate morphisms are *diffeomorphisms*, and so on. The existence of isomorphism between two structures means that they are structurally indistinguishable, regardless the "quality" of the elements of their domains. This is of course an equivalence relation on the class of all models (of a given signature). The equivalence classes are thus *isomorphism types* of the models in question.

2. *Elementary equivalence.* Two structures being interpretations of the same theory T in a specified language L (say, first-order language) are elementarily equivalent if they satisfy exactly the same sentences from L. Thus $\mathbf{A}$ and $\mathbf{B}$ are elementarily equivalent, if for all sentences φ of L: $\mathbf{A} \models \varphi$ if and only if $\mathbf{B} \models \varphi$. Elementary equivalence is an equivalence on the class of all interpretations of

L. Elementarily equivalent structures are thus semantically indistinguishable, they have identical sets of sentences which are true in them.

If two models of a theory are isomorphic, then they are also elementarily equivalent. The converse implication is not true. Thus, a given class of elementary equivalence may contain several classes of isomorphism types of models.

If all models of a given theory T are isomorphic, then we say that T is a *categorical* theory. Thus, a categorical theory describes exactly one model (up to isomorphism).

If all models of a given theory T are elementarily equivalent, then we say that T is a (semantically) *complete* theory. This property can also be characterized syntactically: a theory T is (syntactically, or deductively) *complete*, if for any sentence ψ from the language of T, either $T \vdash \psi$ or $T \vdash \neg\psi$, i.e. if either ψ or $\neg\psi$ is a theorem of T.

Due to some metalogical results (the Löwenheim-Skolem theorem) categoricity is a rare phenomenon among first-order theories. One considers a weaker notion instead. We say that a theory T is *categorical in the infinite power* κ (in brief: κ-categorical), if there exists a model of T of power κ and all models of T of this power are isomorphic.

The extremal axioms considered in monomathematics are thus concerned with categoricity and completeness. In chapter 4 of this book we will discuss the results from general model theory connected with these two properties of theories. In the next chapter, in turn, we will say more about the origin of these notions.

Let us illustrate the notions introduced above with a few examples, not very sophisticated mathematically:

1. Let us consider the structures $(\mathbb{R}_+, \cdot)$ (i.e. the positive real numbers with multiplication) and $(\mathbb{R}, +)$ (real numbers with addition). As the reader surely knows from school, the natural logarithm function $\ln : \mathbb{R}_+ \to \mathbb{R}$ is a bijection and an isomorphism between these structures, because $\ln(x \cdot y) = \ln(x) + \ln(y)$.

2. The ordered structures $(\mathbb{Z}, <)$ (integers with natural ordering) and $(\mathbb{Q}, <)$ (rational numbers with natural ordering) are both countable but not isomorphic (and hence also not elementarily equivalent). The first of these orderings is discrete, while the second is dense.

3. The ordered structures $(\omega, <)$ and $(\omega + (\omega^* + \omega) \cdot \eta, <)$ are not isomorphic but they are elementarily equivalent in the first-order logic. We may think of them as structures consisting of ordinal numbers but in the usual mathematical discourse they may be represented as: the natural numbers $\mathbb{N}$ with the natural ordering $<$ and the ordered structure consisting of the initial segment ordered exactly as $\mathbb{N}$ followed by so many copies of the integers $\mathbb{Z}$ as there are rational numbers $\mathbb{Q}$ (i.e. we put $\mathbb{N}$ first and then we put copies of $\mathbb{Z}$, the set of all these copies ordered as the rational numbers $\mathbb{Q}$). This example shows that the first-order Peano arithmetic is unable to distinguish between the standard natural numbers and structures which – besides such numbers at the proper place – also include "alien intruders".

4. It is an old result of Cantor that any two countable dense orderings without endpoints are isomorphic. This shows one aspect of the uniqueness of the ordered set $\mathbb{Q}$ of the rational numbers.

5. First-order Peano arithmetic is a *wild* theory, i.e. in any infinite power κ it has the maximum possible number of models, that is 2^κ. In particular, it has continuum many countable models.

6. Examples of κ-categorical (for chosen values of κ) theories are, among others:

   (a) Theory of atomless Boolean algebras. It is $\aleph_0$-categorical.

   (b) Theory of algebraically closed fields of characteristic 0 (or p, where p is a prime number) is $\aleph_1$-categorical.

7. The following are examples of important complete and incomplete theories:

   (a) The set of all sentences true in the standard model of Peano arithmetic is a complete theory.

   (b) However, the set of all sentences provable in Peano arithmetic is not complete.

   (c) Theory of identity for infinite sets is complete.

   (d) Theory of zero and successor with the axiom schema of induction (thus without axioms characterizing addition and multiplication) is complete. However, it is not finitely axiomatizable.

(e) The Presburger arithmetic, that is the theory with zero, successor and addition (without multiplication) and with the axiom schema of induction is complete. However, it also is not finitely axiomatizable.

8. *Łoś-Vaught test.* If T is a consistent theory without finite models which is κ-categorical in some infinite power κ, then T is complete.

9. *Morley theorem.* If a first-order theory in a countable language is categorical in some uncountable cardinality, then it is categorical in all uncountable cardinalities.

1.2.3 Standard models

The terms: *intended* and *standard* model are often used interchangeably in literature. We dare to propose the following distinction.

By the *intended model* of a theory in the pre-axiomatic stage we may mean the mathematical structure which is described by the theory in question. Thus, natural numbers (with standard ordering, operations of addition and multiplication, the relation of divisibility without remainder, etc.) form the intended model of arithmetic in its pre-axiomatic stage. Similarly, the real numbers (with arithmetic operations, standard ordering, metric determined by absolute value, etc.) form the intended model of the pre-axiomatic theory concerning them. The intended model of the pre-axiomatic set theory is the universe of all sets with membership relation and other relations defined by it. The situation is a little bit different with the intended model of Euclidean geometry. We know that the origins of Greek geometry precede Euclid's system presented in the *Elements*. Early geometry (starting from about 3000 B.C.) collected observations and proposed certain principles related to measurement of lengths, angles, areas, and volumes. The Pythagorean theorem was known already to Babylonians about 1500 years before Pythagoras. The first deductions about geometric topics are attributed to Thales. But can we claim that the Greeks before Euclid had in mind an abstract geometrical structure (say, consisting of points, straight lines and planes)? Morris Kline writes:

> The Greeks wrote some histories of mathematics and science. Eudemus (4th cent. B.C.), a member of the Aristotle's school, wrote a history of arithmetic, a history of geometry, and a history of astronomy. Except for fragments quoted by later writers, these

> histories are lost. The history of geometry dealt with the period preceding Euclid's would be invaluable were it available. (Kline 1972, 26)

Up to the 19th century the structure described in the *Elements* was thought to be *the correct* geometry, reflecting the structure of the physical world. As shown by Pasch and Hilbert, the system of Euclid used some implicit assumption not expressed in the postulates but necessary for conducting several constructions. Yet the structure of points, straight lines and planes characterized by Euclid's postulates could be considered as the intended model of geometry with certain stability.

From the above remarks one can conclude that the notion of the intended model understood according to this proposal is a little bit vague and in a sense an intuitive notion. We want to point to the analogy with an intuitive notion of a *computable function* in this respect. There is no precise mathematical definition of what is a computable function, but there are several representations of this notion, for example recursive functions, Turing machines, Post systems, Church's λ-calculus, and Markov algorithms, and so on. That all these representations ultimately define the same class of functions is an important result. This is taken as a firm confirmation of the Church-Turing thesis saying that the intuitive notion of a computable function coincides with any of these representations. Of course the thesis itself is not a mathematical theorem.

Developing our analogy further, one may ask what is the formal mathematical counterpart of the intuitive notion of the intended model. We propose to call this counterpart the *standard* model and characterize the last notion as this model which is closest to the intended model with respect to its properties. A formal (axiomatized) theory may have many non-isomorphic models. Now, if it has exactly one model (up to isomorphism), then it is obviously reasonable to call it its standard model. If an axiomatic theory which replaced a pre-axiomatic one has many models, then the model which is isomorphic with the intended model (of the pre-axiomatic theory) could be justly called the standard model of the axiomatic theory.

Examples of standard models in this sense are:

1. The standard natural numbers (with arithmetic operations). It is the standard model of first-order Peano arithmetic. It is one of the continuum many countable models of this theory. In second-order logic, this model is unique, up to isomorphism.

2. The completely ordered field of real numbers. In second-order logic, this model is unique, up to isomorphism. It is the maximal Archimedean field.

3. The Cartesian model of the system of Euclidean geometry, as axiomatized for example in Hilbert's system from his *Grundlagen der Geometrie*. Cf. also the axiomatization in Borsuk, Szmielew 1975.

It is worthwhile recalling how the terms *intended* and *standard* model are used in literature in the cases of particular mathematical theories. Sometimes these terms are used interchangeably, as if they were denoting the same object. This is the case with the standard natural numbers as the intended and standard model of arithmetic. The pressure of tradition also influences terminology. For instance, in set theory one uses the following distinctions (cf. Jech 2003):

1. Let L be the language of set theory, i.e. a language with the predicate symbol $\in$ denoting the membership relation. Let S and T be theories in the language L. For instance, S may be the Zermelo-Fraenkel set theory ZFC with the axiom of choice and T another theory (possibly identical with S). Assume that M is a model for S, and N is a substructure of M such that:

   (a) $\in_N = \in_M \cap N^2$ (here $\in_N$ and $\in_M$ denote the denotations of $\in$ in N and M, respectively).

   (b) $N \models T$.

   (c) The domain of N is a transitive class in M (a class is transitive, if each member of it is a subset of it).

   (d) N contains all ordinal numbers of M.

   We then say that N is an *inner* model of T in M. By a *standard* model one often understands a model satisfying the first two of the above conditions.

2. If a theory T is identical with S (or only contains S), then it is reasonable to say that N is a model for S "inside" the model M.

3. If there exists a standard model for ZFC (this assumption is stronger than the mere assumption of the existence of a model), then there exists a smallest standard model called the *minimal* model, which is contained in all standard models.

4. Any model of ZF has a least inner model of ZF, which follows from Gödel's considerations concerning the constructible universe. In other words, every model M of ZF has an inner model L^M satisfying Gödel's axiom of constructibility, and it is the smallest inner model of M containing all the ordinals of M.

5. The class of all sets is an inner model containing all other inner models.

Roughly speaking, standard model of set theory is thus meant as one in which the membership relation is the "real" membership. We will come back to models of set theory in chapter 7 of this book.

1.2.4 Non-standard models

Assuming that we have an idea what the intended and standard models are, one may ask: what are the non-standard models? To which theories is this notion applicable? What mathematical properties have the non-standard models?

The distinction between standard and non-standard models is not the same as the distinction between objects which are considered as *standard* (*natural*, *normal*, etc.) objects and those which are *unintended* (*unwilling*, *imaginary*, etc.). It is the research practice of the given epoch that determines which mathematical objects bear the name *natural* or *standard* at that time. The introduction (discovery, invention) of new objects is sometimes called *innovation*. Haim Gaifman has discussed the following *innovations* in mathematics in his paper Gaifman 2004 devoted to the non-standard models:

1. The discovery of irrationals.

2. The incorporation of negative and complex numbers in the number system.

3. The extension of the concept *function* in the nineteenth century.

4. The discovery of non-Euclidean geometry.

Gaifman stresses that in the above cases we should not speak of non-standard models:

> Let me sum up the four historical cases and how they differ from non-standard models. First, the discovery of incommensurables is a discovery that a certain presupposition about spatial magnitudes was false. There is only one model, the standard one; we were simply mistaken about one of its basic features. Second, the enlargement of the positive number system by incorporating negative and complex numbers amounts to utilizing the possibilities inhering already in the positive numbers; there is no change of the standard model, but an unfolding of it. Third, the extension of the function concept to that of an arbitrary mapping (given as a set of pairs) is an explication of a previously loose concept, which is needed because new examples do not conform to previous expectations. There are no two models, but one developing conception. Fourth, the discovery of non-Euclidean geometry is the discovery that the concept of geometric space is ambiguous and admits an additional specification besides the received one; the difference is expressed as the denial of an accepted postulate. Here indeed there are several models and non-Euclidean geometry is, as Gauss noted, strange. If 'non-standard' is another word for 'strange' then it is "non-standard". But, as Gauss, Bolyai and Lobachevsky made clear, and as subsequent developments have borne out, non-Euclidean geometry is a legitimate conception of geometric space. If physical space is to be the arbiter of truth (as Gauss suggested) then neither the Euclidean nor the hyperbolic geometry is the winner. But whatever the verdict of physics, the different geometries constitute different specification of the general mathematical concept of geometric space. (Gaifman 2004, 13)

The period of domestication for new mathematical objects may be different in specific cases, for example a few hundred years in the case of negative and imaginary numbers versus a relatively short time in the case of hypercomplex or p-adic numbers.

Gaifman argues further in the paper quoted above that some structures can be justly named *standard*. This is obviously the case with natural numbers. To this one may add, according to Gaifman: *well orderings* and the class L of all *constructible sets*. On the other hand, the *full powerset operation* escapes the list of standard concepts. Even the full powerset of the set of all natural numbers seems to be inaccessible from the point of view of linguistic and logical tools at our disposal, because these tools are countable and the family $\wp(\mathbb{N})$ is uncountable. The notion of powerset is not absolute.

It is clear that one may talk about non-standard models of a theory only after the standard model itself has been adopted. This in turn implies that a theory in question is fully developed, that is is ultimately an axiomatic theory. Only at this stage can one notice (discover) that there are possibly many models interpreting the theory. Thus the investigation of non-standard models may begin only after a certain metatheoretical reflection has been carried out.

The first non-standard model of arithmetic was constructed by Thoralf Skolem in 1934. It is an interesting question as to whether Skolem was partly motivated by results concerning the scales of infinities introduced earlier by Cantor and, independently, by Du Bois-Reymond. Skolem's non-standard model of arithmetic, in which there exist non-standard, infinite numbers, was constructed with the use of an algebraic tool (in works by Hewitt and Łoś) later known as the ultrafilter construction. Skolem considered functions from natural numbers to natural numbers defined arithmetically and made use of an ultrafilter being an extension of the filter of cofinite sets of natural numbers. Elements of the model are classes of equivalence of such functions, where two functions are equivalent when the set of arguments for which they are equal belongs to the ultrafilter. Then it can be shown that the equivalence class of the identity function (i.e. the function $f(x) = x$ for all natural numbers x) is an element of the model which comes after all elements corresponding to the equivalence classes of all standard natural numbers.

There are several ways of proving the existence of non-standard models of arithmetic. A simple way is to use the compactness theorem. Another possibility makes use of the algebraic construction of an ultraproduct. One can also consider a full binary tree of expansions of arithmetic and show that each branch of this tree corresponds to a model of arithmetic – one of them is the standard model, while all others are non-standard ones. The details are described in chapter 6.

One may ask the question: to which domains in modern mathematics are the concepts of intended and standard model applicable at all? These notions seem to play less important role today than at the time when mathematicians started to develop the axiomatic method. In this respect let us observe the following:

1. *Arithmetic.* Second-order Peano arithmetic is categorical, its standard model is determined uniquely. First-order Peano arithmetic is a wild theory, in each infinite power κ it has a maximum possible number of models, i.e. 2^κ. It thus has continuum pairwise non-isomorphic countable models and only one of them is standard.

The Tennenbaum's theorem asserts that the standard model of PA is its only recursive model.

2. *Analysis.* Most research in analysis in the last century accepted the fields of real and complex numbers as standard number fields, but several other structures were also taken into account as numerical basis for developing analysis. The hyperreal field is currently used more and more often. While the domain of its applications is called *non-standard* analysis, the hyperreal field is becoming a kind of a standard structure.

3. *Set theory.* First-order Zermelo-Fraenkel set theory is incomplete, and if it is consistent (which cannot be proved inside it), then it has a huge spectrum of models, for instance differing in the values of cardinal numbers. The constructible universe is of course a distinguished model of set theory (and attractive to many "normal" mathematicians, meaning those who do not work in the foundations of set theory), but there is no pressure among mathematicians to restrict their attention to constructible sets only.

4. *Algebra.* Since the development of abstract algebra, algebraic considerations have certainly belonged to polymathematics, so the question about intended models looses sense here. There are several results characterizing certain algebraic structures either up to isomorphism or as minimal or maximal with respect to chosen properties. Here are a few examples:

   (a) The field of real numbers is the maximal Archimedean field.
   (b) The field of real numbers is the only field ordered in the complete way (up to isomorphism).
   (c) The field of surreal numbers includes all ordered fields as subfields.

5. *Geometry.* Euclidean geometry (as axiomatized by Hilbert) is categorical, so it has the standard (Cartesian) model. Absolute geometry is not categorical. Currently several systems of geometry are known and asking about standard models in this respect seems to miss the point. This is even more obvious in the case of topological investigations.

1.3 Classification and representation theorems

1.3.1 Classification theorems

Classification theorems are general results (in polymathematics) which characterize how many types of structure are present in the class of all possible structures of a given sort. Classifications are based on equivalence relations, for example two structures being isomorphic, homeomorphic, diffeomorphic, and so on. In some domains complete inventories of types of structure are possible, but sometimes a given classification applies not to all structures in the domain, thus enumerating a few exceptions (as in the case of finite simple groups). Here are some examples of well-known classification theorems:

1. *Classification of finite simple groups.* Every finite simple group is isomorphic to one of the following groups:

    (a) a cyclic group of prime order
    (b) an alternating group of degree at least 5
    (c) a simple group of Lie type (among which several further classifications are made)
    (d) one of the 26 sporadic simple groups.

2. *Classification of surfaces.* Every closed compact connected surface without boundary is homeomorphic with one of the three following:

    (a) two-dimensional sphere
    (b) connected sum of g thori, $g \geqslant 1$
    (c) connected sum of k projective planes, $k \geqslant 1$.

3. *Classification of Riemann surfaces.* Each simply connected Riemann surface is conformally equivalent to one of the following surfaces:

    (a) *elliptic* – the Riemann sphere $\mathbb{C} \cup \{\infty\}$
    (b) *parabolic* – the complex plane $\mathbb{C}$
    (c) *hyperbolic* – the open disc $D = \{z \in \mathbb{C} : |z| < 1\}$.

4. *Ostrowski Theorem.* Any field complete with respect to an Archimedean norm is isomorphic with either $\mathbb{R}$ or $\mathbb{C}$ and the norm is equivalent to the usual norm determined by the absolute value.

1.3.2 Representation theorems

Representation theorems, in turn, show in general that structures from a given collection can all be represented by certain (in a strictly defined sense) uniform structures. With a representation theorem at our disposal we are often able to prove much more easily theorems about the investigated objects. Examples of representation theorems are, among others:

1. *Stone Representation Theorem.* Each Boolean algebra is isomorphic to a field of sets.
2. *Cayley Representation Theorem.* Each finite group is isomorphic to some group of permutations.
3. *Mostowski Contraction Lemma.* Every extensional and well-founded structure is isomorphic to a transitive structure.
4. *Nash Theorem.* Every Riemann manifold can be isometrically embedded in some Euclidean space.
5. *Whitney Representation Theorem.* For any m-dimensional differential manifold there exists its embedding into a $2m$-dimensional Euclidean space ($m > 0$).
6. *Representation Theorem for Lie algebras.* Every Lie algebra over a field is isomorphic to a subalgebra of some associative algebra.
7. *Gödel Representability Theorem.* Every recursive function is representable in Peano arithmetic.

In a sense, each completeness theorem in logic is a kind of a representation theorem: one represents syntactic concepts related to deduction by mathematical structures connected with semantic entailment. Also algebraic characterization of elementary equivalence (using families of partial isomorphisms) serves as an example of representation of semantic notions.

1.3.3 Invariants

In 1872 Felix Klein presented in his *Vergleichende Betrachtungen über neuere geometrische Forschungen* (Klein 1872) a research program concerning classification of the systems of geometry. The main idea was based on investigations of groups of transformations. Invariants of such transformations determine the geometries in question. The program was very influential, not only in geometry.

We are not going to present a general definition of a complete system of invariants – for our purposes considering a few examples of invariants that illustrate the general idea shall suffice.

One may for instance ask which transformations in geometry act in such a way that:

1. straight lines are transformed into straight lines,
2. the surface area does not change,
3. the angles do not change, etc.

Similar questions may be asked for algebraic or topological structures. Furthermore, one may fix certain invariants and characterize all transformations with these invariants. For instance, the Euler characteristic and orientability form a complete system of invariants for the classification of all two-dimensional closed manifolds. Gauss' curvature is an invariant of local isometries of two-dimensional surfaces. A few examples of transformations and their invariants in the corresponding domains are given in the table below:

Domain	Transformation	Invariant
Topological spaces	Homeomorphism	Dimension
Sets	Translation	Lebesgue measure
Complex numbers	Multiplication	Absolute value
Ring of polynomials	Linear transformation of variables	Degree of a polynomial
Projective space	Projective transformation	Cross ratio of a quadruple of points

Transformations investigated in the theory of invariants form groups, as already mentioned. The next table gives examples of such groups, together with the systems of geometry (or another theory) determined by them and some of the invariants (not all) in question:

Group of transformations	Geometry	Some invariants
identity	of position	position
isometries	metrical	distance
similarities	of similarities	angle
affine	affine	colinearity
homeomorphisms	topology	connectedness
bijections	set theory	cardinality

The theory of invariants was developed by Salmon, Sylvester, Cayley. By the end of the 19th century Hilbert had proposed a general algebraic theory of invariants.

A theory in a logical sense is a fixed point of some operation of consequence, that is a set of formulas (in a fixed language) which is closed under accepted rules of inference. In mathematics the term theory has recently been understood as meaning an axiomatic theory, but there was a time when by a theory one understood the set of all true sentences about invariants.

1.4 Which mathematical objects are standard?

In mathematical literature we often encounter statements claiming that some object is *well behaving*, sometimes one also uses a comparative, such as: object X behaves better than object Y. For instance: Hausdorff spaces behave better than general topological spaces, Borel sets behave better than arbitrary sets, differentiable functions behave better than continuous functions, analytic functions (i.e functions from the class C^ω) behave better than smooth functions (i.e functions from the class C^∞), recursive functions behave better than arbitrary functions, and so forth. All such statements express our attitude toward accessibility to mathematical objects, familiarity with them and their usefulness in applications. Observe that:

1. *Well behaviour* of mathematical objects is always related to some investigated theory or its applications. There is nothing like *absolute well behaviour* – properties of objects are evaluated from a pragmatic point of view.

2. *Well behaviour* has nothing to do with *being in majority*. Analytic functions are rare among continuous functions, almost all functions are not differentiable anywhere, there are only $\aleph_0$ recursive

functions, etc. Well behaviour corresponds rather to the property of being a prototypic object in the considered domain. Such objects occurred as something like patterns at the beginning of the corresponding theory.

3. The property in question is also related to the history of mathematics. Objects may become well behaving only when a theory underlying them is sufficiently developed. For example, complex numbers were viewed as well behaving only after proving that they form a field.

One of the important procedures in mathematics is to represent objects from a given domain in so-called *normal*, *standard*, and *canonical* forms. This procedure often enables us to conduct more easily the deductions and calculations concerning these objects. Examples are:

1. *Logic.* Conjunctive and disjunctive normal forms, prefix normal form, Skolem normal form, etc.

2. *Set theory.* Cantor normal form (for ordinal numbers).

3. *Algebra.* Jordan normal form (for matrices over an algebraically closed field).

4. *Number theory.* Canonical representation of integers (as products of powers of prime numbers).

5. *Analysis.* Canonical differential forms.

6. *Formal languages.* Chomsky, Greibach, Kuroda normal forms.

7. *Recursion theory.* Kleene normal form.

We would like to draw a distinction between standard objects (elements of standard models, as described above) and *generic* objects and generic properties. A property holding for "typical" objects from a given class is called *generic* and objects possessing it are called *generic* objects from this class. In this sense, "generic" usually means the same thing as "almost all", where the latter term obtains a precise mathematical meaning, depending on the context. A dual concept to "almost all" is "negligible" and its meaning is derived from the meaning associated with "almost all". Here are a few examples illustrating these notions:

1. "Almost all" in measure theory (and also in probability theory) means "on a set with measure one" (for instance, Lebesgue measure). A dual notion ("negligible") is here of course a set with measure zero.

2. In topology "almost all" usually means "on an open dense set". A dual notion is a "nowhere dense set". Furthermore, a meagre set (a set of the first category) is a countable union of nowhere dense sets. A comeagre set is one whose complement is meagre, which means that it is the intersection of countably many sets with dense interiors. Sets which are not meagre are called sets of the second category. It should be added that sets of the second category need not be comeagre – a set may be neither meagre nor comeagre (and then it is of the second category).

Observe that "almost all" in the sense of measure theory does not coincide with "almost all" in the topological sense: for example the rational numbers form a set of Lebesgue measure zero and at the same time a dense subset in the set of all real numbers (with the usual topology).

In several domains of mathematics we talk about *degenerate* objects and *limiting* cases. An object is degenerate if some of its properties take the most possible values. This is not a precise definition, but it can be supplied with a lot of intuitive examples:

1. A point is a degenerate circle (with radius equal to zero).

2. A parabola can degenerate into two distinct or coinciding parallel lines.

3. An ellipse can degenerate into a single point or a line segment.

4. A hyperbola can degenerate into two intersecting lines.

5. In inversive geometry a line is a degenerate circle (with infinite radius).

6. A degenerate triangle is one with collinear vertices and zero area.

7. A sphere with radius zero is a point, a degenerate sphere.

8. A degenerate continuum consists of one point.

9. Examples of degenerate polygons are: a digon and a monogon.

In general, degenerate objects in some class belong to a class of simpler objects (e.g. a point as a degenerate circle). We distinguish between degenerate objects and objects called *exceptions*. We talk about exceptional objects usually in the case when they do not fit into some classification of all objects from a domain in question. For example, 26 sporadic groups form exceptions in the classification of all finite simple groups.

The term *exceptional* can be applied not only to objects that lack properties taken into account during classifications but also to such objects which are distinguished due to possession of regularities absent in all other objects from a given class. This is for instance the case with Platonic solids. In three dimensions there are exactly five Platonic solids, as has been well known from antiquity. They are convex polyhedra with faces formed by regular polygons such that at each vertex there is the same number of edges meeting at it. The Platonic solids are exceptional between convex polyhedra in the sense that they possess a lot of nice symmetries. Such regularly symmetric cells exist also in higher dimensions – for example in the fourth dimension there are six of them and in all higher dimensions there are exactly three of them.

Mathematical objects named *pathological* (sometimes also: *paradoxical*) appear either as unexpected and, moreover, unwilling, or they are constructed on purpose. Thus, there seem to be at least two typical situations in which one speaks about pathological objects in mathematics which were surprisingly unwilling:

1. *A clash with established intuitions.* Discoveries of new kinds of objects may contradict intuitions shared by the mathematicians in the given epoch. These new objects are treated with suspicion, as was the case for instance with negative and imaginary numbers at the beginning of their appearance in mathematics.

2. *New definitions of concepts formerly understood in an intuitive way.* When a new definition is proposed, then it may happen that it embraces objects which were not recognized before and whose properties differ from already familiar, prototypic objects. This was the case for example when a general definition of a function was proposed (a function understood as a set of pairs of objects). Mathematicians discovered then, among other things, that most functions are nowhere differentiable, which contradicted intuitions connected with commonly used functions.

A new definition may cause some properties considered paradoxical to become domesticated and treated as natural. As an example one can give here Dedekind's definition of infinite sets, which, in a sense, transformed a paradoxical property (being equinumerous with a proper subset of itself) into a design feature of precisely defined objects.

Pathological objects are also constructed on purpose, for instance in order to show the role of particular assumptions of theorems or to make our mathematical intuitions more sublime. Many examples can be found, for example in general topology.

Many objects considered as pathological or strange are already quite well described – a very nice example is the book Kharazishvili 2006 devoted to strange functions in real analysis.

A prototypic example of an object originally thought of as pathological and then becoming normal, standard, "domesticated" is the Cantor set. In recent times no professional mathematician would have considered it a pathology. This is due to its fundamental role in, for example, topology.

A few examples of "famous" pathological objects are:

1. *Alexander horned sphere*. This topological object is homeomorphic with the sphere S^2. However, it divides the space $\mathbb{R}^3$ in a different way to S^2: its inside is homeomorphic with the inside of S^2 but its outside is not homeomorphic with the outside of S^2.

2. *Exotic spheres*. We recall that an exotic sphere is a differentiable manifold M that is homeomorphic but not diffeomorphic to the standard Euclidean n-sphere. This means that M is topologically indistinguishable from the Euclidean sphere, but admits a smooth structure which is essentially different from the standard such structure.

3. *Exotic* $\mathbb{R}^4$. We recall that an exotic $\mathbb{R}^4$ is a differentiable manifold that is homeomorphic to the Euclidean space $\mathbb{R}^4$, but not diffeomorphic to it. Let us also recall that dimension 4 is exceptional in this respect: exotic structures do not occur on $\mathbb{R}^n$ for $n \neq 4$. But $\mathbb{R}^4$ itself admits a continuum of distinct smooth structures on it.

To sum up this section let us enumerate the types of mathematical objects discussed above:

1. *Standard*. Objects which are well recognized. Their investigation has a long tradition, they have a range wide of applications, the

knowledge about them is considered fundamental. Standard objects are well behaving, but they may nevertheless be in minority in the investigated domain. Examples: natural numbers, continuous functions.

2. *Generic.* Objects which are (in a specified sense) genuine, typical in the investigated domain. "Generic" usually means "almost all". Generic objects form thus a majority in the investigated domain. "Almost all" means usually "belonging to a set with measure one". An opposite property is "negligible", i.e. "belonging to a set with measure zero".

3. *Exception.* Objects which do not fit into a given classification. Example: sporadic groups.

4. *Extremal object.* Objects which have a certain property (or group of properties) to the extreme value (minimal or maximal). Example: Platonic solids as objects with nice properties concerning symmetry.

5. *Degenerate.* This is, in a sense, a property contrary to being generic. Examples: a point as a degenerate circle (or interval).

6. *Pathology.* There are two cases in which we talk about pathological objects. First, such an object can arise in an unexpected way (at a given moment in the history of mathematics), as for example the complex solution of a polynomial with rational coefficients (before the acceptance of complex numbers) or an irrational magnitude (at the time when Pythagoreans did not expect such a situation). Second, a pathological object may be created intentionally, on purpose. In this case we create (discover?) a pathology in order to show for instance the limitations of a given theorem or for making our intuitions about investigated objects more subtle. Several examples of this situation were typical in the early days of general topology.

7. *Counterexample.* Such objects show that the scope of a certain result is limited. Counterexamples need not be pathological but they often are. Nice collections of counterexamples exist in the main mathematical theories – (cf. Gelbaum and Olmsted 1990, 2003; Steen and Seebach 1995; Wise and Hall 1993). The role of counterexamples is discussed for example in the classical work Lakatos 1976.

8. *Surprise*. A discovered object may be surprising in a sense that it is unexpected but not unwilling (thus differing from a pathology). Examples of such surprises can be found in many domains, for example in logic (a finite logical matrix which is not finitely axiomatizable), analysis (exotic structures), and functional analysis (a Banach space without a basis), and so forth.

It should be stressed that the types of objects mentioned above do not form a classification of mathematical objects, but they are recognized with the essential use of pragmatic criteria. We think it reasonable to assume that there are no objects pathological in an absolute sense: calling an object *pathological* is a mood of speaking which reveals our attitude to the object in question and which is related to historical context.

Chapter 2

Historical remarks concerning extremal axioms

In this chapter we discuss the origins of extremal axioms in mathematics. The story goes back to the 19th century, when the first axioms of this kind were formulated: in geometry, arithmetic, algebra, and analysis. The special case of set theory is also taken into account. Specific extremal axioms will be discussed in more detail in chapters 5–7. Before that, it would be pertinent to recall the origins of such notions as: isomorphism, categoricity, completeness. Such is the subject of this chapter and the next. We pay special attention to the paper Carnap and Bachmann 1936 which was the first attempt at a general approach to extremal axioms.

2.1 Origin of the notion of isomorphism

When did the notion of *isomorphism* enter mathematical considerations? We recall briefly here the opinions of certain mathematicians, logicians and historians of mathematics.

Bourbaki ascribes awareness of this concept to Évariste Galois (Bourbaki 1980, 72). Indeed, it seems that only after recognition of the fact of existence different structures of a given general sort (e.g. groups) was it possible to ask questions whether they are indistinguishable and if so, then in what sense.

Morris Kline also attributes the concept in question to Galois and its development to his followers:

> Galois made the largest step in introducing concepts and theorems about substitution groups. His most important concept was the

> notion of a normal (invariant or self-conjugate) subgroup. Another group concept due to Galois is that of an isomorphism between two groups, This is one-to-one correspondence between the elements of the two groups such that if $a \cdot b = c$ in the first one, then for the corresponding elements in the second $a' \cdot b' = c'$. (Kline 1972, 765–766)
>
> The knowledge of (finite) substitution groups and their connection with the Galois theory of equations up to 1870 was organized in a masterful book by Jordan, his *Traité des substitutions et des équations algébriques* (1870). [...] The *Traité* presented new results and made explicit for substitution groups the notions of isomorphism (*isomorphisme holoédrique*) and homomorphism (*isomorphisme mériédrique*), the latter being a many-to-one correspondence between two groups such that $a \cdot b = c$ implies $a' \cdot b' = c'$. (Kline 1972, 767)
>
> Though Eugen Netto, in his book *Substitutionstheorie und ihre Anwendung auf die Algebra* (1882), confined his treatment to substitution groups, his wording of his concepts and theorems recognized the abstractness of the concepts. Beyond putting together results established by his predecessors, Netto treated the concepts of isomorphism and homomorphism. The former means one-to-one correspondence between two groups such that if $a \cdot b = c$, where a, b, and c are elements of the first group, then $a' \cdot b' = c'$, where a', b', and c' are the corresponding elements of the second group. A homomorphism is a many-to-one correspondence in which again $a \cdot b = c$ implies $a' \cdot b' = c'$. (Kline 1972, 1139)

Two papers by John Corcoran (Corcoran 1980, 1981) are devoted to the origin of the notions of isomorphism and categoricity. Corcoran stresses the fact that even if one accepted the fact that categoricity implied completeness (at the beginning of the 20th century), there was no conviction that such a fact should be proved:

> Veblen 1904 presents an axiomatization of Euclidean geometry and "proves" its completeness. The "methods" that Veblen uses are readily adaptable to the then-available axiomatizations of number theory and of analysis so as to yield completeness immediately without the need for mathematical results beyond those then-available. Thus, in 1904 the completeness of the three main branches of mathematics was "established".
>
> It is a disappointment for modern readers to find that Veblen 1904 (esp. p. 346) infers completeness from categoricity without an argument. Moreover, careful comparative study of Huntington 1902, Veblen 1904 and Huntington 1905 makes it highly probable, if not

> certain, that Veblen took completeness and categoricity to be coextensive, again without argument. (Corcoran 1981, 116–117)

The paper Corcoran 1980 discusses how the early mathematical logicians gradually achieved the solution of the problem of characterization of a given structure uniquely with respect to isomorphism. He considers also a certain kind of completeness (*atom-completeness*) which, in a suitable language (*slightly-augmented language*), together with a special sort of induction implies categoricity.

Corcoran writes that recently the distinction between categoricity and completeness had related to the clear recognition of some other metamathematical concepts, notably those from recursion theory (which of course was not evident at the beginning of the 20th century):

> Some early postulate theories (e.g. Veblen 1904, 346) were clear about the *conceptual* distinction between characterization and axiomatization *and* about the possibility of an axiomatically inadequate categorical characterization at least to the extent of explicitly mentioning the possibility that a categorical characterization need not be a (deductively) complete axiomatization. This possibility, of course, entails the possibility of 'logically' incomplete underlying logics (wherein semantic consequences of a given set of axioms are not deducible as theorems).
>
> At that time, however, there was no suspicion of the idea of recursiveness, nor, *a fortiori*, of the relevance of recursiveness and recursive enumerability to problems of axiomatizability. Now we can see that if the set of truths of an interpretation is not recursively enumerable then there is no way to give a complete axiomatization even if the logic is complete. It follows immediately from the Gödel incompleteness result that a (recursive) set of sentences which provides a categorical characterization need not to provide a complete axiomatization. Moreover, in such cases, it follows that there are infinitely many other categorical characterizations each of which provides a better axiomatization in the sense of providing the basis for the deduction of additional theorems not deducible from the first characterization. (Corcoran 1981, 203)

Categoricity depends on the language used, which is evident. In the first-order language without the predicate of identity no structure can be described in a categorical way (due to the Löwenheim-Skolem theorem). If we add the identity predicate to that language, then only finite structures obtain categorical descriptions.

It is worth mentioning that an important result concerning categoricity is an early theorem by Cantor, stating that any two countable dense linear orderings without ends are isomorphic. The *back-and-forth technique* (or a *zig-zag technique*) used in the proof of this theorem has since found numerous applications.

Sometimes we are interested not in the mere existence of an isomorphism between two structures but we also ask about the number of such possible isomorphisms. For instance, in the case of two Peano algebras there exists exactly one isomorphism between them.

2.2 The notions of completeness

One of the desiderata of early axiomatical description of mathematical theories was an exhaustive and unique description of particular mathematical domains. Uniqueness was related to the notion of isomorphism: isomorphic structures are indistinguishable. Exhaustiveness, in turn, was characterized via several notions of completeness. The latter property was understood initially in an informal way. Sometimes it was also identified with categoricity. Only after the development of formal logic and the emergence of metalogic was it possible to provide precise formulations of different notions of completeness, in particular, to distinguish between its syntactic and semantic versions.

The main authors who contributed to this issue were at first Giuseppe Peano, Richard Dedekind, David Hilbert, Abraham Fraenkel, Edward Huntington, Oswald Veblen, and Rudolf Carnap and later also Kurt Gödel and Alfred Tarski. A concise and adequate discussion of this development is presented in the paper Awodey and Reck 2002a. The authors enumerate several notions of completeness:

1. *Semantic completeness.* A theory T is *semantically complete* (with respect to a given semantics) if it satisfies any of the following equivalent conditions:

    (a) For all sentences φ and all models M, N of T, if $M \models \varphi$, then $N \models \varphi$.

    (b) For all sentences φ, either $T \models \varphi$ or $T \models \neg\varphi$.

    (c) For all sentences φ, either $T \models \varphi$ or $T \cup \{\varphi\}$ is not satisfiable.

    (d) There is no sentence φ such that both $T \cup \{\varphi\}$ and $T \cup \{\neg\varphi\}$ are satisfiable.

2. *Deductive completeness.* A theory T is *deductively complete* (with respect to a given deductive consequence $\vdash$) if it satisfies any of the following equivalent conditions:

   (a) For all sentences φ, either $T \vdash \varphi$ or $T \vdash \neg\varphi$.

   (b) For all sentences φ, either $T \cup \{\varphi\}$ or $T \cup \{\neg\varphi\}$ is inconsistent.

   (c) There is no sentence φ such that both $T \cup \{\varphi\}$ and $T \cup \{\neg\varphi\}$ are consistent.

3. *Relative completeness.* A theory T is *relatively complete* (with respect to a given set S of sentences from the language of T), if every sentence belonging to S is provable from T.

4. *Logical completeness.* A theory is *logically complete* (with respect to a given semantics) if for all sentences φ, if $T \models \varphi$, then φ is provable from T.

The above formulations fulfill the contemporary standards of correctness. However, as is evident in the source texts of Dedekind, Fraenkel and Carnap (to name just a few), there was a long path before formulations were achieved with such a degree of clarity. The prototypes of several notions of completeness emerged at the time when the systems of formal logic were still in *statu nascendi.* The overall view of formal logic was also in the process of reformulation.

A few (intuitive) commentaries to the above definitions are in order. The property 1(a) means that all models of the theory in question are semantically indistinguishable, that is they have exactly the same set of sentences true in them. This indistinguishability relation is called *elementary equivalence*, as was already mentioned in the previous chapter.

The property 1(b) expresses the fact that the theory T determines semantically each sentence φ of its language, in the sense that either φ itself or its negation $\neg\varphi$ is a semantical consequence of T.

Observe that $T \models \neg\varphi$ is equivalent to non-satisfiability of $T \cup \{\varphi\}$ and hence the property 1(c) may be thought of as a reformulation of 1(b).

The property 1(d) means that there is no sentence φ such that T "forks" at φ, in the sense that both theories $T \cup \{\varphi\}$ and $T \cup \{\neg\varphi\}$ have models.

Turning now to deductive completeness let us note that 2(a) means that each sentence φ is deductively determined by T, meaning that either φ or $\neg\varphi$ can be proved in T.

If the system of logic under question is sound and complete (as in the case of first-order logic), then the semantic and syntactic notions of completeness of theories are equivalent. This is not true in general, for example in the case of second-order logic.

The table below shows which deductive and semantic completeness properties can be attributed to certain well-known theories (we will come back to these theories in the second part of the book):

Theory	Semantical completeness	Deductive completeness	Categoricity
First-order Peano arithmetic	no	no	no
Second-order Peano arithmetic	yes	no	yes
First-order theory of completely ordered field	no	no	no
Second-order theory of completely ordered field	yes	no	yes
Tarski's first-order theory of real numbers	yes	yes	no

The notion of relative completeness defined in item 3 above is of special importance, because the term 'provable' used in its wording may be related not only to deduction in some fixed deductive system but also to the informal notion of mathematical proof used in mathematical practice well before the codification of systems of formal deduction. Finally, the notion of logical completeness is a special case of relative completeness.

As we will see in the next two sections, mathematicians working on early axiomatics were approaching the meaning of the above notions. Sometimes they also related the emerging ideas of completeness to the extremal axioms concerning the investigated domains.

In the paper Grzegorczyk 1962 certain special types of completeness are considered – among others so-called *descriptive completeness*. If we have the possibility of attaching a name to each element of a model (e.g. an individual constant or a closed term from the language of the theory), then we obtain additional possibilities of establishing isomorphism between structures.

An important contribution to the issue of categoricity and completeness is the text of the talk given by Alfred Tarski at the Harvard Logic Club in January 1940, and published under the title *On the completeness and categoricity of deductive systems* (Tarski 1940).

The paper Lindenbaum and Tarski 1936 discusses, among others, conditions under which the semantical indistinguishability of models (completeness) implies that they are isomorphic. These problems are discussed more recently also in papers devoted to the so-called Fraenkel-Carnap property (cf. Weaver and George 2002, 2003, 2005).

We will come back to the problems of relations between categoricity and completeness in chapter 4 of this book, where we will discuss results from classical and modern model theory concerning dependencies between the number of non-isomorphic models and the number of abstraction classes of the relation of elementary equivalence.

2.3 Extremal axioms: first formulations

Two very important number domains – those of natural and real numbers – were axiomatized in the late 19th century. In the first case we have axiomatizations by Hermann Grassmann, Giuseppe Peano, Gottlob Frege and Richard Dedekind, given independently of each other. In the second case we have more attempts, but the most influential ones are the axiomatization by David Hilbert and the constructions proposed by Georg Cantor and Richard Dedekind. These topics will be discussed in more detail in chapters 5 and 6 of this book; here we limit ourselves to a short indication only.

Hilbert's *Grundlagen der Geometrie* (first edition: Hilbert 1899) became a paragon for axiomatic descriptions widespread in mathematics from the end of 19th century. Hilbert's system contains *The Axiom of Completeness* whose very formulation expresses the maximality condition imposed on the domain of the theory. In the 1902 edition of the *Grundlagen der Geometrie* the axiom of completeness had the following form:

> AXIOM OF COMPLETENESS. *(Vollständigkeit): To a system of points, straight lines, and planes, it is impossible to add other elements in such a manner that the system thus generalized shall form a new geometry obeying all of the five groups of axioms. In other words, the elements of geometry form a system which is not susceptible of extension, if we regard the five groups of axioms as valid.*

In the later editions of *Grundlagen der Geometrie* this axiom – referring to points, lines and planes – was replaced by the *Axiom of Line Completeness* which referred to lines only, as objects which cannot be expanded without violation of the other axioms. One could say that the first formulation of Hilbert's axiom did not follow the modern standards of logical correctness, but one should keep in mind that such standards were only worded with precision much later. A similar remark applies to the axiom of completeness in the axiomatic system of real numbers proposed in Hilbert 1900.

The American Postulate Theorists – Oswald Veblen and Edward Huntington – published papers at the beginning of the 20th century on the foundations of geometry in which versions of extremal axioms are mentioned (Huntington 1913, Veblen 1904). Huntington also published a paper about continuous magnitudes (positive real numbers) where he uses a form of the axiom of continuity (Huntington 1902). These authors used explicitly the notion of categoricity and formulated interesting claims concerning semantic completeness.

Minimality conditions, in turn, can be observed in the case of Peano's axiomatization of natural numbers (Peano 1889) and in Dedekind's construction of natural numbers (Dedekind 1888). This role is played in Peano's system by the (second-order) axiom of induction which says, roughly speaking, that if a set of natural numbers contains the first element of the domain and if it contains any element, then it contains the successor of that element, then the set in question is identical to the set of all natural numbers. Dedekind's construction of a simply infinite chain (a minimal set containing the initial element and closed with respect to the successor operation) also serves the purpose of minimality of domain. With this construction at hand, Dedekind was able to prove the axiom of induction. It should be added that Dedekind explicitly stated the necessity of the exclusion of "alien intruders" from the domain of all natural numbers (see his famous letter to Keferstein, published in Ewald 1996).

In the cases of geometric space and arithmetic continuum, maximality conditions were proposed while in the case of natural numbers minimality conditions were considered adequate. The situation was quite different in set theory: both minimality and maximality conditions were considered for the set theoretic universe.

Abraham Fraenkel has formulated his axiom of restriction in in set theory in the 1920s. This was supposed to express the idea that the universe of sets should contain only those sets whose existence could be proved from the axioms (Fraenkel 1922, 1928). This was a condition of minimality of the universe of all sets and similarly to Hilbert's axiom of completeness in geometry was not formulated in the object language.

John von Neumann formulated an axiom concerning proper classes in his system presented in von Neumann 1925 and stating that: *All classes which are not sets can be put into a one-one correspondence with each other (respectively, with the class of all sets)*. This is a kind of a minimality condition.

Ernst Zermelo, in turn, expressed the idea that set theory (understood as the system from Zermelo 1930) must not be restricted to a single interpretation but rather it should be a transfinite hierarchy of "normal domains". Moreover, he postulated the existence of a transfinite sequence of strongly inaccessible cardinals which was a kind of maximality condition.

Some other extremal axioms formulated in the discussed period (up to 1936, that is the publication of Carnap and Bachmann 1936) had lesser influence than the ones listed above. This applies, for example, to the proposal in Finsler 1926, critically discussed in Baer 1928. Further extremal axioms, formulated after 1936 (for instance Gödel's axiom of constructibility, Suszko's axiom of canonicity, axioms of the existence of large cardinal numbers, and Ehrlich's generalization of the Dedekind's continuity axiom) will be discussed in the second part of the book.

2.3.1 Examples of maximal axioms for real numbers

The axiomatic theories of real numbers were developed in the 19th century. Let us consider three examples.

Richard Dedekind

The work *Stetigkeit und irrationale Zahlen* was published by Dedekind in 1872. Here the set of all rational numbers $\mathbb{Q}$ is the starting point. Dedekind makes an essential use of the order properties of rational numbers in his definition of the set of real numbers. His construction is applicable to arbitrary ordered sets, but let us first consider only the case of the ordered system $(\mathbb{Q}, <)$: the countable set of all rational numbers densely ordered by the usual less-than relation $<$.

By a *Dedekind cut* (in the system $(\mathbb{Q}, <)$) we mean any pair (A, B) of non-empty subsets of $\mathbb{Q}$ such that $A \cup B = \mathbb{Q}$, $A \cap B = \emptyset$ and $a < b$

for all $a \in A$ and $b \in B$. If (A, B) is a Dedekind cut, then A is called its *lower class* and B is called its *upper class*. Every rational number is associated with some Dedekind cut. However, there exist Dedekind cuts which do not correspond to any rational number, for example:

$$\{x \in \mathbb{Q} : x < 0 \text{ or } (0 \leqslant x \text{ and } x^2 < 2)\}.$$

Dedekind defined the real numbers as the family of all Dedekind cuts (in the system $(\mathbb{Q}, <)$). The ordering of this set, as well as the arithmetical operations on it are defined with the use of the corresponding ordering and arithmetical operations on rational numbers. Most importantly, Dedekind showed that real numbers defined in this way are ordered in the complete way (meaning that their ordering is continuous).

Dedekind cuts can be defined for any linearly ordered set $(X, <)$. A pair (A, B) of non-empty subsets of X is a cut in $(X, <)$, if $A \cup B = X$, $A \cap B = \emptyset$ and $a < b$ for all $a \in A$ and $b \in B$. The cuts (A, B) can be of the following forms:

1. There is the greatest element in A and the smallest element in B.
2. There is the greatest element in A but there is no smallest element in B.
3. There is the smallest element in B but there is no greatest element in A.
4. There is no greatest element in A and there is no smallest element in B.

The situation described in 1 is present, for example, in the case of the integers and does not take place when the ordering $<$ is dense. The remaining cases all occur in the case of densely ordered rational numbers. Case 4 is what is commonly called a *gap* in the ordering under question. Dedekind's construction of the real numbers understood as cuts of the system $(\mathbb{Q}, <)$ thus shows that the ordering of the reals has no gaps in this sense.

In §3 of his treatise Dedekind writes:

> When we compared the domain R of rational numbers with a straight line, we found in the former a gappiness, incompleteness, discontinuity; but to the straight line we ascribe absence of gaps, completeness, continuity. In what then does this continuity consist? (Dedekind 1872, citing the translation in Ewald 1996 vol. II, 771)

Dedekind's answer to that question is given in the next paragraph of the text and the author claims that this is exactly the essence of continuity:

> 'If all points of the straight line fall into two classes such that every point of the first class lies to the left of every point of the second class, then there exists one and only one point which produces this division of all points into two classes, this severing of the straight line into two portions.' (Dedekind 1872, citing the translation in Ewald 1996 vol. II, 771)

This is what is now commonly called Dedekind's axiom of line completeness. Observe that the above principle is applied to a geometrical object (a straight line). Dedekind's construction of the real numbers provides then a possibility of a one-one correspondence between real numbers in his sense and points of a straight line.

David Hilbert

The short note *Über den Zahlbegriff* (Hilbert 1900) contains an explicit axiomatization of the system of real numbers. There are four groups of axioms: I. Axioms of linking, II. Axioms of calculation, III. Axioms of ordering, IV. Axioms of continuity. The last group consists of the Archimedean axiom and the following extremal axiom (citing the translation of the Hilbert paper in Ewald 1996, vol. II, 1094):

> **IV 2.** (*Axiom of Completeness.*) It is not possible to add to the system of numbers another system of things so that the axioms I, II, III, and IV 1 are also satisfied in the combined system; in short, the numbers form a system of things which is incapable of being extended while continuing to satisfy all the axioms.

Hilbert adds that the axioms of group IV imply Bolzano's theorem about the existence of a point of condensation and therefore his system is in agreement with the usual system of real numbers. The axiom of completeness in this system expresses the idea that the system of real numbers is the maximal Archimedean field.

Edward Huntington

The paper *A complete set of postulates for the theory of absolute continuous magnitude* (Huntington 1902) contains a set of postulates characterizing positive real numbers. Huntington words the goal of his work as follows:

> The object of the work which follows is to show that these six postulates form a *complete set*; that is, they are (I) *consistent*, (II) *sufficient*, (III) *independent* (or *irreducible*). By these three terms we mean: (I) there is at least one assemblage in which the chosen rule of combination satisfies all the six requirements; (II) there is essentially *only one* such assemblage possible; (III) none of the six postulates is a consequence of the other five. (Huntington 1902, 266)

This concise description summarizes the main features of the works of the American Postulate Theorist under consideration. Huntington proves the consistency of this set of postulates showing that the positive real numbers with the operation of addition satisfy all the postulates in question.

The axiom of continuity in this system is formulated on the page 267 of Huntington 1902 and takes the following form:

> *Postulate 5*: If S is any infinite sequence of elements a_k such that $a_k < a_{k+1}$, $a_k < c$ $(k = 1, 2, 3, \ldots)$ (where c is some fixed element), then there is one and only one element A having the following two properties:
>
> 1. $a_k \leqslant A$ whenever a_k belongs to S;
> 2. if y and A' are such that $y + A' = A$, then there is at least one element of S, say a_r, for which $A' < a_r$.

Huntington then adds that this formulation is related to the principle used in the lectures of Weierstrass devoted to the definition of an irrational number.

In the footnote at page 210 of *A set of postulates for ordinary complex algebra* (Huntington 1905) the author remarks:

> In the case of any categorical set of postulates one is tempted to assert the theorem that if any proposition can be stated in terms of the fundamental concepts, either it is itself deducible from the postulates, or else its contradictory is so deducible; it must be admitted, however, that our mastery of the processes of logical deduction is not yet, and possibly never can be, sufficiently complete to justify this assertion.

Huntington also recalls Hilbert's axiom of completeness for the real numbers in this footnote. Observe that in the quotation above Huntington speculates that categoricity may imply syntactic completeness. The formal machinery required for deciding such questions was developed much later.

2.3.2 Examples of minimal axioms for natural numbers

Natural numbers belong to the most fundamental objects in mathematics. Let us present two examples of their formal characterization carried out in the 19th century.

Richard Dedekind

The famous Dedekind's paper *Was sind und was sollen die Zahlen?* was published in 1888. This essay is also important from the point of view of the development of set theory. Dedekind introduced in this work the notion of *infinite set*: a set is infinite (in Dedekind's sense), if it can be one-one mapped onto its proper subset.

Dedekind gives in that work a categorical characterization of natural numbers (infinite systems, in his terminology). Theorems 132 and 133 deal with the *similarity* of infinite systems, which corresponds to their isomorphism. One may deduce from his Remark 134 that he considers semantic completeness of the system a consequence of its categoricity (but one does not yet have the logical tools to express this relationship overtly). Categoricity of a system was for Dedekind a necessary condition for adequacy of its description.

Let S be any set and $f : S \to S$ a function defined on S. We say that $K \subseteq S$ is a *chain* (in S and with respect to f), if $f(K) \subseteq K$, which means that the image of K with respect to f is included in K. We say that a set S is *simply infinite*, if there exists a function $f : S \to S$ and an element $a \in S$ such that:

1. $f(S) \subseteq S$
2. $a \notin f(S)$
3. if $f(x) = f(y)$, then $x = y$
4. S is the least set containing the element a and closed with respect to f.

Dedekind introduces a certain simply infinite system N with the *base element* 1 and an *ordering* determined by the function ϕ and calls this system *natural numbers*. Thus the system of natural numbers was defined as the smallest simply infinite chain (that is the intersection of all simply infinite chains).

Theorem 132 states that all simply infinite sets are similar (isomorphic) to the system N and theorem 133 establishes that each system which is similar to N is simply infinite. Dedekind claims (Remark 134) that it is clear that every theorem concerning natural numbers holds true for all simply infinite systems. Dedekind proves the principle of induction for his system.

It should be noted that Dedekind did not use any system of formal logic in his investigations. The systems of Peano and Dedekind appeared in print at almost the same time (Dedekind 1888, Peano 1989).

Giuseppe Peano

In 1889 Giuseppe Peano proposed an axiom system for the arithmetic of natural numbers in *Arithmetices principia, nova methodo exposita* (Peano 1889). This essay is written in a formal language proposed by Peano which was later accepted also by other logicians (notably by Whitehead and Russell in the *Principia Mathematica*). Theorems are proved in that language, though Peano did not specify explicitly the rules of inference.

There are three arithmetical primitive terms in Peano's system: the set of all natural numbers $\mathbb{N}$, the constant 1 (unity) and the successor function, which is one-argument function from $\mathbb{N}$ to $\mathbb{N}$. Let the successor of n be denoted by n^*. The system includes also the predicate of identity and the binary predicate $\in$ corresponding to the relation of being an element. The latter is understood here in an intuitive way – Peano's system from 1889 was introduced before the first axiomatization of set theory by Zermelo in 1908.

Peano's original system has nine axioms. However, if we omit axioms concerning the identity predicate, then the purely arithmetical axioms are as follows:

1. $1 \in \mathbb{N}$
2. $n \in \mathbb{N} \to n^* \in \mathbb{N}$
3. $n \in \mathbb{N} \to n^* \neq 1$
4. $(n \in \mathbb{N} \wedge m \in \mathbb{N} \wedge n^* = m^*) \to n = m$
5. $(Z \subseteq \mathbb{N} \wedge 1 \in Z \wedge \forall n(n \in \mathbb{N} \to n^* \in \mathbb{N})) \to Z = \mathbb{N}$

Here Z is a second-order variable running over sets of natural numbers. Hence the last axiom (axiom of induction) may be written as a single formula in the language of the theory. The system characterizes $\mathbb{N}$ as a minimal set containing 1 and closed with respect to the operation of successor. One might say that the axioms of this system express the fact that natural numbers form the least infinity.

One can define in a unique way operations of addition and multiplication in this system and show that these operations, applied to the elements of $\mathbb{N}$, have their results also in $\mathbb{N}$.

One can also prove that this axiom system is categorical. The operations of addition and multiplication can be defined on the basis of the above axioms.

In recent times a *Peano algebra* has been understood as meaning any system (A, a, f) such that:

1. $a \in A$
2. $f : A \to A$
3. $a \notin rng(f)$
4. f is an injection
5. For any set $X \subseteq A$, if the following two conditions hold:
   (a) $a \in X$
   (b) $\forall x(x \in X \to f(x) \in X)$,

   then $X = A$.

The existence of Peano algebras follows from the axiom of infinity in set theory. Any two Peano algebras are isomorphic (and there exists exactly one isomorphism between them).

2.3.3 Examples of maximal axioms for geometry

Hilbert's axiom of completeness in geometry inspired other mathematicians. Let us consider two further examples of systems of geometry proposed at the beginning of the 20th century.

In the first decades of the 20th century in *Transactions of the American Mathematical Society* there appeared several works which are distinguished by the accepted methodology as well as the choice of problems from the foundations of mathematics. They were devoted to the construction of axiom systems for chosen mathematical structures (systems of geometry, groups, number systems, and so forth). Moreover, most of these works explicitly mentioned the role of such notions as categoricity and completeness. Their authors are named the *American Postulate Theorist*. The most distinguished among them are: Eliakim Hastings Moore, Edward Vermilye Huntington, Oswald Veblen and Leonard Eugene Dickson.

The works of the American Postulate Theorists fulfilled certain common methodological standards:

1. *Choice of structure.* In the case of Euclidean geometry and the system of continuous magnitudes (corresponding to the real numbers) a mathematical structure is given in advance and the postulates were proposed in order to describe it in a unique way. In the case of, say, groups, the proposed postulates are thought of as characterizing the whole class of groups.

2. *Choice of primitive concepts.* Several alternative sets of primitive concepts were taken into account and compared.

3. *Formulation of postulates.* Alternative versions of postulates were investigated and compared with respect to their simplicity.

4. *Definitions of the further concepts.* All further concepts, beside the primitive ones, were introduced by explicit definitions.

5. *The consistency problem.* In order to show that a given system of postulates is consistent the authors presented a model of it (a structure in which all the postulates are true).

6. *Independence of the postulates.* This property of the postulates was established in a well-known way: a sentence ψ is independent of the set of sentences Ψ if there exists a structure in which all sentences from Ψ are true and ψ is false.

7. *The categoricity problem.* Whenever possible, (attempts at) proofs of the categoricity of the system in question were provided.

8. *Reflections concerning completeness.* This concept was not given a precise definition, but the American Postulate Theorists have formulated interesting remarks about its informal version.

Edward Huntington and Oswald Veblen proposed original systems of geometry in the papers Huntington 1913 and Veblen 1904. We are not going to discuss their systems in detail here, and shall only recall their formulations of the axioms of completeness.

The primitive terms of the system presented in Veblen 1904 are: *point* and *ordering*. The latter is a ternary relation of *lying between*. Straight lines and planes, as well as all geometrical objects usually considered become definable. The axiom of completeness is introduced with a commentary:

> If there exist infinitely many points, then there exists a pair of points AC such that if $[\sigma]$ is any infinite set of segments of the straight line AC with the property that each point which is the point A or the point C or a point of the segment AC is a point of some segment from $[\sigma]$, then there exists a finite subset $\sigma_1, \sigma_2, \dots, \sigma_n$ of the set $[\sigma]$ with the same property. [...]
>
> The proposition here adopted as the continuity axiom is referred to by Schoenflies as the Heine-Borel theorem. So far as I know, it was first stated formally (as a theorem of analysis rather than of geometry) by Borel in 1895 but is involved in the proof of the theorem of uniform continuity given by Heine in 1871. The idea of its equivalence with the Dedekind cut axiom was the result of a conversation with Mr. N.J. Lennes. (Veblen 1904, 347–348)

Veblen axioms have simpler formulations than those in the systems of Pieri or Hilbert. However, Veblen was wrong in the claim that he could define the congruence of segments (which is a four argument relation) in terms of the ternary relation of betweenness. For comments concerning this problem see for example Hodges 2007, Smith 2010, and Scanlan 2003. Scanlan quotes the opinion of Tarski from Tarski 1956 (reprinted in Tarski 1986b) on this subject:

> ... when we analyze the argument in [*Veblen 1904*] which is used to justify this claim, we notice that it actually leads to the opposite conclusion – to the statement that [betweenness] cannot serve as the only primitive notion for Euclidean geometry. In fact, in [*Veblen 1904*] it is proved correctly that, in terms of [betweenness], we can define various notions which have the same properties as the ordinary metric notions... except that they are relativized to a few arbitrarily chosen points and vary with these points. Hence it is seen that Euclidean geometry (with any given number of dimensions) has various models in which the relation [betweenness] is the same while the corresponding metric notions differ. (Tarski 1986b vol. 3, 618, citing Scanlan 2003, 311)

Of some interest to our main subject are Veblen's metamathematical remarks. In the work in question, Veblen uses the term *categoricity* (replacing Huntington's earlier term *sufficiency*). Similarly to Dedekind, he maintains that categoricity implies semantic completeness (but without proving this). He says in this context:

> Inasmuch as the terms *point* and *order* are undefined one has a right, in thinking of the propositions, to apply the terms in connection with any class of objects of which the axioms are valid propositions. It is part of our purpose however to show that there is *essentially only one* class of which the twelve axioms are valid. In more exact language, any two classes K and K' of objects that satisfy the twelve axioms are capable of a one-to-one correspondence such that if any three elements A, B, C of K are in the order ABC, the corresponding elements of K' are also in the order ABC. Consequently any proposition which can be made in terms of points and order either is in contradiction with our axioms or is equally true of all classes that verify our axioms. The validity of any possible statement in these terms is therefore completely determined by the axioms; and so any further axiom would have to be considered redundant. Thus, if our axioms are valid geometrical propositions, they are sufficient for the complete determination of Euclidian geometry. (Veblen 1904, 346)

To the sentence before the last one above there is attached a footnote: *Even were it not deducible from the axioms by a finite number of syllogisms.* Here by *syllogism* one should understand proof methods accepted in the mathematics of that period.

Another important remark made by Veblen concerns categoricity of his system:

> A system of axioms such as we have described is called *categorical*, whereas one to which it is possible to add independent axioms (and which therefore leaves more than one possibility open) is called *disjunctive*. The categorical property of a system of propositions is referred to by Hilbert in his "Axiom der Vollständigkeit", which is translated by Townsend into "Axiom of Completeness". E.V. Huntington, in his article on the postulates of the real number system, expresses this conception by saying that his postulates are sufficient for the *complete definition* of essentially a single assemblage. It would probably be better to reserve the word *definition* for the substitution of one symbol for another, and to say that a system of axioms is categorical if it is sufficient for the complete *determination* of a class of objects or elements. (Veblen 1904, 346–347)

Let us also add that in a popular article Veblen 1906 we find also a very interesting remark showing that Veblen was in a sense aware of the fact that semantic relations may escape from a purely syntactic description:

> But if [a proposition] is a consequence of the axioms, can it be derived from them by a syllogistic process? Perhaps not. (Veblen 1906, 28, citing Awodey and Reck 2002a, 19)

The paper Huntington 1913 was published not in the *Transactions of the American Mathematical Society* but in *Mathematische Annalen*. It nevertheless also belongs to a series of papers following the directions accepted by the American Postulate Theorists. Huntington explains in the *Prefatory Note* of this paper some important differences between his approach and certain systems of geometry proposed at that time:

> The chief points of difference between the present set of postulates and the well known sets given by Pasch, Veronese, Peano, Pieri, Hilbert, Veblen, Schweitzer, and others are: 1) the use of solid body instead of the point as an undefined concept; 2) the extreme simplicity of the undefined relation of inclusion; the systematic definitions of the straight line, the plane, and the 3-space, which can be readily extended, if desired, to space of n dimensions; and 4) the attempt to separate the 'existence postulates' from the postulates expressing 'general laws'. (Huntington 1913, 523)

The system from Huntington 1913 has two primitive terms: the set K of all *spheres* and the relation R of *inclusion* (holding between spheres). The notion of a *point* becomes definable, as well that of a *segment*. A sphere A is a point, if there is no sphere included in it. Notice that this definition does not say anything about – intuitively speaking – the size of points, it only says that points are minimal elements in the domain of the relation of inclusion.

Huntington stresses the originality of his definition of a segment. If X is a point such that each sphere which contains the points A and B (that is such that A and B are included in it) contains X as well, then one says that X belongs to the *segment* $[AB]$ (or $[BA]$).

Among the *existential* postulates accepted by Huntington one can find the following counterpart of Dedekind's axiom concerning a point on a line:

> If S_1, S_2,…,S_n,… is an infinite sequence of spheres, each of which lies within the preceding one, then there exists a point X which lies within them all. (Huntington 1913, 539)

It should perhaps be stressed here that Huntington considers the categoricity property in an abstract way, not related to any a priori given intended model:

> *Sufficiency of the postulates to determine a unique type of system.* A third and most important part of our work is to show that any two systems (K, R) which satisfy all the postulates are *formally equivalent*, or *isomorphic*, with respect to the variables K and R. This means that if (K', R') and (K'', R'') are any two 'geometrical systems' – that is, any two systems that satisfy all the postulates – then it will be possible to set up a one-to-one correspondence between the elements A', B', C', ... of K' and the elements A'', B'', C'', ... of K'' in such a way that whenever A' and B' satisfy the relation $A'R'B'$, then the corresponding elements A'' and B'' will satisfy the relation $A''R''B''$. By the establishment of this isomorphism, we show that *any theorem involving only the variables K and R, which is true in one of the systems will be true in the other also*; hence *all the systems* (K, R) *which satisfy the postulates may be said to belong to a single type.*
>
> With the proof of this theorem, the series of deductions which we draw from the postulates is brought to a natural conclusion; and in view of this theorem, we may then define *abstract geometry* as the *study of the properties of a particular type of system* (K, R), *namely, that type which is completely determined by the Postulates 1–18 and E1–E7.* (Huntington 1913, 528)

Huntington proves the independence of his axioms using several arithmetical and geometrical constructions. Some of his counterexamples are very original (if not bizarre): he considers e.g. spheres painted in blue and red or collections of egg-shaped convex solids. One of these counterexamples seems difficult to follow, at least for us:

> *Example 11.* To construct an example for Postulate 11, consider first an ordinary plane, with all its circles, points, and lines, and suppose the interior of a part of this plane – say a square $ABCD$ – is stretched or deformed in such a way that all the points within the square are crowded towards one corner C, without altering their relations of order, or causing any break in continuity. In this deformed plane, by a circle or line we mean, of course, a figure which was a true circle or line before the deformation. Secondly, consider another plane, containing a square $APCQ$ without any deformation, and place the two planes so that they intersect along the line AC.
>
> Then as our class K we take all the circles that lie in these two planes, and as our relation R, the ordinary relation of inclusion.

> In this system, Postulates 1–10 are clearly satisfied, but not Postulate 11. To see that Postulate 11 is not true, let PQ meet AC in M, while the deformed line BD meets AC in a different point N; then B, P, D, Q cannot be coplanar; but if Postulate 11 were true, we should have a right to infer, from a consideration of the 'four-points' $ACBP$ and $CADQ$, that $BP \parallel DQ$. All the other general laws, 12–18, are satisfied (many of them vacuously). (Huntington 1913, 551)

2.4 The work of Carnap and Bachmann

As far as we know, the term *extremal axiom* was introduced in Carnap and Bachmann 1936. Axioms of this sort were already known: the authors mention Peano's induction axiom in arithmetic, completeness axiom in Hilbert's system of geometry and Fraenkel's axiom of restriction in set theory. The paper Carnap and Bachmann 1936 (translated into English as Carnap and Bachmann 1981) is an attempt to characterize a general form of such axioms. At the beginning of the paper they write:

> Some important axiom systems are so constructed that first a series of axioms is given, making certain statements about the basic concepts of the axiomatic theory, and then at the end an axiom of a special sort appears which apparently speaks about the foregoing axioms and not about the special concepts of the theory. The most famous axiom system of this sort is Hilbert's axiom system of Euclidean Geometry. It ends with the famous "completeness axiom" which runs as follows [The footnote given here by the authors reads: D. Hilbert, *Grundlagen der Geometrie* (Leipzig and Berlin). We take the Hilbert completeness axiom in the form it has in editions 2–6, not the 'linear formulation' of the 7th edition of 1930.]:
>
> 'The elements (points, lines, planes) of geometry constitute a system of things which cannot be extended while maintaining simultaneously the cited axioms, i.e., it is not possible to add to this system of points, lines, and planes another system of things such that the system arising from this addition satisfies axioms AI-V1.'
>
> Axioms of this sort, which ascribe to the objects of an axiomatic theory a maximal property – in that they assert that there is no more comprehensive system of things that satisfies a given series of axioms – we call a maximal axiom. The same axiomatic role as that of maximal axiom is played in other axiom systems by minimal axioms which ascribe a minimality property to the objects

> of the discipline. Maximal and minimal axioms we call collectively extremal axioms. (Carnap and Bachmann 1981, 68–69)

Before discussing the content of this paper it is necessary to make some comments. The extremal axioms mentioned so far were formulated by mathematicians and this was done either in absence of a system of formal logic or in a situation where logical investigations were just emerging. This differs from the case of the paper Carnap and Bachmann 1936. The main difference is that the authors' proposals – motivated mathematically – are formulated with respect to an original approach to formal axiomatics and metalogic developed by Carnap just before the publication of the paper. There are several factors of importance here:

1. The authors work in type theory (a simple theory of types). This was the prevailing paradigm at that time.

2. The authors make an essential use of Carnap's ideas contained in his *Abriss der Logistik* (Carnap 1929) and his *Untersuchungen zur allgemeinen Axiomatik* – notes from 1928 that as a whole were unpublished at that time and edited as late as in 2000 (Carnap 2000). Some published results related to that work are Carnap's short notes: *Eigentliche und uneigentliche Begriffe* (Carnap 1927), *Bericht über Untersuchungen zur Allgemeinen Axiomatik* (Carnap 1930).

3. The meaning of the term "model" as proposed by Carnap differs essentially from the contemporary meaning of this term. There are also subtle questions concerning domain variation, definability of models in the theory of types, and range of quantification, and so on.

4. The authors relate investigations of extremal axioms with several metalogical notions proposed by Carnap at that time. In particular, this concerns the question as to how the extremal axioms are related to categoricity and different notions of completeness.

It seems that Carnap's metalogical proposals as well as his work on extremal axioms were underestimated until quite recently. There are numerous papers published in the last two decades which show the originality of Carnap's ideas, for example: Awodey and Carus 2001, Awodey and Reck 2002a, 2002b, the introductory notes to Carnap 2000 by Bonk and Mosterin, Schiemer 2010a, Schiemer 2010b, Schiemer 2012, Schiemer

2013, Schiemer and Reck 2013, and Schiemer, Zach and Reck 2015. Some of these works make an essential use of still unpublished materials from the Carnap's *Nachlass*. Our remarks below take into account the results reported in the papers mentioned above. We must point out that the presentation of these issues in Schiemer 2010b is much more elaborate than our own sketchy remarks.

Let us come back to the paper Carnap and Bachmann 1936. The authors work in the simple theory of types. This theory corresponds to that which can be expressed in the modern system of higher-order logic.

Axiom systems are understood as propositional functions from the language of the simple theory of types. The primitive terms are predicates of that language understood as free variables. An axiom system may contain individual and predicate variables. Thus an axiom system is a (finite) conjunction of formulas containing the primitive terms. By an *admissible model* of an axiom system Carnap understood the result of a substitution of predicate constants of the appropriate sort (arity) for predicate variables. A sequence of predicate constants is a *model* of an axiom system if the substitution of these constants in places of predicate variables gives *analytic truth*. The latter notion is to be understood as presented in Carnap's *Logische Syntax der Sprache* (Carnap 1934). Compare the following remark:

> As basic signs for which no definite meaning is presupposed, we use variable signs of the corresponding types. Axioms, and therefore also axiom systems viewed as conjunctions of axioms, are then propositional functions of these basic variables, therefore of the form "$F_1(M)$", where "F_1" is a predicate and "M" is an abbreviation for a finite sequence of variables. Accordingly we must also regard the theorems as propositional functions, and we call "$F_1(M)$" a theorem of the theory characterized as "$F_1(M)$", if the universal implication
>
> $$(M)(F_1(M) > F_2(M))$$
>
> is an analytic sentence of the language S. (Carnap and Bachmann 1981, 70)

Logical terms used by Carnap in *Untersuchungen zur allgemeinen Axiomatik* include: proper logical constants (quantifiers and logical connectives), constants of arithmetic (number, successor, numerals), constants of set theory (class, membership), and constants of relation theory.

Domains of models in Carnap's sense are understood as the unions of fields of the relations constituting the models. Since models are determined by (finite) sequences of relations, they themselves can be conceived of as relations of a higher type. The inclusion between models is defined in terms of the inclusion of their relations, and as such a model M is a submodel of a model N, if the relations in M are partial relations of the corresponding relations in N.

Besides models, Carnap and Bachmann consider *structures*. They are defined in terms of *complete isomorphisms*. The definition of the last concept is inductive and rather involved – see Carnap and Bachmann 1981, 73–74. The idea is simple: complete isomorphism is isomorphism on each level of the hierarchy of logical types. Elements of the domain of a model may be of any (finite) logical type. The complete isomorphisms are defined, so to speak, from top to bottom: from higher logical types to the lower ones. Let us denote the relation of complete isomorphism by $\cong_v$.

Now, structures are determined by the abstraction classes of the relation $\cong_v$ which obviously is an equivalence relation. Any structure is a higher-order relation. The authors say that a model M_1 *has* a structure S_1, if the sentence "$S_1(M_1)$" is analytic.

A structure S is *dividable* if every model with this structure contains a proper part which belongs to the same structure S. An example of a dividible structure is the ordering of type ω (like the ordering of natural numbers with the relation less-than). Structures which are not dividable are called *undividable*. For instance, any finite segment of the natural numbers ordered by the relation less-than has an undividable structure.

The propositional function $F(M)$ is called *structural*, if it is closed under the relation of complete isomorphism. In case of structural propositional functions it thus makes sense to say that a given structure is characteristic of the propositional function in question.

One can associate with each axiom system $F(M)$ its *structural diagram*. It represents the relation of *proper substructure, restricted to the structures belonging to "$F(M)$"*. We say that S is a proper substructure of T (in symbols: $S \sqsubset_F T$), if S is distinct from T and every model that has a structure S is isomorphic to a proper part of any model that has structure T. Elements belonging to the domain and not to the counterdomain of the relation $\sqsubset_F$ are called *beginning structures*, while those belonging to the counterdomain and not to the domain of the relation $\sqsubset_F$ are called *end structures*. By *isolated structures* one means structures belonging to "F" but not belonging to the field of $\sqsubset_F$. Beginning and isolated structures are called *minimal* while end and isolated structures are called *maximal*.

An informal characterization of extremal axioms is presented in the following formulation:

> We understand by a *maximal axiom belonging to a structural function "$F(M)$"* (putting it first informally) a propositional function: "There is no extension N of M such that $F(N)$ holds", and by a *minimal axiom belonging to "$F(M)$"* a propositional function: "There is no N such that $F(N)$ holds and M is an extension of N". (Carnap and Bachmann 1981, 76)

There are two types of extensions: those of models and of structures. N is a model extension of M, if M is a submodel of N and M and N are distinct. In turn, N is a structure extension of M, if M is a submodel of N and M and N are not completely isomorphic.

Finally, the authors formulate four kinds of extremal axioms which may belong to a propositional function "$F(M)$" taking into account these two types of extension:

1. Maximal model axiom: $Max_m(F, M)$ expresses the fact that there is no model extension N of M.

2. Maximal structure axiom: $Max_s(F, M)$ expresses the fact that there is no structure extension N of M.

3. Minimal model axiom: $Min_m(F, M)$ expresses the fact that there is no model N such that M is a model extension of N.

4. Minimal structure axiom: $Min_s(F, M)$ expresses the fact that there is no model N such that M is a structure extension of N.

Given an axiom system $F_1(M)$, there are thus four new axiom systems obtained by adding one of the extremal axioms to $F_1(M)$:

1. Conjunction of $F_1(M)$ and $Min_s(F, M)$. Here belong minimal structures of "F_1".

2. Conjunction of $F_1(M)$ and $Min_m(F, M)$. Here belong undividable minimal structures of "F_1".

3. Conjunction of $F_1(M)$ and $Max_s(F, M)$. Here belong maximal structures of "F_1".

4. Conjunction of $F_1(M)$ and $Max_m(F, M)$. Here belong undividable maximal structures of "F_1".

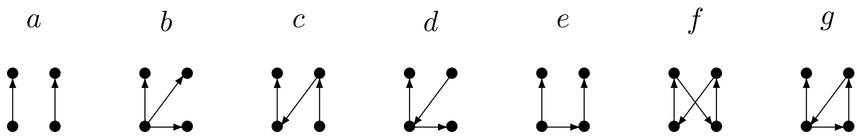
The authors claim that Hilbert's completeness axiom for Euclidean geometry (as well as his completeness axiom for the real numbers) can be formulated as a maximal model axiom while the Fraenkel's axiom of restriction can be formulated as a minimal model axiom.

The authors provide also a few examples constructed on purpose, illustrating the concepts introduced above:

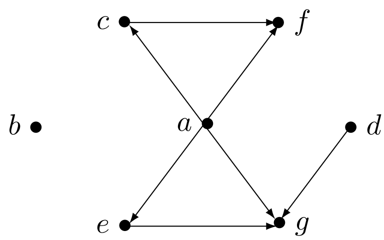
Example 1. Consider the following axioms $(*)$ for a binary relation R:

1. $(x)(y)(z)\ ((R(y, x) \wedge R(z, x)) \rightarrow y = z)$
2. $(x)(y)\ (R(x, y) \rightarrow \neg R(y, x))$
3. $(x)\ \neg R(x, x)$
4. The field of R has exactly four elements.

There are exactly seven structures of these axioms whose universe is a four element set. They may be portrayed as in the following diagram:



Some connections between them can be presented on their *structural diagram*, where an arrow corresponds to the relation $\sqsubset$ of being a proper substructure:



Here a, b and d are minimal while b, f and g are maximal structures. This means that if we add a minimal structure axiom to $(*)$, then we get a, b and d as structures satisfying such an expanded axiom system. Similarly, a maximal structure axiom added to $(*)$ gives b, f and g as structures satisfying such an expanded axiom system.

EXAMPLE 2. Consider the following axioms (★) for a system called elementary arithmetic by the authors:

1. $(x)(y)\ (R(x, y) \rightarrow (\exists z)\ R(y, z))$ (R is endless),
2. $(x)(y)(z)\ (((R(x, y) \wedge R(x, z)) \rightarrow y = z) \wedge ((R(x, y) \wedge R(z, y)) \rightarrow x = z))$ (R is one-one),
3. R has exactly one initial element (i.e. the difference of domain and counterdomain of R has exactly one element),
4. Minimal structure axiom for the above axioms.

The authors claim that models of 1–3 (more exactly: structures satisfying 1–3) are: ω (order type of natural numbers) and ω together with any number (from zero to infinitely many) of R-cycles with from one to infinitely many elements (R-cycle with infinitely many elements is understood as the order type $\omega^* + \omega$ of the integers). ω is a minimal-structure model of (★).

Carnap and Bachmann do not consider algebraic examples though it seems that there are interesting cases which could be taken into account. Let us recall that one defines a *real closed field* as an ordered field where the order relation is a maximal one compatible with the operations of the field. This is a kind of maximality condition. In this case we have several alternative (equivalent to the above) characterizations of real closed fields. For instance, a real closed field is an ordered field in which every positive element has the root and every polynomial of odd degree has a root (the latter requires an infinite set of conditions in the first-order language). Thus, the maximality condition (concerning order) can be replaced by an equivalent one without any explicit reference to models of the theory and hence without any necessity of quantifying over the space of models. We will come back to this issue in chapter 5.

The authors discuss connections between extremal axioms and categoricity (monomorphy): on page 82 of Carnap and Bachmann 1981 they write that monomorphy implies that the models in question cannot be further extended (in such a way that all axioms remain valid) but the converse implication does not hold, and they recall in this respect the result of Baldus from Baldus 1928 (cf. chapter 5).

Carnap devoted much attention to relationships between several notions of completeness, essentially developing the ideas informally described by Fraenkel. In *Bericht über Untersuchungen zur Allgemeinen Axiomatik*, published in the first issue of *Erkenntnis* in 1930 (Carnap 1930) as well as in *Untersuchungen zur allgemeinen Axiomatik* (Carnap 2000) he gives definitions of the following metalogical properties:

1. Categoricity (in his terms: *Monomorphie*). A system of axioms is *monomorph* if it has only one structure (see the definition of structure above), that is, if any two models of it are completely isomorphic. Systems which are not monomorph are called *polymorph.*

2. Semantic completeness (in his terms: *Nicht-Gabelbarkeit*, which means non-forkability). An axiom system $F(\mathcal{R})$ is *forkable* at a (structural) propositional function $G(\mathcal{R})$ if there exists a model for the conjunction of $F(\mathcal{R})$ and $G(\mathcal{R})$ as well as a model for the conjunction of $F(\mathcal{R})$ and the negation of $G(\mathcal{R})$. Systems which are not forkable for any propositional function (of an appropriate kind) are called *non-forkable.*

3. Decidability (in his terms: *Entscheidungsdefinitheit*). A satisfiable axiom system $F(\mathcal{R})$ is *decidable* if for any (structural) propositional function $G(\mathcal{R})$ either $G(\mathcal{R})$ or the negation of $G(\mathcal{R})$ is a theorem of $F(\mathcal{R})$. Due to the assumptions adopted by Carnap in his approach to axiom systems this property coincides, however, with non-forkability.

4. Constructive decidability (in his terms: *k-Entscheidungsdefinitheit*). An axiom system $F(\mathcal{R})$ is *k-decidable*, if there is a procedure which can give in a finite number of steps for any (structural) propositional function $G(\mathcal{R})$ either the proof of the implication $F \rightarrow G$ or the implication $F \rightarrow \neg G$.

The works Carnap 1930 and Carnap 2000 contain his *Gabelbarkeitssatz*, a claim that the properties of polymorphy and forkability are equivalent. This attempt was flawed. There are several factors of this failure, discussed for example in Awodey and Carus 2001:

1. lack of a sharp distinction between object language and a metalanguage
2. lack of a sharp distinction between syntax and semantics
3. the unjustified assumption that any consistent theory has a definable model in the simple theory of types.

That project, though unsuccessful was nevertheless important in the history of logic. Kurt Gödel was one of a few logicians who had read Carnap's manuscript, just before he proved his famous incompleteness results. Let us add that a correct proof of a very similar theorem was given in Lindenbaum and Tarski 1936 – there the authors provide criteria (the existence of a model definable in type theory) under which non-forkability implies monomorphy. Tarski himself also discussed categoricity and completeness in Tarski 1940. More recently such relationships have been investigated in model theory. The works George 2006, Weaver and George 2002, 2003, 2005 are devoted to problems of definability and the so-called Fraenkel-Carnap property, which concerns (roughly speaking) the conditions under which semantic completeness implies categoricity.

Carnap himself returned to extremal axioms in his later works – for instance in Carnap 1954 (translated as Carnap 1958) on page 154 he proposes a formulation of Fraenkel's axiom of restriction which says, roughly speaking, that each subrelation of the relation of membership which satisfies the axioms of set theory is identical to the relation of membership itself.

2.5 Further developments

Let us finish this chapter with brief indications of the further development of the issues under consideration. Clear characterizations of such fundamental notions as categoricity, categoricity in power, several forms of completeness, decidability, and computability were obtained during several decades in the 20th century. Limitative theorems revealed the possibilities and restrictions of first-order logic. In particular, it became clear that some methodological ideals can not be obtained simultaneously. The descriptive power of a system of logic (and the underlying formal language), that is its ability to characterize models categorically is, in a sense, inversely proportional to its deductive power. This topic will be described in some detail in the next chapter. Investigations concerning categoricity and completeness were further developed in the classical model theory (as initiated by Tarski, Vaught and others) as well as later in the modern model theory (whose beginning may be connected with the Morley categoricity theorem). This topic, in turn, will be described in some detail in the fourth chapter of this book.

Chapter 3

The expressive power of logic and limitative theorems

We recall here some important facts from the history of mathematical logic. We also discuss the difference between the expressive and deductive power of logic. We pay attention to the role of limitative theorems in the foundations of mathematics which show the possibilities and limitations of the deductive method. A comparison of logical systems can be presented in a semantic framework of abstract logics.

3.1 Expressive versus deductive power of logic

We should begin with a few words concerning the history of mathematical logic and the foundations of mathematics. However, we will limit ourselves to a brief indication of the events relevant to our main subject:

1. Mathematical logic began in the 19th century with the works of George Boole, Augustus De Morgan, Gottlob Frege, Ernst Schröder, Charles Sanders Peirce and others.

2. Hermann Grassmann proposed an axiom system for the integers in Grassmann 1861. Richard Dedekind proposed a formal characterization of natural numbers in Dedekind 1888. Giuseppe Peano proposed an axiom system for natural numbers in Peano 1889. Richard Dedekind proposed constructions of the real numbers in Dedekind 1872 (at that time several other authors also proposed formal description of real numbers).

3. David Hilbert proposed a system of axioms for Euclidean geometry in Hilbert 1899 and an axiom system for the real numbers in Hilbert 1900.

4. The *Principia Mathematica* (1910–1913) written by Bertrand Russell and Alfred North Whitehead still assumes that there is only one system of logic. The formalism used in the book is that of type theory.

5. Ernst Zermelo proposed the first axiomatization of set theory in Zermelo 1908.

6. The theorem proved in 1915 by Leopold Löwenheim and later improved by Thoralf Skolem (Löwenheim 1915, Skolem 1920) is the first result which may be called metalogical. In modern terms the theorem says that if a (first-order) theory has at least one (infinite) model, then it has a countable model.

7. Thoralf Skolem, Wilhelm Ackermann, Abraham Fraenkel, John von Neumann contributed to the axiomatic foundations of set theory (second and third decades of the 20th century).

8. Rudolf Carnap contributed to the emergence of metalogical investigations (third and fourth decades of the 20th century).

9. Alfred Tarski developed a general approach to metalogic (third and fourth decades of the 20th century).

10. David Hilbert and Wilhelm Ackermann proposed the standard of first-order logic in their *Grundzüge der theoretischen Logik* in 1928 (Hilbert and Ackermann 1928).

11. Ernst Zermelo proposed the second axiomatization of set theory in Zermelo 1930.

12. Kurt Gödel proved his famous theorems: completeness of the system of first-order logic and incompleteness of the system of first-order arithmetic (Gödel 1930, 1931).

13. Thoralf Skolem constructed a non-standard model of arithmetic using a construction resembling ultraproduct (Skolem 1933, 1934). Anatoly Maltsev used the compactness property to show the existence of a non-standard model of arithmetic in 1936.

14. David Hilbert and Paul Bernays wrote a comprehensive monograph *Grundlagen der Mathematics* (Hilbert and Bernays 1934/1939).

15. Kurt Gödel introduced the constructible universe and proved the consistency of the axiom of choice and the continuum hypothesis with respect to the axioms of Zermelo-Fraenkel set theory (Gödel 1940).

16. Andrzej Mostowski introduced generalized quantifiers in Mostowski 1957 and several other logicians (Alfred Tarski, Carol Karp, and M.A. Dickmann) contributed to the foundations of infinitary logic at that time.

17. Alfred Tarski, Abraham Robinson, Robert Vaught and others developed classical model theory (the beginning of the second half of the 20th century).

18. Paul Cohen introduced the method of forcing and proved the independence of the axiom of choice and the continuum hypothesis from the axioms of Zermelo-Fraenkel set theory in Cohen 1966.

There are many more results which could be added to this list as far as our main subject (attempts towards the unique characterization of intended models) is concerned. Some of them will be mentioned in chapters 4–7 of this book.

Different systems of logic can be compared with respect to their properties and to their strength, among other things. For our purposes the following distinction will be useful:

1. *Expressive power.* This property is related to the possibilities of characterization of mathematical structures in terms of syntactic and semantic tools accessible in the language under consideration. We are interested in which mathematical concepts are definable in the system of logic under consideration. For instance, the concepts of continuity and infinity are not definable in first-order logic, but they are definable in second-order logic.

2. *Deductive power.* This property embraces the deductive possibilities which are possessed by a given system of logic. Several factors are taken into account in this respect, for instance we may demand that logic should be sound and complete (that is, that their theorems coincide with their tautologies). Or we may restrict our

attention to finitary inferences only, thus demanding that the logic in question is based on a finitary operator of consequence and satisfies the compactness theorem.

The above two metalogical properties are, in a sense, inversely proportional. Logic with great expressive power is usually poor with respect to its deductive power and vice versa, if a logic has many nice deductive properties, then it usually lacks the possibilities of a unique characterization of important mathematical notions.

One should perhaps stress the fact that systems of logic are evaluated differently from the point of view of logicians and that of mathematicians. Logicians are primarily interested in deductive aspects of systems of logic, while mathematicians seem to put more stress on their expressive power (obviously not forgetting the deductive properties). Let us quote the opinion of Barwise in this respect:

> But if you think of logic as the mathematicians in the street, then the logic in a given concept is what it is, and if there is no set of rules which generate all the valid sentences, well, that is just a fact about the complexity of the concept that has to be lived with. (Barwise 1985, 7)

In this book we are interested mainly in the opinions of mathematicians concerning the possibilities of a unique characterization of the intended models of their theories.

3.2 Metalogic and metamathematics

We use the terms *metalogic* and *metamathematics* in the standard, widely accepted sense. Metalogic is a theoretical reflection about logic itself. This reflection is articulated in mathematical terms. Similarly, metamathematics is a theoretical reflection about mathematical theories and is also a mathematical discipline. Philosophical analysis concerning logic and mathematics may be treated as a companion to metalogic and metamathematics. It is commonly believed that the term *metamathematics* was introduced by David Hilbert. However, Paul Du Bois-Reymond also used this term (roughly in the contemporary meaning) in Du Bois-Reymond 1866.

3.2.1 Examples of metalogical properties

Some of the most important metalogical properties of systems of logic and theories formulated in their languages are the following (some of of which have already been discussed earlier in this book):

1. *Soundness (of a system of logic).* If any formula which is provable in logic L is a tautology of L, then we say that L is sound (with respect to the methods of proof accepted in L).

2. *Completeness (of a system of logic).* If any tautology of the logic L is provable in L, then we say that L is complete.

3. *Consistency (of a theory).* A theory T is consistent if there is no formula φ of the language of T such that both φ and $\neg\varphi$ are provable in T.

4. *Satisfiability (of a set of formulas).* A set of formulas (of the language of first-order logic) is satisfiable if it has a model.

5. *Compactness (of a system of logic).* We say that a logic L is (syntactically) compact if it is based on a finitary consequence relation. If a logic L has this property, then the following conditions are equivalent, for any set X of formulas in the language of L:

   (a) X is consistent.

   (b) Each finite subset of X is consistent.

   We may also formulate the compactness property in semantic terms. A logic L is (semantically) compact, if for any set X of formulas in the language of L the following conditions are equivalent:

   (a) X is satisfiable.

   (b) Each finite subset of X is satisfiable.

   In both formulations the implication from (a) to (b) is trivial – it is the implication from (b) to (a) which expresses the essence of compactness.

6. *Categoricity.* A theory is categorical if all its models are isomorphic.

7. *Categoricity in power.* A theory is categorical in an infinite power κ if it has a model of power κ and all its models of power κ are isomorphic.

8. *Completeness (of a theory).* Let us recall the definitions from the previous chapter:

   (a) *Deductive (syntactic) completeness.* A theory T is complete in this sense if for all sentences φ, either $T \vdash \varphi$ or $T \vdash \neg\varphi$. Here $\vdash$ denotes syntactic consequence relation.

   (b) *Semantic completeness.* A theory T is semantically complete (with respect to a given semantics) if for all sentences φ and all models M, N of T, if $M \models \varphi$, then $N \models \varphi$. It follows from this definition that a theory is complete if all its models are elementarily equivalent.

9. *ω-consistency.* Let T be a (first-order) theory which interprets arithmetic, in a sense that there is a translation of the language of arithmetic into the language of T. This means in particular, that we have terms $\overline{n}$ in the language of T which are numerals corresponding to natural numbers. We say that T is ω-inconsistent if for some formula $\varphi(x)$ of the language of T with one free variable T proves all sentences $\varphi(\overline{n})$ as well as the sentence $\exists x \neg\varphi(x)$. If T is not ω-inconsistent, then we say that T is ω-consistent. The property of ω-consistency is stronger then consistency.

10. *Decidability.* A theory T is decidable if the set of (Gödel numbers of) all its theorems is a recursive set.

11. *Recursive axiomatizability.* A theory T is recursively axiomatizable if its axioms (more precisely, numbers coding these axioms, such as Gödel numbers) form a recursive set.

The list of metalogical properties investigated recently is of course much longer. The reader may consult any modern textbook to get more information on this issue, for instance: Cori and Lascar 2001 or Hinman 2005.

3.2.2 Origin of metalogic

According to *The Search for Mathematical Roots, 1870–1940* written by Igor Grattan-Guiness, the term *metalogic* was used by Itelson during the second International Philosophical Congress in Geneva (Grattan-Guiness 2000, 562, citing Couturat 1904, 1041); however Itelson did not understand this term in the modern sense as he thought of logic as a universal and unique system. And according to the same author, Gerhard Stammler also used this term in a similar sense (as Itelson) in his *Begriff Urteil Schluss. Untersuchungen über Grundlagen und Aufbau der Logik* (Stammler 1928), where on page 317 he writes:

> There is no metalogic as extralogical grounding of logic. Logic stands for itself. (Citing Grattan-Guiness 2000, 489)

William Thomson in 1842 (*Outline of the laws of thought*, page 23) used the term *metalogic* in still another meaning, as: *a science regulating the processes and symptoms of thought which are not universal* (Thomson 1842, 23, citing Grattan-Guiness 2000, 562).

Rudolf Carnap in the notes *Versuch einer Metalogik* (Carnap 1931) used the term *metalogic* in its modern sense. His notes were known to a few of his contemporaries, notably to Kurt Gödel. Carnap's typescript *Untersuchungen zur Allgemeine Axiomatik* from 1928 is stored at the Archives for Scientific Philosophy, University of Pittsburgh. The greater part of these notes was published as Carnap 2000. Many metalogical notions and results originated from an Alfred Tarski seminar in Warsaw in the second decade of the twentieth century.

The origin of metalogic became possible only after rejection of certain beliefs concerning logic itself, notably the belief that there was only one logic outside of which one could not go. This universality assumption was present in the works of Gottlob Frege, Alfred North Whitehead and Bertrand Russell. Also the algebraic tradition, related to the works of Charles Sanders Pierce, Ernst Schröder and Thoralf Skolem had some limitations – for instance it did not make a precise distinction between syntax and semantics.

Let us recall from the previous chapter that the papers Awodey and Reck 2002a, 2002b, Awodey and Butz 2000, and Awodey and Carus 2001 contain an interesting analysis of the origin and development of such metalogical notions as categoricity and completeness.

The first of these articles discusses different aspects of completeness, for example completeness of the (syntactic) relation of provability $\vdash$ with respect to the (semantic) relation $\models$ of satisfiability, or the semantic completeness of theories. The authors discuss the axiomatic systems of: Dedekind and Peano (concerning natural numbers), Hilbert (geometry with special emphasis on the *Vollständigkeit Axiom*), Hilbert and Dedekind (real numbers), Huntington (positive real numbers), and Veblen (Euclidean and projective geometry). Awodey and Reck stress the fact that between 1910 and 1930 such logicians as Hilbert, Gödel, Carnap, and Tarski did not work in the system of first-order logic but rather in (simple) type theory which, in a sense, is a logic of higher order.

The authors appreciate the main role played by Hilbert and Tarski in the development of metalogic. They add, however, that the achievements of Fraenkel and Carnap in this respect seem to be underestimated. In particular, Fraenkel included in his *Einleitung in die Mengenlehre* from 1928 many interesting remarks concerning a possible understanding of the notion of completeness.

Awodey and Reck advocate the system of second-order logic as adequate for mathematical research. In particular, in Awodey and Reck 2002b and Awodey and Butz 2000 one can find new concepts of completeness and categoricity, based on topological semantics.

According to Zach 1999, the (only partly published) Habilitationschrift of Paul Bernays *Beiträge zur axiomatischen Behandlung des Logik-Kalküls* from 1918 is the first place where formulas (of propositional calculus) were separated from their interpretations (valuations) and the completeness theorem of the system in question was formulated as a metatheorem.

The problem of completeness of classical predicate calculus was explicitly formulated for the first time in *Grundzüge der theoretischen Logik* from 1928, written by David Hilbert and Wilhelm Ackermann (Hilbert and Ackermann 1928). The problem was solved by Kurt Gödel in his doctoral dissertation from 1930. The circumstances of this result are described in the introductory note Dreben and van Heijenoort 1986. Gödel made some use of the techniques developed earlier by Thoralf Skolem but Skolem himself did not explicitly comment on the problem of completeness.

In the third decade of the 20th century metalogical investigations were conducted with growing intensity, due to the famous incompleteness results of Kurt Gödel from 1931 and ideas developed by Alfred Tarski (cf. Tarski 1933, 1934). The history of that research is well described in literature and we believe that there is no need to comment on it here. We recommend the collection of papers edited by Stanisław Surma (Surma 1973a) devoted to the history of the completeness problem in logic (in particular Surma 1973b, 1973c).

3.2.3 Paradigms of metalogic

Metalogical investigations are conducted in several styles. We think that the following paradigms can be recalled in this respect:

1. *Algebraic paradigm.* The central concept is that of a general consequence operator. Logic is understood as determined by such an operator (rather than by its set of theorems). Theories are fixed points of this operator. The set of all consequence operators (in a given language) forms a lattice. This paradigm was initiated by Alfred Tarski and developed later by Polish logicians (Jerzy Łoś, Roman Suszko, Witold Pogorzelski, Ryszard Wójcicki, Andrzej Wroński, Janusz Czelakowski, Jan Zygmunt, Jerzy Perzanowski, Piotr Wojtylak, Wojciech Dzik and others).

2. *Standard paradigm.* Under this label we cover the classical metalogical investigations, including first of all metatheorems concerning soundness, completeness, compactness, consistency, and independence of axioms, etc. The famous limitative theorems concerning incompleteness phenomena should of course also be included here, as well as some classical results from model theory, proof theory and recursion theory.

3. *Model-theoretic paradigm.* Also known as *soft model theory* or *abstract model theory*, this is a semantic approach to metalogic. It concerns a relation $\models$ between sentences and structures (both in a given signature). In a sense, sentences can then be identified with classes of structures (in which these sentences are true). In this approach the expressive power of different logics can be compared in a simple manner. Some examples will be given at the end of this chapter. Barwise and Feferman 1985 is a comprehensive presentation of this approach to metalogic.

Metalogical investigations are also classified in the well-know partition of research areas in mathematical logic and the foundations of mathematics:

1. *Model theory.* The classical model theory is sometimes considered a composition of universal algebra and logic. It is devoted to the investigations of theories and their models. Some special kinds of models (prime, atomic, saturated, homogeneous, and universal, among others) are distinguished. It proves several preservation

theorems as well as theorems concerning categoricity in power and completeness. Modern model theory focuses its attention on the problems of definability, spectra of theories and investigation of the space of types.

2. *Recursion theory.* This is devoted to several mathematical representations of the (intuitive) notion of computability. There are many, mutually equivalent, such representations, including recursive functions, Turing machines, Post systems, Church λ-calculus, and Markov algorithms. Higher recursion theory deals, among others, with the degrees of undecidability.

3. *Proof theory.* This theory was originated by David Hilbert and developed by many authors, notably Gerhard Gentzen. It is concerned with different proof techniques (axiomatic method, resolution, tableaux systems, natural deduction, and sequent calculi, among others). The more advanced part of proof theory deals for example with the consistency strength of theories.

4. *Set theory.* Most mathematicians believe that one can base all of mathematics on set theory. This theory has an elementary part, devoted to most fundamental constructions: operations on sets, properties of relations and functions, ordinal and cardinal numbers and their arithmetic. More advanced branches of set theory deal with its models, large cardinal numbers, partition theorems, and so forth. Descriptive set theory is devoted to the investigation of special kinds of sets in Polish spaces.

The monumental work Barwise 1977 is already seen as a classic exposition of these research areas.

3.2.4 Remarks on the concept of compactness

We are interested mainly in the notions of completeness and categoricity, but these notions are related to some other metalogical notions as well. Let us say a few words about the notion of compactness – its origin in logic and its connections with some purely mathematical notions. Our main sources for these remarks are the papers: Dawson 1993 and Buldt 2002. We are interested in the semantic version of the compactness property: a set of formulas is satisfiable if and only if each of its finite subsets is satisfiable.

Let us first recall some historical facts concerning the compactness theorem:

1. The first proof of the compactness theorem (for countable languages) is contained in the doctoral dissertation of Gödel from 1929 (published in 1930). Gödel uses normal forms of formulas (introduced by Skolem) and explicitly distinguishes between syntactic and semantic properties (unlike Skolem). Gödel formulates his result as a lemma in the proof of the completeness theorem.

2. In the years 1936–1941 Anatoly Maltsev proved the compactness theorem also for uncountable languages. His "general local theorem" is obtained by using purely algebraic tools.

3. The first explicit formulation of the compactness theorem (for languages of arbitrary cardinality) can be found in the doctoral dissertation of Leon Henkin from 1947. In Henkin 1996 the author described the context of discovery of his completeness theorem.

4. Leon Henkin, Abraham Robinson, Alfred Tarski and Anatoly Maltsev investigated the compactness property in the years 1946–1956 in the context of characterization of several algebraic structures.

5. Tarski initiated the topological perspective of the investigation of the compactness property in logic and Andrzej Mostowski stressed the existence of connections between topology and metamathematics as early as in the third decade of the 20th century.

Dawson points to the following differences between investigations of compactness in topology and logic:

1. *top-down fashion* in topology: a known infinite cover of a given space by open sets is (nonconstructively) reduced to an unknown finite cover;

2. *bottom-up fashion* in logic: satisfiability of an infinite set of sentences (which is at first sight not evident) is reduced to satisfiability of each of its finite subsets (which, as a rule, is clearly visible).

Connections between topology and logic as far as the compactness property is concerned are related both to the interpretation of the compactness theorem in topology and its proof. Dawson's article recalls several mathematical notions and results which are related to the compactness property, for example the Heine-Borel theorem, a generalized concept of a limit, notions of a filter and ultrafilter, connections between the theory of Boolean algebras and topology, and Tychonoff's theorem.

The topological perspective in logic may be illustrated by such results as: the topological proofs of completeness theorem for propositional and first-order logic, the Löwenheim-Skolem theorem, compactness theorem, and proof of equivalence of the compactness theorem with the Boolean prime ideal theorem.

In the late nineteen-fifties Alfred Tarski and his collaborators (Morel, Scott, Frayne) used the ultraproduct construction (owed essentially to Jerzy Łoś) in their investigations in model theory and obtained, among others, a proof of the compactness theorem. The role of the notion of ultraproduct in logic is described for example in Zygmunt 1973.

Argument based on compactness is used by way of example in the proof that if a (first-order) theory has finite models of arbitrary cardinality, then it also has an infinite model.

It should be noticed that of special importance are the results which provide mathematical (for example algebraic or topological) representations of logical concepts, in particular of semantical concepts. If one can provide a mathematical representation of a semantical concept, then the latter becomes more precise.

The compactness theorem does not hold, as a rule, in logics stronger than first-order logic. Generalizations of the compactness property are also considered in the so-called *admissible fragments.*

The compactness theorem also has interesting applications in non-standard analysis (cf. Robinson 1996). We will come back to this subject in chapter 5.

3.3 Limitative theorems

By limitative theorems one understands such theorems which reveal certain restrictions as far as the methods or methodological ideals are concerned. They express for instance the distinction between proof and truth, the non-definability of certain concepts, incompleteness phenomena, and undecidability, among others. The results obtained by Gödel, Tarski, and Turing are standard examples of limitative theorems. But the Löwenheim-Skolem theorem already has this character, too: it expresses the fact that first-order logic does not distinguish infinite powers. Limitative character is also visible in the Lindström theorems which characterize first-order logic as a maximal logic with certain given metalogical properties.

3.3.1 Examples of limitative theorems

The most famous limitative theorems are the following:

1. *Downward Löwenheim-Skolem theorem.* If a theory (in a first-order language) has at least one infinite model, then it has a countable model.
2. *Upward Löwenheim-Skolem-Tarski theorem.* A consistent theory with no finite models has models of arbitrary high cardinality.
3. *I Gödel theorem.* If the Peano arithmetic is ω-consistent, then it is not complete. There exist undecidable arithmetical statements.
4. *Rosser theorem.* If the Peano arithmetic is consistent, then it is not complete. There exist undecidable arithmetical statements.
5. *II Gödel theorem.* If the Peano arithmetic is consistent, then its consistency cannot be proved in it.
6. *Tarski theorem.* The set of (Gödel numbers of) all sentences of the language of Peano arithmetic true in the standard model of arithmetic is not definable in the language of this theory.
7. *Turing theorem.* The halting problem is undecidable.
8. *Church theorem.* The predicate calculus is not decidable.
9. *Trakhtenbrot theorem.* There exists a finite signature such that the set of all sentences of the first-order logic of this signature which are true in all finite structures belonging to the class of structures of this signature is not recursively enumerable.
10. *Tennenbaum theorem.* The standard model of Peano arithmetic is its only recursive model.
11. *I incompleteness theorem for set theory ZF.* For each theory T with $\in$ as its only non-logical constant such that $T \cup ZF$ is consistent the following hold:

    (a) T is undecidable.

(b) If T is recursively axiomatizable, then T is incomplete.

12. *II incompleteness theorem for set theory ZF*. Let $Con(T)$ be a sentence expressing the fact that T is consistent. Then for every consistent and recursively axiomatizable theory T such that $ZF \subseteq T$, the sentence $Con(T)$ is not provable in T. In particular, $Con(ZF)$ is not provable in ZF.

There is a huge amount of literature concerning the limitative theorems. We are not going to comment on them here, assuming that the reader is well acquainted with their presentation in textbooks.

3.3.2 Noncompossibility theorem

The article Tennant 2000 concerns the expressive and deductive power of logic in a historical perspective. Its main result shows that certain methodological ideals (as for instance completeness and categoricity) cannot be achieved simultaneously. Tennant focuses his attention on monomathematics – a term introduced by him, referring to those mathematical theories which are thought of as characterizing intended models. We have discussed the distinction between monomathematics and polymathematics in the first chapter of this book.

The methodology of monomathematics has, according to Tennant, some restrictions. The axioms accepted for characterization of the intended model should form a computable (recursive) set and should express our well-grounded intuitions about the model in question. The proofs should be finite syntactic objects. The same applies to refutations. Here by a refutation of a set of sentences Δ one means a proof of contradiction from Δ. Both proofs and refutations should be sound, in the sense that if there exists a proof of φ from Δ, then φ is true in all structures in which all the sentences from Δ are true. This also means that if there exists a refutation of Δ, then there is no model in which all the sentences from Δ are true.

Let $\mathcal{A}$ be a recursive set of axioms formulated in the language L appropriate for talking about the intended model M of all axioms from $\mathcal{A}$. We assume that L contains the identity predicate $=$ and by a $=$-literal we understand each formula which is either an identity or a negation of an identity. Tennant defines the following version of completeness:

Weak completeness. For each expansion L^* of L by a finite amount of non-logical constants there exists a system of sound refutations such that for any recursive and satisfiable set X of =-literals from L^*: if $\mathcal{A} \cup X$ does not have a model, then there exists a refutation of a finite subset of the set $\mathcal{A} \cup X$.

This condition is weaker than compactness and weaker than strong completeness of logical consequence. Under certain natural assumptions it is equivalent (on the basis of RCA_0, meaning the system of *recursive comprehension arithmetics*) with recursive enumerability of the set of valid sentences. Tennant considers the following categoricity condition:

Categoricity. Each structure which is a model of all axioms from $\mathcal{A}$ is isomorphic to the intended model M.

The above condition does not say anything about the number of isomorphisms under consideration. Tennant proves the following theorem:

The Noncompossibility Theorem. If M is an infinite countable structure in which each element is definable (in L^*), then there does not exist a recursive set $\mathcal{A}$ satisfying the condition of weak completeness and such that $M \models \mathcal{A}$ which categorically describes M.

The idea of the proof is rather simple. One considers the set:

$$X = \mathcal{A} \cup \{\neg a = t_m : m \in M\},$$

where a is a new constant and each t_m is a term denoting element $m \in M$. The assumption that X does not have a model implies that some of its finite subsets Z do not have a model. As Z is finite and M is countable, there exists an element in M which is not denoted by any term occurring in Z. We expand the model M by assuming that a denotes this very element and hence we obtain a model for Z which contradicts our initial assumption. Thus the set X has a model, but this model cannot be isomorphic with M, because the denotation of a in it is different from all denotations of the terms t_m. This finishes the proof of the theorem.

Tennant's result thus has evident consequences for the role of extremal axioms which were proposed in order to characterize intended models. Tennant writes that his theorem holds for first-order and second-order logic. Observe that the theorem does not have the flavor of Skolem's paradox which is related to the fact that in first-order logic it is not possible to characterize categorically uncountable structures. Due to the Noncomposibility Theorem it is not possible to achieve simultaneously:

1. a categorical (that is, structurally unique) characterization of the intended model
2. a well-established, sound deductive machinery which could offer us "the full truth" about the intended model.

3.4 Abstract logics and Lindström's theorems

Theorems obtained by Lindström in the nineteen-sixties are also considered as limitative theorems. They are formulated in the model theoretic paradigm of metalogic. Here we limit ourselves to a brief account of the main results. For a more detailed exposition, see for instance: Barwise and Feferman 1985, Ebbinghaus, Flum and Thomas 1996, Krynicki, Mostowski and Szczerba (eds.) 1995, Shapiro (ed.) 1996, Stegmüller and Varga von Kibéd 1984, and Westerståhl 1989. Lindström's original works are: Lindström 1966a, 1966b, 1966c, 1969.

This approach sees the explicit characterization of the notion of the expressive power of logic. Firstly, an abstract logic is taken as meaning a system consisting of a language and a relation which holds between sentences of that language and relational structures. This relation is the relation of satisfiability characterized axiomatically. The intuition behind it is the following: sentences may be treated as classes of relational structures (meaning those structures in which the sentences are true). Secondly, we may say that a logic $\mathcal{L}_1$ has expressive power less than or equal to that of a logic $\mathcal{L}_2$, if for every sentence ψ_1 of a logic $\mathcal{L}_1$ there exists a sentence ψ_2 of a logic $\mathcal{L}_2$ such that ψ_1 and ψ_2 have exactly the same models. This relation is a partial ordering and thus determines in a known way an equivalence relation (of having the same expressive power) and a strict ordering (of having less expressive power).

Abstract logics are systems of the form $\mathcal{L} = (L, \models^{\mathcal{L}})$ which satisfy certain natural conditions characterizing L and $\models^{\mathcal{L}}$ and concerning signatures, isomorphic structures and reducts of structures, which we omit here. Let $L(\sigma)$ be the language of signature σ, $Str(\sigma)$ be the class of all structures of the signature σ and let for any sentence φ and any set of sentences Φ of that language:

$$Mod^{\sigma}_{\mathcal{L}}(\varphi) = \{\mathfrak{A} : \mathfrak{A} \in Str(\sigma) \wedge \mathfrak{A} \models^{\mathcal{L}} \varphi\}$$

$$Mod^{\sigma}_{\mathcal{L}}(\Phi) = \bigcap\{Mod^{\sigma}_{\mathcal{L}}(\psi) : \psi \in \Phi\}.$$

If σ is known from the context, then we may omit it. We can now define the following relations between $\mathcal{L}_1 = (L_1, \models^{\mathcal{L}_1})$ and $\mathcal{L}_2 = (L_2, \models^{\mathcal{L}_2})$:

1. We write $\mathcal{L}_1 \leqslant \mathcal{L}_2$ if and only if for each signature σ and every $\varphi \in L_1(\sigma)$ there exists $\psi \in L_2(\sigma)$ such that: $Mod^{\sigma}_{\mathcal{L}_1}(\varphi) = Mod^{\sigma}_{\mathcal{L}_2}(\psi)$. We say in this case that $\mathcal{L}_1$ has *expressive power less or equal to that of* $\mathcal{L}_2$.

2. If $\mathcal{L}_1 \leqslant \mathcal{L}_2$, but $\mathcal{L}_2 \leqslant \mathcal{L}_1$ does not hold, then we write $\mathcal{L}_1 < \mathcal{L}_2$ and say that $\mathcal{L}_2$ has *greater expressive power than* $\mathcal{L}_1$.

3. If $\mathcal{L}_1 \leqslant \mathcal{L}_2$ and $\mathcal{L}_2 \leqslant \mathcal{L}_1$, then we write $\mathcal{L}_1 \sim \mathcal{L}_2$ and say that $\mathcal{L}_1$ and $\mathcal{L}_1$ have *the same expressive power.*

In order to formulate the Lindström theorems one should introduce several further notions (regularity conditions, relativization condition, conditions concerning the Boolean structure of the considered languages, quantifier rank of a sentence, and so on) as well as some facts about families of partial isomorphisms and definability, among other things. We omit these details here. Let $\mathcal{L}_{\omega\omega}$ denote the classical first order logic. We introduce only the following notions:

1. a sentence $\varphi \in L(\sigma)$ is called *satisfiable* in $\mathcal{L}$ if $Mod^{\sigma}_{\mathcal{L}}(\varphi) \neq \emptyset$;

2. a set $\Phi \subseteq L(\sigma)$ is called *satisfiable* in $\mathcal{L}$ if $\bigcap\limits_{\varphi \in \Phi} Mod^{\sigma}_{\mathcal{L}}(\varphi) \neq \emptyset$;

3. a sentence $\varphi \in L(\sigma)$ is called *true* in $\mathcal{L}$ if $Mod^{\sigma}_{\mathcal{L}}(\varphi) = Str(\sigma)$;

4. we write $\Phi \models^{\mathcal{L}} \varphi$ if each $\mathcal{L}$-model of all sentences from Φ is also a $\mathcal{L}$-model of φ (i.e. if $Mod^{\sigma}_{\mathcal{L}}(\Phi) \subseteq Mod^{\sigma}_{\mathcal{L}}(\varphi)$).

We say that a logic $\mathcal{L}$ has:

1. *Löwenheim-Skolem property* if each $\mathcal{L}$-satisfiable sentence $\varphi \in L(\sigma)$ has a countable model.

2. *compactness property* if for any $\Phi \subseteq L(\sigma)$: if each finite subset of Φ is $\mathcal{L}$-satisfiable, then Φ is $\mathcal{L}$-satisfiable.

We may now formulate the main result obtained by Lindström:

> Lindström first theorem. *Let $\mathcal{L}$ be a regular abstract logic and let $\mathcal{L}_{\omega\omega} \leqslant \mathcal{L}$. If $\mathcal{L}$ is compact and has the Löwenheim-Skolem property, then $\mathcal{L} \sim \mathcal{L}_{\omega\omega}$.*

This theorem thus states that the classical first-order logic has a special maximality property: a proper compact extension of it which has the Löwenheim-Skolem property does not exist. In other words, the classical first-order logic is the maximal logic possessing these two properties.

In order to formulate the *Lindström second theorem* one should introduce many further notions related to recursion theory. We omit this topic here, although the reader may consult Lindström's original papers or the textbooks mentioned above. Let us say only that the theorem states the maximality property of the classical first-order logic with respect to the Löwenheim-Skolem property and the property that the set of tautologies is recursively enumerable.

3.5 Examples

In this section we briefly comment on the expressive and deductive power of certain systems of logic. We take into account:

1. *First-order logic.* We assume that readers are familiar with the rudiments of this system of logic.

2. *Second-order logic.* In this logic one allows quantification over sets of individuals (or, more generally, over predicates).

3. *Infinitary logics.* If one allows infinitary long conjunctions and alternatives, then one obtains an infinitary logic (one can also admit infinitely long quantifier prefixes). If one allows infinitary rules of inference (as for instance the ω-rule), then one also obtains an infinitary logic.

4. *Logics with generalized quantifiers.* One can extend the collection of classical logical constants (i.e. propositional connectives and quantifiers) by adding new elements to it. Generalized quantifiers may constitute such new elements – they enable us to talk about, for example, infinitely many objects satisfying a formula.

We do not consider systems of non-classical logic (e.g. modal or many-valued logics). This is because these systems of logic have fewer applications in mathematical research than first or second-order logic.

3.5.1 First-order logic

First-order logic (often abbreviated to FOL) has very nice deductive properties but its expressive power is extremely weak. It should be added that it is precisely this logic to which other logical systems are compared when evaluated as having good or bad deductive properties. This is due to tradition and to the fact that first-order logic has been explored for a long time and in many aspects. The same applies to the classical propositional calculus – it may be thought of as a standard logical system. Metalogical notions were initially developed precisely for exploring the properties of such standard systems as the two mentioned above.

The language of FOL is also a language in which many important mathematical theories are formulated, notably the Zermelo-Fraenkel set theory. It should be kept in mind that there is a difference between the expressive power of a system of logic and the expressive power of a theory formulated in the language of that logic. The first may be poor while the second very rich, as is the case exactly with the language of FOL and set theory formulated in that language.

First-order logic is consistent, sound, complete, and compact, and the Löwenheim-Skolem theorem holds for it. These facts (and many other) are advantages as far as the deductive power of FOL is concerned. On the other hand, the expressive power of FOL is extremely weak: one can not characterize in FOL such fundamental mathematical notions as infinity, continuity, or a set of Lebesgue measure zero, and others. Due to the Löwenheim-Skolem theorem no non-trivial theory in the language of FOL can be categorical. Unlike the classical propositional calculus, FOL is not decidable. This last fact is neither an advantage nor a disadvantage – it is simply a state of affairs which must be accepted. However, the monadic fragment of FOL is decidable. The following facts about FOL are well known:

1. The notion of infinity is not expressible in the language of FOL. We recall Dedekind's definition of an infinite set: a set is infinite if and only if it is equinumerous with its proper subset. In the case of FOL, a denotation of a unary predicate is infinite if and only if there exists a bijection between this denotation and its proper subset. Quantification over functions is not allowed in the language of FOL.

2. If a set of sentences has arbitrarily large finite models, then it has an infinite model. A nice proof of this fact uses the ultraproduct construction.

3. There exist sentences which have only infinite models. Indeed, if R is a binary predicate, then the conjunction of the following conditions has infinite models only:

   (a) R is asymmetric: $\forall x \forall y (R(x,y) \rightarrow \neg R(y,x))$

   (b) R is transitive: $\forall x \forall y \forall z (R(x,y) \wedge R(y,z) \rightarrow R(x,z))$

   (c) R is serial: $\forall x \exists y R(x,y)$

The first-order Peano arithmetic cannot characterize its standard model. There are continuum many countable models of this theory which are not elementarily equivalent (and consequently also continuum many pairwise non-isomorphic countable models).

The field of real numbers is uniquely characterized in a second-order language as a completely ordered field. Tarski proposed such an axiomatization in 1936. This field can also be characterized uniquely with respect to elementary equivalence – in this case we employ the axioms for a real closed field. Let us recall that a field F is called *formally real*, if for any its elements $a_1, a_2, \ldots, a_n$ the condition

$$a_1^2 + a_2^2 + \ldots + a_n^2 = 0$$

implies that $a_1 = a_2 = \ldots = a_n = 0$. This, in turn, is equivalent to any of the following conditions:

1. -1 is not a sum of squares in F.

2. F can be linearly ordered (and the ordering is compatible with arithmetical operations in F).

A field F is a *real closed field* if it satisfies any of the following conditions (there exist more conditions equivalent to them):

1. F is a formally real field which does not have any algebraic extension being a formally real field.

2. -1 is not a square in F and the field $F(\sqrt{-1})$ is algebraically closed.

3. There exists a linear order $\leqslant$ such that $(F, \leqslant)$ is an ordered field in which each positive element has the square root and each polynomial of an odd degree with coefficients from the field F has a root in F.

4. F is not algebraically closed but its algebraic closure is a finite field extension.

Real closed fields are elementarily equivalent with the field of real numbers, meaning that they are semantically indistinguishable from the field of real numbers. If we add to the axioms of ordered fields (formulated in a first-order language) the following axioms:

1. an axiom stating that each positive element has the square root and

2. an axiom schema expressing the fact that each polynomial of an odd degree has at least one root,

then we get (a first-order) axiomatic characterization of real closed fields. Tarski demonstrated in 1951 that this theory admits the elimination of quantifiers and hence it is complete and decidable.

3.5.2 Second-order logic

In general, two types of semantics are considered for second-order logic. One can either admit that the second-order variables run over all subsets of the finite Cartesian products of the universe (of a model) or restrict attention to a subset of that set. Let us denote by L^2 the language of the full second-order logic and by L^{2m} the language of the monadic second-order logic (in which one quantifies only unary predicates). We skip the details concerning the syntax of these languages (cf. Barwise and Feferman 1985). We are going to recall a few examples (known from literature) of sentences from these languages which characterize certain mathematical notions not definable in FOL.

1. Let ψ_1 be the conjunction of the following sentences (in the language with one binary predicate $\prec$):

   (a) A sentence asserting that $\prec$ is a strict linear ordering.

   (b) $\forall X\ (\exists y\ X(y) \rightarrow \exists y\ (X(y) \wedge \neg \exists z\ (X(z) \wedge z \prec y)))$.

Then ψ_1 holds in a structure $\mathfrak{A} = (A, \prec^{\mathfrak{A}})$ (appropriate for the language in question) if and only if $\prec^{\mathfrak{A}}$ is a well ordering of A.

2. Let ψ_2 be the induction axiom, in the language with the constant 0 and a symbol s for the successor function. The models of ψ_2 are exactly those structures $\mathfrak{A}$ for which the domain of $\mathfrak{A}$ is the set $\{\overline{n}^{\mathfrak{A}} : n \in \omega\}$, where $\overline{n}$ is the numeral for the number n and $\overline{n}^{\mathfrak{A}}$ is the interpretation of $\overline{n}$ in $\mathfrak{A}$.

3. Let ψ_3 be the conjunction of the following sentences:

   (a) ψ_1

   (b) $\forall x\ (\exists y\ y \prec x \rightarrow \exists y\ (y \prec x \wedge \forall z\ (z \prec x \rightarrow (z \prec y \vee z = y))))$.

   The second sentence states that each element possessing a predecessor has an immediate predecessor (with respect to the ordering $\prec$). If $\mathfrak{A} \models \psi_3$, then $\prec^{\mathfrak{A}}$ is a well ordering of the universe of this structure which is either finite or of ordinal type ω. Hence the sentence ψ_3 has finite and countably infinite models, but does not have uncountable models.

4. Let $x \preceq y$ means that $x \prec y \vee x = y$. Let ψ_4 be the conjunction of the following sentences:

   (a) A sentence stating that $\prec$ is a dense linear ordering without endpoints.

   (b) $\forall X\ (\exists y\ (X(y) \wedge \exists z \forall y\ (X(y) \rightarrow y \preceq z)) \rightarrow \exists z\ (\forall y\ (X(y) \rightarrow y \preceq z) \wedge \forall w\ (\forall y\ \ (X(y) \rightarrow y \preceq w) \rightarrow (z \preceq w))))$.

   The second sentence here expresses the fact that the ordering under consideration is complete. Thus models of ψ_4 are those structures $\mathfrak{A}$ in which $\prec$ is interpreted as a dense linear complete ordering without endpoints (each non-empty set bounded from above has the least upper bound). For such structures $\mathfrak{A}$ there exists an embedding of the real numbers (with their natural order) in $\mathfrak{A}$. It follows from this that the sentence ψ_4 has models of cardinality at least $2^{\aleph_0}$ but it does not have models of smaller cardinalities.

5. Let ψ_5 be the sentence asserting that each injection is a surjection: $\forall F\ (\forall x \forall y\ (F(x) = F(y) \rightarrow x = y) \rightarrow \forall y \exists x\ F(x) = y)$. Then ψ_5 holds in finite structures only.

6. Let ψ_6 be the sentence $\exists R\ \psi_3$, where in ψ_3 the symbol $\prec$ is replaced by a binary predicate R. Then $\mathfrak{A} \models \psi_6$ if and only if there exists (in its universe) a well ordering without limit points, and hence exactly in the case when $\mathfrak{A}$ is countable.

7. Let ψ_7 be the sentence asserting that there exists a well ordering of the universe such that each of its initial segments is countable. Then $\mathfrak{A} \models \psi_7$ if and only if the cardinality of $\mathfrak{A}$ is not greater than $\aleph_1$.

8. One can construct a sentence in the language of the second-order logic which is valid (holds in all structures) if and only if the continuum hypothesis is true. Let $\psi_8(Y)$ be a formula stating that the set Y is at most countable. We get this formula from the sentence ψ_3 by restricting all its first-order quantifiers to the set Y in the well-known way. We are looking for the sentence which can express the following property: if the universe is equinumerous with the set of real numbers (which has cardinality continuum), then each of its subsets is either at most countable or equinumerous with the whole universe (which means that there is no intermediate cardinality between $\aleph_0$ and $2^{\aleph_0}$). Let ψ_9 be the sentence:

   $\psi_4 \rightarrow \forall Y\ (\psi_8(Y) \vee \exists G\ (\forall x\ Y(G(x)) \wedge \forall x \forall y\ (G(x) = G(y) \rightarrow x = y) \wedge \forall y\ (Y(y) \rightarrow \exists x\ G(x) = y)))$.

   Then ψ_9 is true if and only if the continuum hypothesis is true. This example shows the strength of the semantics of second-order logic. There is no possibility of any effective, finitary proof system existing for the full second-order logic.

From these examples it is also clear that monadic second-order logic does not satisfy the compactness theorem and also that the Löwenheim-Skolem theorem does not hold for this logic.

In FOL the set of (Gödel numbers of) all consequences of any recursive set of (Gödel numbers of) sentences is recursively enumerable. In monadic second-order logic (and hence also in the full second-order logic) the set of valid sentences (tautologies) is not recursively enumerable.

3.5.3 Logics with generalized quantifiers

Andrzej Mostowski introduced *numerical quantifiers* Q_α, where α is an ordinal number (Mostowski 1957). The semantics of the formula

$Q_\alpha x\varphi(x)$ is determined in the following way: $\mathfrak{M} \models Q_\alpha x\varphi(x)$ if and only if the cardinality of the set $\{a \in dom(\mathfrak{M}) : \mathfrak{M} \models \varphi(x)[a]\}$ is at least $\aleph_\alpha$. Here (as well as below) $dom(\mathfrak{M})$ denotes the domain of the structure $\mathfrak{M}$ and $\mathfrak{M} \models \varphi(x)[a]$ means that $\mathfrak{M}$ satisfies the formula $\varphi(x)$ when x is valuated as $a \in dom(\mathfrak{M})$ (we also say that a satisfies $\varphi(x)$ in $\mathfrak{M}$). Model theoretic notation will be described in detail in the next chapter.

In particular, the formula $Q_0 x\varphi(x)$ holds in the structure $\mathfrak{M}$ if and only if there exist infinitely many objects in $dom(\mathfrak{M})$ which satisfy the formula $\varphi(x)$. Thus in the language with this quantifier one can express the notion of infinitude, which was impossible in the language of FOL.

A natural requirement which should be imposed on quantifiers Q_α in order to catch their intuitive meaning is: if $\mathfrak{M} \models Q_\alpha x\varphi(x)$ and $\mathfrak{M}$ is isomorphic with $\mathfrak{N}$, then $\mathfrak{N} \models Q_\alpha x\varphi(x)$.

One can understand Mostowski's quantifiers in a relational way: by a generalized quantifier we mean a family Q of pairs of sets (U, X), where $X \subseteq U$ and if $(U, X) \in Q$ and $|X| = |Y|$, $|U - X| = |V - Y|$, then $(V, Y) \in Q$. Here $|X|$ denotes the cardinality of the set X. Thus:

$$Q_0 = \{(U, X) : |X| \geqslant \aleph_0\}.$$

Mostowski has proved that $L_{\omega\omega}(Q_0)$ (that is, the extension of first-order logic $L_{\omega\omega}$ by adding to it the quantifier Q_0 with semantics described above) is not effectively (recursively) axiomatizable. Let θ be the sentence: $\forall x \neg Q_0 y\ (y < x)$ from $L_{\omega\omega}(Q_0)$ and let Π be the axiomatics of the Peano arithmetic (in the language $L_{\omega\omega}$) and $\mathfrak{N}_0$ the standard model of Π. Then for each sentence ϕ from $L_{\omega\omega}$ we have: $\mathfrak{N}_0 \models \phi$ if and only if for all $\mathfrak{M}$, if $\mathfrak{M} \models \Pi$, then $\mathfrak{M} \models (\theta \rightarrow \phi)$. And because there is no effective procedure for deciding the left side of this equivalence, there does not exists a complete proof technique for $L_{\omega\omega}(Q_0)$. The problem, posed by Mostowski, regarding whether the logic $L_{\omega\omega}(Q_1)$ is axiomatizable was answered in the affirmative by Keisler. Mostowski also provided a model-theoretic characterization of the first-order quantifiers – any extension of $L_{\omega\omega}$ which satisfies the condition that each sentence which has an infinite model has models of each infinite power is equivalent to the first-order logic.

Mostowski's approach was soon generalized by Per Lindström in a series of papers Lindström 1966a, 1966b, 1966c, 1969. Lindström's general definition has the following form: by a (local) generalized quantifier on U of the type $\langle k_1, \ldots, k_n\rangle$ we mean any n-ary relation between sets $U^{k_1}, \ldots, U^{k_n}$. Mostowski's quantifiers are then of the type $\langle 1\rangle$, because they correspond to a family of subsets of the universe of a model. Quantifiers whose semantics are described in terms of numerical relations between subsets of the universe of a model are of type $\langle 1, 1\rangle$. Examples of such relations between subsets of a given universe U are:

1. $\{(X, Y) : |X| = |Y|\}$
2. $\{(X, Y) : |X| \leqslant |Y|\}$
3. $\{(X, Y) : |X - Y| > |X \cap Y|\}$.

Leon Henkin introduced yet another generalized quantifier – one in which the quantifier prefix is not linear but partially ordered (Henkin 1961). As we remember, the prefix normal form of formulas of first-order logic is linear and one can eliminate the existential quantifiers from it by introducing Skolem functions. But in the case of a linear prefix, each existential quantifier depends on all universal quantifiers preceding it. It is thus impossible to express the idea of making *independent* choices. A formula with the Henkin quantifier has the following form:

$$\begin{array}{l} \forall x \text{---} \exists y \\ \forall u \text{---} \exists v \end{array} \Big> \varphi(x, y, u, v)$$

Let us denote this formula by $Q_H xyuv\, \varphi(x, y, u, v)$. The semantics of this quantifier (which is a quantifier of the type $\langle 4\rangle$ in Lindström's sense) is determined by the following condition: $\mathfrak{A} \models Q_H xyuv \varphi(x, y, u, v)$ if and only if there exist functions f and g (defined on $dom(\mathfrak{A})$ and with values in $dom(\mathfrak{A})$) such that $\mathfrak{A} \models \alpha(x, f(x), u, g(u))$.

The language with the Henkin quantifier has greater expressive power than FOL. One can show that Q_0 is definable in terms of the Henkin quantifier. Further generalizations as far as the partially ordered quantifiers are concerned were proposed by Barwise (cf. Barwise and Feferman 1985).

Let us make a brief recollection of a few facts concerning the expressive power of languages with additional generalized quantifiers.

The quantifier Q_0 (there exist infinitely many elements). The semantics of this quantifier was given above. Here are a few properties of the language L_{Q_0}:

1. The standard model of arithmetic can be characterized uniquely (up to isomorphism) in L_{Q_0}. It is sufficient to add to the axioms

of discrete linear ordering (determined by the successor function) the sentence: $\forall x \neg Q_0 y\ y < x$.

2. The upward Löwenheim-Skolem theorem does not hold in L_{Q_0}.
3. L_{Q_0} is not recursively axiomatizable.
4. The compactness theorem does not hold in L_{Q_0}.
5. The downward Löwenheim-Skolem theorem holds in L_{Q_0}.
6. One can get the completeness of L_{Q_0} (as a system with infinitary proofs) by adding to it an infinitary rule of inference:

$$\frac{\exists^{\geqslant 1} x\, \varphi(x), \exists^{\geqslant 2} x\, \varphi(x), \ldots}{Q_0 x\, \varphi(x)}.$$

7. Any countable $\aleph_0$-categorical theory in L_{Q_0} without finite models is complete.
8. The theory of dense linear ordering is complete in L_{Q_0}.
9. L_{Q_0} is a fragment of the infinitary logic $L_{\omega_1\omega}$ (cf. the next section), which is clearly visible from the equivalence:

$$Q_0 x\, \varphi(x) \equiv \bigwedge_{n<\omega} \exists^{\geqslant n} x\, \varphi(x).$$

The quantifier Q_1 (there exist uncountably many elements). The expression $Q_1 x\, \varphi(x)$ should be read: there exist uncountably many x such that $\varphi(x)$. The semantics of this quantifier is determined by the following condition: $\mathfrak{A} \models Q_1 x\, \varphi(x)$ if and only if the set $\{a \in dom(\mathfrak{A}) : \mathfrak{A} \models \varphi[a]\}$ is uncountable.

Only uncountable interpretations only are taken into account, for obvious reasons. In L_{Q_1} one can define countability (in L_{Q_0} one can define finiteness). Some properties of L_{Q_1} are as follows:

1. The theory of dense linear ordering is not complete in L_{Q_1}.
2. The upward Löwenheim-Skolem theorem does not hold in L_{Q_1}. The downward Löwenheim-Skolem theorem holds in the following version: if a theory (of power at most $\aleph_1$) in L_{Q_1} has a model, then it has a model of power $\aleph_1$.

3. Each $\aleph_1$-categorical theory in L_{Q_1} (of power at most $\aleph_1$) is complete.
4. The theory of real closed fields of characteristic zero is complete in L_{Q_1}.
5. L_{Q_1} is axiomatizable.

Chang quantifier. The expression $Q_c x\, \varphi(x)$ should be read: there exist as many x which satisfy $\varphi(x)$ as there are elements in the whole universe of the model. The semantics of this quantifier is determined by the following condition: $\mathfrak{A} \models Q_c x\, \varphi(x)$ if and only if the set $\{a \in dom(\mathfrak{A}) : \mathfrak{A} \models \varphi[a]\}$ has the same power as the set $dom(\mathfrak{A})$. Here are some properties of L_{Q_c}:

1. In models of the power $\aleph_1$ the quantifiers Q_c and Q_1 have the same interpretation.
2. The theory of dense linear ordering is not complete in L_{Q_c}.
3. The theory of real closed fields of characteristic zero is complete in L_{Q_c}. This theory admits the elimination of the quantifiers $\exists$ and Q_c.
4. If a countable theory in L_{Q_c} has a model whose power is a successor cardinal number, then it has a model of power $\aleph_1$.
5. If a countable theory in L_{Q_c} has a countable model, then it has models in each infinite power.

The quantifier "there are more A then B". This is still another generalization of the notion of quantifier – in this case a quantifier is joined with two formulas. The expression $Q_M x\, \varphi(x)\psi(x)$ should be read: there are more x such that $\varphi(x)$ than x such that $\psi(x)$. The semantics of this quantifier is determined by the following condition: $\mathfrak{A} \models Q_M x\, \varphi(x)\psi(x)$ if and only if the power of the set $\{a \in dom(\mathfrak{A}) : \mathfrak{A} \models \varphi[a]\}$ is bigger than the power of the set $\{a \in dom(\mathfrak{A}) : \mathfrak{A} \models \psi[a]\}$. Here are some properties of L_{Q_M}:

1. The standard model of first-order Peano arithmetic can be characterized uniquely (up to isomorphism) in L_{Q_M}.
2. L_{Q_M} is not axiomatizable.

3. L_{Q_M} does not satisfy: the compactness theorem, the upward and downward Löwenheim-Skolem theorem.

4. Q_0 and Q_c are definable in L_{Q_M}.

The Henkin quantifier Q_H. Its semantics was given above. Here are some properties of L_{Q_H}:

1. The quantifiers Q_0, Q_c and Q_M are definable in L_{Q_H}.

2. L_{Q_H} is not recursively axiomatizable.

3. The compactness theorem does not hold in L_{Q_H}.

4. Neither the downward nor upward Löwenheim-Skolem holds in L_{Q_H}.

3.5.4 Infinitary logics

We say that a logic is infinitary, if at least one of the following holds:

1. we allow infinitary syntactic constructions (for example infinite conjunctions and disjunctions, infinite quantifier prefixes),

2. we allow infinitary rules of inference (for instance ω-rule) and, consequently, infinitary proofs.

Languages with an uncountable number of constants are sometimes considered in model theory. The underlying logic remains, nevertheless, first-order.

At the beginning of the second half of the 20th century the interest in infinitary logics began to grow. Some important mathematical notions which are not expressible in first-order logic $L_{\omega\omega}$ are expressible in languages $L_{\alpha\beta}$, in which we allow conjunctions and disjunctions of length α and quantifier prefixes of length β (here α and β are ordinal numbers), for instance:

1. The standard model of Peano arithmetic can be characterized in a unique way (up to isomorphism) in $L_{\omega_1\omega}$.

2. The class of all finite sets can be characterized in $L_{\omega_1\omega}$.

3. The theory of ordered Archimedean fields is finitely axiomatizable in $L_{\omega_1\omega}$.

4. The truth predicate for formulas in a language with a countable number of symbols is definable in $L_{\omega_1\omega}$.

5. The notion of well ordering is not definable in $L_{\omega_1\omega}$, but it is definable by a single formula from $L_{\omega_1\omega_1}$ which is the conjunction of the sentence defining linear ordering and the sentence:
$$(\forall x_n)_{n\in\omega}\exists x\,(\bigvee_{n\in\omega}(x = x_n) \wedge \bigwedge_{n\in\omega}(x \leqslant x_n)).$$

6. *Scott theorem.* Any countable structure with countably many relations can be uniquely characterized (up to isomorphism) in $L_{\omega_1\omega}$.

7. The sentence $Q_0 x\ \psi(x)$ from L_{Q_0} (saying that there are infinitely many x such that $\psi(x)$) has exactly the same models as the following sentence from $L_{\omega_1\omega}$:
$$\neg \bigvee_{n\in\omega} \exists x_1 \ldots \exists x_n \forall x\ (\psi(x) \rightarrow (x = x_1 \vee \ldots \vee x = x_n)).$$

We omit the details concerning the precise formulation of syntax of infinitary languages (cf. Barwise and Feferman 1985). It is clear that one must use set theory for the characterization of infinitary operations in the languages $L_{\alpha\beta}$ with conjunctions and disjunctions of length at most α and quantifier prefixes of length at most β. By $L_{\infty\omega}$ one understands the language with arbitrarily long conjunctions and disjunctions and finite quantifier prefixes.

Infinitary logics in which we admit infinite quantifier prefixes are close to the full second-order logic and, consequently, they are not complete. The logic $L_{\omega_1\omega}$ is very special among infinitary logics with finite quantifier prefixes. It has the completeness property under the assumption that its infinitary rule of inference allows the conjunction $\bigwedge \Phi$ to be obtained from *countable* set of premisses Φ. This condition is essential, because one can prove that there exists an uncountable set of sentences from $L_{\omega_1\omega}$ which does not have a model though each of its countable subsets has a model. This example also shows that the compactness theorem does not hold in $L_{\omega_1\omega}$ (and the same concerns all $L_{\alpha\beta}$, where $\alpha \geqslant \omega_1$). Nevertheless, we can formulate a suitable generalization of the notion of compactness and it is the case that such a generalized version of the compactness theorem for infinitary logics is related to the question of the existence of large cardinal numbers.

Some differences between first-order logic and infinitary logics are the following:

1. The downward Löwenheim-Skolem theorem has a counterpart in $L_{\omega_1\omega}$ and, more generally, in all infinitary logics. The upward Löwenheim-Skolem theorem does not hold in these logics.
2. The compactness theorem does not hold in $L_{\omega_1\omega}$ (and, more generally, in no logic $L_{\kappa\lambda}$, where $\kappa \geqslant \aleph_1$).
3. The Craig interpolation lemma holds in $L_{\omega_1\omega}$ (while it does not hold in any other infinitary logic).

Chapter 4

Categoricity and completeness results in model theory

In this chapter we recall briefly some of the most important results concerning categoricity, categoricity in power and completeness obtained in classical model theory. We also add concise information about certain problems investigated in modern model theory. The initial version of this chapter was a rather lengthy text in which we provided numerous definitions from the textbooks of model theory. We have decided to skip that material in this final version. The reader interested in model theory may consult many good textbooks on model theory, both classical and modern, for example: Barwise 1977, Bell and Machover 1977, Bell and Slomson 1969, Cori and Lascar 2001, Ebbinghaus, Flum and Thomas 1996, Hedman 2004, Hinman 2005, Hodges 1993, Marcja and Toffalori 2003, Marker 2002, and Poizat 1999. Some of them also contain valuable information concerning the history of model theory – in this respect one can consult for instance: Hodges 2000 or Vaught 1974.

4.1 Goals of model theory

The term *model theory* was introduced by Alfred Tarski (Tarski 1954). Systematic investigations of models began at around that time. The roots of model theory, however, should be traced to much earlier results. They include, among others:

1. A theorem proved in 1915 by Leopold Löwenheim (and improved in 1920 by Thoralf Skolem) which, in modern terms, means that if a first-order theory has an infinite model, then it has a countable model.

2. Kurt Gödel proved completeness and compactness theorems for countable languages in 1930.

3. The existence of non-standard models of arithmetic was proved by Thoralf Skolem and, independently, Anatoly Maltsev in the nineteen-thirties.

4. A very fundamental role for the origin and development of model theory was played by Alfred Tarski's semantical research in the nineteen-thirties: the definition of satisfiability and truth in a model as well as investigations into relational structures and definable sets of real numbers. This also concerns the method of quantifier elimination investigated by Tarski and his collaborators at his seminar in Warsaw from 1927 onwards.

5. Kurt Gödel's incompleteness results from 1931 also had an obvious influence on mathematical and logical investigations into theories and their models.

The road from the mathematical investigation of structures (as, for example, the works of Dedekind, Hilbert, Huntington, Veblen and others) through the important Tarski's work on formal semantics in the nineteen-thirties up to the beginning of model theory from the nineteen-fifties (Alfred Tarski, Abraham Robinson, Leon Henkin, Robert Vaught and others) has its own history, described for example in Hodges 2000.

Classical systematic model theory originated in a situation where the standard of first-order languages was prevalent. The goals of the theory were connected to describing the relations holding between theories formulated in such languages and structures which interpreted them. Relations between structures, in particular relations expressed in semantic terms, were also investigated. Several operations on models were invented and one investigated which types of formulas were preserved under such operations. Of great importance was the problem of mathematical (e.g. algebraic) representation of semantic notions. Some special types of model entered the scene. There was also growing interest in the problems of definability as well as the conditions related to categoricity (and categoricity in power), completeness, and decidability, and so forth.

The notions of maximality and minimality of models and theories can be expressed in several styles in model theory. One may for instance ask which subsets of the domain of a model are definable in the corresponding theory. It appears that theories with a small amount of subsets definable in models have interesting metalogical properties. "Poor" and "rich" models (with respect to semantic properties) can also be described in terms of types realized in those models – atomic models are "poor", while saturated models are "rich". Types are, intuitively speaking, descriptions of "positions" of n-tuples of elements of the domain of a model (for a precise definition see below). Investigation of the spaces of types of theory recently emerged as one of the most important research activities in model theory.

We declared in the Preface that we assume the readers' familiarity with the rudiments of semantics of the first-order languages. As such we are not going to recall explicitly all necessary concepts, but shall instead limit ourselves to establishing the notation.

We will consider first-order languages with a countable signature (that is, with a countable lists of predicates, function symbols and individual constants). The first-order language with signature σ will be denoted by $L(\sigma)$. The class of all relational structures which are interpretations of such language will be denoted by $Str(\sigma)$. The domain of a relational structure $\mathfrak{M}$ will be denoted by $dom(\mathfrak{M})$. The meaning of the following symbols is standard:

1. $\mathfrak{M} \models_a \varphi(x_1, \ldots, x_n)$: the formula $\varphi(x_1, \ldots, x_n)$ is satisfied in the structure $\mathfrak{M}$ by the valuation $a = (a_i)_{i\in\omega}$. In what follows, we will write simply $\mathfrak{M} \models \varphi[a_1, \ldots, a_n]$ when the formula $\varphi(x_1, \ldots, x_n)$ with free variables $x_1, \ldots, x_n$ is satisfied in the structure $\mathfrak{M}$ by the sequence $[a_1, \ldots, a_n]$, where x_i is interpreted as a_i. It is also convenient to write $\mathfrak{M} \models \varphi[\vec{a}]$, when a is a finite sequence $(a_1, \ldots, a_n)$ and similarly $\varphi(\vec{x})$, when x is a finite sequence $(x_1, \ldots, x_n)$.

2. For any predicate $P \in \sigma$ and any structure $\mathfrak{M} \in Str(\sigma)$ by $P^{\mathfrak{M}}$ we mean the realization of P in $\mathfrak{M}$ (and similarly for function symbols and individual constants).

3. If $\vec{a}$ is a sequence of elements from the domain of $\mathfrak{M}$, then by $(\mathfrak{M}, \vec{a})$ we mean the structure $\mathfrak{M}$ in which the elements of $\vec{a}$ are new distinguished elements.

4. If $\mathfrak{M}$ is a structure of signature σ and $A \subseteq dom(\mathfrak{M})$, then by $L(\sigma \cup A)$ we mean the first-order language obtained from $L(\sigma)$ by adding to it constants c_a for all $a \in A$.

5. $\mathfrak{M} \models \varphi$ means that the sentence φ is true in the structure $\mathfrak{M}$, that is, it is satisfied by all valuations ($\mathfrak{M}$ is a model of φ).

6. $\mathfrak{M} \models \Phi$ means that $\mathfrak{M} \models \varphi$ for all $\varphi \in \Phi$, where Φ is a set of sentences of the language in question ($\mathfrak{M}$ is a model of the set Φ).

7. $Mod(\varphi) = \{\mathfrak{M} \in Str(\sigma) : \mathfrak{M} \models \varphi\}$. Thus, $Mod(\varphi)$ is the class of all models of φ.

8. $Mod(\Phi) = \bigcap_{\varphi \in \Phi} Mod(\varphi)$. Thus $Mod(\Phi)$ is the class of all models of the set of sentences Φ.

9. $Th(\mathfrak{M}) = \{\varphi : \mathfrak{M} \models \varphi\}$. Thus $Th(\mathfrak{M})$ is the theory of $\mathfrak{M}$.

10. $Th(K) = \bigcap_{\mathfrak{M} \in K} Th(\mathfrak{M})$. Thus $Th(K)$ is the set of all sentences true in all structures from a class K of structures.

If the structures $\mathfrak{M}$ and $\mathfrak{N}$ are isomorphic, then we write $\mathfrak{M} \cong \mathfrak{N}$. If $\mathfrak{M}$ is a substructure of $\mathfrak{N}$ (which means that $dom(\mathfrak{M}) \subseteq dom(\mathfrak{N})$ and $\mathfrak{M}$ interprets the symbols from the signature in the same manner as $\mathfrak{N}$ interprets them in $dom(\mathfrak{M})$), then we write $\mathfrak{M} \subseteq \mathfrak{N}$. We say that structures $\mathfrak{M}$ and $\mathfrak{N}$ are elementarily equivalent, in symbols $\mathfrak{M} \equiv \mathfrak{N}$, if $Th(\mathfrak{M}) = Th(\mathfrak{N})$, that is, if their sets of true sentences coincide. We say that $\mathfrak{M}$ is an elementary substructure of $\mathfrak{N}$ (and $\mathfrak{N}$ is an elementary extension of $\mathfrak{M}$), in symbols $\mathfrak{M} \prec \mathfrak{N}$, if $\mathfrak{M} \subseteq \mathfrak{N}$ and $\mathfrak{M} \models \varphi[\vec{a}]$ if and only if $\mathfrak{N} \models \varphi[\vec{a}]$ for any formula $\varphi(\vec{x})$ and any n-tuple $\vec{a}$ of elements of $dom(\mathfrak{M})$. If $\mathfrak{M} \equiv \mathfrak{N}$, then $\mathfrak{M} \prec \mathfrak{N}$, but not vice versa. We say that a function $f : dom(\mathfrak{M}) \rightarrow dom(\mathfrak{N})$ is an elementary embedding of $\mathfrak{M}$ into $\mathfrak{N}$, if f is an injection and for every formula $\psi(\vec{x})$ and every sequence $\vec{a}$ of elements of $dom(\mathfrak{M})$: $\mathfrak{M} \models \psi[\vec{a}]$ if and only if $\mathfrak{N} \models \psi[f(\vec{a})]$.

The fundamental tools used in model theory include applications of the completeness and compactness theorems for first-order logic, as well as the Löwenheim-Skolem theorem. Essential use is made of the fact that one can add new individual constants to the formal language in question which are then interpreted in models.

Theories in model theory are understood as explained in chapter 2, meaning either as sets of axioms or as deductively closed sets of sentences or sets of sentences true in a given model.

4.2 Examples of categoricity and completeness results

We recall that a theory is κ-categorical if it has a model of power κ and all its models in that power are isomorphic. A theory is (semantically) complete if all its models are elementarily equivalent. A theory is model complete if for any two its models $\mathfrak{M}$ and $\mathfrak{N}$, if $\mathfrak{M} \subseteq \mathfrak{N}$, then $\mathfrak{M} \prec \mathfrak{N}$.

Classical model theory has elaborated many tools used in the characterization of categoricity in power, completeness and model completeness. For instance, a syntactic procedure of quantifier elimination found applications in the proofs that the following theories are complete:

1. The theory of dense linear ordering without endpoints (Langford).
2. Arithmetic with addition and without multiplication (Presburger).
3. Theories of algebraically closed fields and real closed fields (Tarski).
4. The theory of Boolean algebras (Tarski).
5. The theory of abelian groups (Szmielew).

Let us recall that a theory T admits elimination of quantifiers if for every formula $\psi(\vec{x})$ of the language of this theory there exists a formula $\varphi(\vec{x})$ without quantifiers (hence being a Boolean combination of atomic formulas) such that $\psi(\vec{x})$ and $\varphi(\vec{x})$ are T-equivalent (that is, their equivalence can be proved in T). If T admits elimination of quantifiers, then we can obtain some information about the decidability of T. Of course, the method of elimination of quantifiers cannot be applied in theories which are undecidable (Peano arithmetic, set theory, group theory, field theory, or theory of partial order).

Elementary equivalence between models can be described in algebraic terms, using families of partial isomorphisms which can be back-and-forth extended. This result is ascribed to Roland Fraïssé. Andrzej Ehrenfeucht obtained the result independently formulating it in terms of games between two players.

A fundamental theorem connecting categoricity in power and completeness is the *Łoś-Vaught test*, which says that if T is a consistent theory without finite models which is κ-categorical in some infinite cardinal κ, then T is complete.

The relation of elementary substructure (and hence also the property of model completeness) can be characterized by the *Tarski-Vaught test*:

if $\mathfrak{M} \subseteq \mathfrak{N}$, then $\mathfrak{M} \prec \mathfrak{N}$ if and only if for any formula $\varphi(\vec{x}, y)$ and any sequence of elements $\vec{a}$ from the domain of $\mathfrak{M}$, if $\mathfrak{N} \models \exists y \varphi(\vec{x}, y)[\vec{a}]$, then there exists an element $b \in dom(\mathfrak{M})$ such that $\mathfrak{M} \models \varphi[\vec{a}, b]$.

An early result concerning countable categoricity of complete theories (that is, $\aleph_0$-categoricity) is the theorem proved in 1959 independently by Ryll-Nardzewski, Engeler and Svenonius which says that a complete theory T is $\aleph_0$-categorical if and only if for each natural number n there are only finitely many formulas in n free variables up to the relation of T-equivalence.

In 1954 Jerzy Łoś formulated the conjecture that the only possibilities of categoricity (in power) of theories are the following:

1. $(+, -)$: T is $\aleph_0$-categorical, but is not κ-categorical for all uncountable κ. Example: the theory of dense linear ordering without endpoints.

2. $(-, +)$: T is not $\aleph_0$-categorical, but is κ-categorical for all uncountable κ. Example: the theory of algebraically closed fields of characteristic 0 (or p, where p is a prime number).

3. $(+, +)$: T is κ-categorical for all cardinals κ. Example: the pure theory of identity.

4. $(-, -)$: For all cardinals κ, T is not κ-categorical. Example: the theory of the standard model of first-order Peano arithmetic PA.

In 1965 Michael Morley positively confirmed Łoś's conjecture. The Morley categoricity theorem says that: for any countable theory T, T is κ-categorical in some uncountable power κ if and only if T is κ-categorical for all uncountable powers.

There are obviously numerous further results in classical model theory which concern completeness, model completeness, categoricity in power and other properties of theories and models. We have chosen only a few examples; the inquiring reader may consult the textbooks mentioned at the beginning of this chapter for more comprehensive information.

4.3 Ultraproducts

Several types of operations on models are investigated in model theory. Let us present an example of one such construction, which will be used in chapter 5.

Let us recall that by a filter on a set I we mean a family F of subsets of I such that:

1. $I \in F$ and $\emptyset \notin F$
2. if $A, B \in F$, then $A \cap B \in F$
3. if $A \in F$ and $A \subseteq B \subseteq I$, then $B \in F$.

Intuitively, elements of a filter on I are "big" subsets of I. For instance, all cofinite subsets of the set of all natural numbers (that is, sets with finite complement) form a filter. Also the family of all sets of real numbers whose complements have Lebesgue measure zero forms a filter. Given a set I and any $x \in I$, the set $\{X \subseteq I : x \in X\}$ is a filter, called a principal filter (generated by x). We say that a filter F on I is an ultrafilter on I, if $X \in F$ or $I - X \in F$ for any $X \subseteq I$. One proves that each filter can be extended to an ultrafilter.

Let I be an infinite set and $(\mathfrak{A}_i)_{i \in I}$ a family of structures with the same signature σ. We will define the ultraproduct $\prod\limits_{i \in I} \mathfrak{A}_i / F$ of this family with respect to the ultrafilter F.

Let $\Pi = \{f : I \to \bigcup\limits_{i \in I} dom(\mathfrak{A}_i) : f(i) \in dom(\mathfrak{A}_i) \text{ for all } i \in I\}$. Define the relation $\sim_F$ on Π by: $f \sim_F g$ if and only if $\{i \in I : f(i) = g(i)\} \in F$. Then $\sim_F$ is an equivalence relation on Π.

The domain of the structure $\prod\limits_{i \in I} \mathfrak{A}_i / F$ is the set $\Pi/_{\sim_F}$ of all $\sim_F$-equivalence classes, that is, $\Pi/_{\sim_F} = \{[g]_{\sim_F} : g \in \Pi\}$. For typographical reasons let us accept the abbreviations (in the present definition only): $\mathfrak{A} = \prod\limits_{i \in I} \mathfrak{A}_i / F$ and $[g] = [g]_{\sim_F}$. The interpretations of individual constants, function symbols and predicates in $\prod\limits_{i \in I} \mathfrak{A}_i / F$ are defined as follows.

If c is an individual constant, then $c^{\mathfrak{A}}$ is the $\sim_F$-equivalence class of the function $f_c \in \Pi$, where $f_c(i) = c^{\mathfrak{A}_i}$ for all $i \in I$.

Let G be an n-argument function symbol and let $g_1, \ldots, g_n \in \Pi$. Then:

$$G^{\mathfrak{A}}([g_1(i)], \ldots, [g_n(i)]) = [G^{\mathfrak{A}_i}(g_1(i), \ldots, g_n(i))].$$

Let R be an n-argument predicate. Then:

$$R^{\mathfrak{A}} = \{([g_1], \ldots, [g_n]) \in \Pi/_{\sim_F} : \{i \in I : (g_1(i), \ldots, g_n(i)) \in R^{\mathfrak{A}_i}\} \in F\}.$$

One can prove that all the above definitions are correct, meaning that they do not depend on the choice of elements from the equivalence classes.

It follows immediately from this definition that for each atomic formula $\psi(x_1, \ldots, x_n)$ and all $g_1, \ldots, g_n \in \Pi$ the following conditions are equivalent:

1. $\prod\limits_{i \in I} \mathfrak{A}_i/F \models \psi[[g_1]_F, \ldots, [g_n]_F]$
2. $\{i \in I : \mathfrak{A}_i \models \psi[g_1(i), \ldots, g_n(i)]\} \in F$.

If all the structures $\mathfrak{A}_i$ are identical, say $\mathfrak{A}_i = \mathfrak{B}$ for all $i \in I$, then the ultraproduct $\prod\limits_{i \in i} \mathfrak{A}_i/F$ is called the ultrapower of $\mathfrak{B}$. Sometimes one uses the notation $\mathfrak{B}^I/F$ for ultrapowers.

Some important theorems concerning ultraproducts are the following:

1. *Łoś theorem.* Let $\{\mathfrak{A}_i : i \in I\}$ be a family of structures of the same signature and F an ultrafilter on I. For any formula $\psi(x_1, \ldots, x_n)$ with free variables $x_1, \ldots, x_n$ and for all $g_1, \ldots, g_n \in \Pi$ the following conditions are equivalent:
   (a) $\prod\limits_{i \in I} \mathfrak{A}_i/F \models \psi[[g_1]_F, \ldots, [g_n]_F]$
   (b) $\{i \in I : \mathfrak{A}_i \models \psi[g_1(i), \ldots, g_n(i)]\} \in F$.
2. *Frayne theorem.* $\mathfrak{A} \equiv \mathfrak{B}$ if and only if $\mathfrak{A}$ is elementarily embeddable in some ultrapower $\mathfrak{B}^I/F$.
3. *Keisler theorem.* Under the assumption of the generalized continuum hypothesis, for any structures $\mathfrak{A}$ and $\mathfrak{B}$ of power at most κ^+ (successor of κ) the following conditions are equivalent:
   (a) $\mathfrak{A}$ and $\mathfrak{B}$ are elementarily equivalent.
   (b) There exist ultrafilters F_1 and F_2 such that the ultrapowers $\mathfrak{A}^\kappa/F_1$ and $\mathfrak{B}^\kappa/F_2$ are isomorphic.

The operation of forming ultrapowers can be iterated and combined with the operation of union of chains of models which gives so-called ultralimits. The Kochen theorem states that two structures are elementarily equivalent if and only if they have isomorphic ultralimits.

4.4 Definability

Theories can be also characterized with respect to the families of sets definable in their models. Let $n > 0$ and X be a subset of $dom(\mathfrak{A})^n$. We say that X is definable in $\mathfrak{A}$, if there exists a formula $\psi(\vec{x})$ such that:

$$X = \{\vec{a} \in dom(\mathfrak{A})^n : \mathfrak{A} \models \psi[\vec{a}]\}.$$

If the formula $\psi(\vec{x})$ contains occurrences of individual constants denoting elements of the set $Y \subseteq dom(\mathfrak{A})$, then we say that X is Y-definable in $\mathfrak{A}$ (meaning definable in $\mathfrak{A}$ with parameters from Y).The following facts can be proven:

1. For all $n > 0$ and $Y \subseteq dom(\mathfrak{A})$, the sets Y-definable in $\mathfrak{A}$ form a Boolean algebra.

2. Each finite subset of $dom(\mathfrak{A})^n$ is definable in $\mathfrak{A}$.

3. If $\mathfrak{A}$ is infinite, then there exist subsets of its domain which are not definable in $\mathfrak{A}$.

4. The recursive sets are definable in the standard model of arithmetic.

We say that an infinite structure $\mathfrak{A}$ is minimal if the only sets definable in $\mathfrak{A}$ are finite and cofinite (sets whose complement is finite) subsets of its domain. We say that an infinite structure $\mathfrak{A}$ is strongly minimal if each structure elementarily equivalent with it is minimal. In strongly minimal structures there exist as few definable sets as possible. We say that a complete theory T is strongly minimal, if each model of T is minimal. Each algebraically closed field is a minimal structure. The theory of algebraically closed fields of characteristic 0 or p, where p is a prime number, is strongly minimal.

The minimality property is not preserved under elementary equivalence: there exist minimal structures $\mathfrak{A}$ such that $Th(\mathfrak{A})$ is not a strongly minimal theory. If T is the theory of discrete linear order with a first element but no last element, then the set of all natural numbers with zero, successor and the relation $\leqslant$ is a minimal model of this theory, but no other model of it is minimal. In particular, no non-standard model of Peano arithmetic (reduced to the signature mentioned above) is a model of T.

Using the notions related to definability, one can introduce in model theory such concepts as algebraic closure, independence, basis and dimension which are general counterparts of the corresponding notions from linear algebra. These concepts can be used to prove that:

1. Countable strongly minimal theories are categorical in uncountable powers.
2. Strongly minimal theories are complete.

Let $\mathfrak{A}$ be a strongly minimal structure of signature σ and assume that σ contains a binary predicate $<$ which is interpreted as a linear ordering in $\mathfrak{A}$. By an interval in $\mathfrak{A}$ we mean any of the following subsets of the universe of $\mathfrak{A}$, for some $a, b \in dom(\mathfrak{A})$:

1. $(a, b) = \{x \in dom(\mathfrak{A}) : a < x < b\}$
2. $(a, \infty) = \{x \in dom(\mathfrak{A}) : a < x\}$
3. $(\infty, a) = \{x \in dom(\mathfrak{A}) : x < a\}$

One-element sets are intervals as well. Each interval is of course a set definable in $\mathfrak{A}$. We say that a structure $\mathfrak{A}$ is *o*-minimal if every set definable in $\mathfrak{A}$ is a finite sum of intervals. A theory T is *o*-minimal, if every model of T is *o*-minimal.

o-minimal structures need not be strongly minimal, but these two minimality properties have a lot in common. In *o*-minimal structures there are as few definable sets as possible. Here are a few examples of important *o*-minimal structures:

1. $\mathbb{Q}_< = (\mathbb{Q}, <)$
2. $\mathbb{R}_{ord} = (\mathbb{R}, <, +, \cdot, 0, 1)$
3. $\mathbb{R}_{exp} = (\mathbb{R}, <, exp, +, \cdot, 0, 1)$, where the function $exp(x)$ is interpreted as e^x.

The fact that the first two of these structures are *o*-minimal follows from the fact that the theories: of dense linear ordering without endpoints and $Th(\mathbb{R}_{ord})$ admit elimination of quantifiers. The fact that $\mathbb{R}_{exp}$ is *o*-minimal was proved by Alex Wilkie, who has shown that it is a model complete theory. Whether this theory is decidable remains an open problem.

The structure consisting of all natural numbers with addition, multiplication and natural ordering is not *o*-minimal. Indeed, the sets definable in this structure are very complicated.

4.5 A few words about types

Local properties of models can be described in terms of types. Sentences describe models globally. The elements of models can be characterized by formulas with one free variable which hold true about these elements. In other words, any $a \in dom(\mathfrak{A})$ can be characterized by the set of all formulas $\psi(x)$ for which $\mathfrak{A} \models \psi(x)[a]$. Similarly for collections of n elements of the domain of a structure.

Let Fml_n denote the set of all formulas from the first-order language in question with n free variables. For any structure $\mathfrak{A}$ and any n-element sequence $\vec{a}$ of elements of its domain let:

$$tp_{\mathfrak{A}}(\vec{a}) = \{\psi \in Fml_n : \mathfrak{A} \models \psi[\vec{a}]\}.$$

The set $tp_{\mathfrak{A}}(\vec{a})$ is called the n-type of the element $\vec{a}$ (in the structure $\mathfrak{A}$). Sometimes we say in brief: the type of the element $\vec{a}$, if the number n is clear from the context or irrelevant.

It can be seen that the type of the element $\vec{a}$ in the structure $\mathfrak{A}$ equals the set of all formulas satisfied in $\mathfrak{A}$ by the valuation of the free variables of these formulas by the elements of the sequence $\vec{a}$. It should be clear that different sequences $\vec{a}$ may have the same type in a structure $\mathfrak{A}$.

Let $\Gamma \subseteq Fml_n$. We say that $\mathfrak{A}$ realizes Γ, if $\Gamma \neq \emptyset$ and $\Gamma \subseteq tp_{\mathfrak{A}}(\vec{a})$ for some n-tuple $\vec{a}$ of elements of the domain of $\mathfrak{A}$. If Γ is not realized in $\mathfrak{A}$, then we say that $\mathfrak{A}$ omits Γ. If Γ is realized in some structure, then we say that it is realizable.

By an n-type we mean a realizable set of formulas whose free variables are among $x_1, \ldots, x_n$. A type is an n-type for some n. Obviously, the sets of the form $tp_{\mathfrak{A}}(\vec{a})$ are types – they are called complete types. Types that are not complete are called partial types.

If $\Gamma(x_1, \ldots, x_n) \subseteq Fml_n$ and $c_1, \ldots, c_n$ are constants not occurring in Γ, then let $\Gamma(c_1, \ldots, c_n)$ denote the set of formulas obtained from $\Gamma(x_1, \ldots, x_n)$ by replacing each x_i by c_i. Then $\Gamma(x_1, \ldots, x_n)$ is realizable if and only if $\Gamma(c_1, \ldots, c_n)$ is satisfiable. Moreover, by the compactness argument $\Gamma(x_1, \ldots, x_n)$ is realizable if and only if every finite subset of $\Gamma(x_1, \ldots, x_n)$ is realizable.

For any complete theory T any type that is realized in a model of T is called a type of T. Let $S(T)$ denote the set of all complete types of T and let $S_n(T)$ denote the set of all n-types from $S(T)$. If T is a complete theory and $\mathfrak{A}$ is a model of T, then each type in $S(T)$ is realized in some elementary extension of $\mathfrak{A}$.

Types belonging to $S(T)$ may be realized in some models of T but omitted in other such models. However, there are types which are necessarily realized in all models of a complete theory T.

Let T be a complete theory and let $p \in S_n(T)$. We say that p is an isolated type in $S_n(T)$, if there exists a formula $\theta \in p$ such that p is the only type in $S_n(T)$ which contains θ. One proves that if T is a complete theory and $p \in S(T)$ is an isolated type of T, then p is realized in every model of T. In the case of theories with countable signature one can prove the following *omitting types theorem*: if $p \in S(T)$ is not an isolated type, then there exists a model of T which omits p.

If T is a countable complete theory, then a type from $S(T)$ is isolated if and only if it is realized in every countable model of T. In the case of countable complete $\aleph_0$-categorical theories, every type realized in the countable model of such theory must be isolated. Hence one obtains a characterization of $\aleph_0$-categorical theories in terms of types, or in other words the following version of the theorem mentioned earlier and credited to Ryll-Nardzewski, Engeler and Svenonius holds:

> Let T be a complete theory which has an infinite model. The following conditions are equivalent:
>
> 1. T is $\aleph_0$-categorical.
> 2. The set $S(T)$ is countable and every type in it is an isolated type.
> 3. The set $S_n(T)$ is finite for each natural number n.
> 4. There are only finitely many formulas with n free variables up to the relation of T-equivalence, for each natural number n.

If A is a subset of the domain of $\mathfrak{A} \in Str(\sigma)$ and $\vec{a} = (a_1, \ldots, a_n)$ is an n-tuple of elements of $dom(\mathfrak{A})$, then by the type of $\vec{a}$ over A in $\mathfrak{A}$, denoted $tp_{\mathfrak{A}}(\vec{a}/A)$ we mean the set of all formulas from the language $L(\sigma \cup A)$ with free variables among $x_1, \ldots, x_n$ that are satisfied in $\mathfrak{A}$ when each x_i is replaced by a_i. The types $tp_{\mathfrak{A}}(\vec{a}/A)$ are called complete types over A.

Let us consider a few examples of types for particular theories which are often presented in the textbooks of model theory.

1. The structure $(\mathbb{Q}, <)$, where $\mathbb{Q}$ is the set of all rational numbers and $<$ is the usual ordering of these numbers, is a model of the theory of dense linear ordering without endpoints. This theory admits elimination of quantifiers. Consider any four rational numbers a_1, a_2, a_3, a_4 such that $a_1 < a_2 < a_3 < a_4$. Then the following formulas belong to the 4-type of the sequence (a_1, a_2, a_3, a_4): $x_1 < x_2$, $x_2 < x_3$ and $x_3 < x_4$. Thus, the 4-type of the sequence (a_1, a_2, a_3, a_4) is the same as the 4-type of any increasing sequence of four rational numbers.

2. In the theory of dense linear orderings without endpoints we have three complete 2-types of this theory, which contain, respectively, the formulas: $x_1 < x_2$, $x_1 = x_2$ and $x_2 < x_1$. In general for any n there exist finitely many n-types of this theory. In particular, there exists only one 1-type of this theory.

3. Let us consider the theory of the structure of natural numbers with zero and successor. For any natural number n we have the term $\overline{n}$, meaning the numeral denoting the number n. Such a term is obtained by applying n times the operation of successor to the individual constant 0. There exist at least countably many complete 1-types Ψ_n which contain, respectively, the formula $x = \overline{n}$, for each $n \in \omega$.

4. Let us consider the complete theory of the standard model of arithmetic PA. Let P denote the set of all prime numbers. For any set $X \subseteq P$ let:

$$\Psi_X = \{\exists y\ (p \cdot y = x) : p \in X\} \cup \{\neg\exists y\ (p \cdot y = x) : p \in P - X\}.$$

Then each set Ψ_X is a 1-type of the theory in question. If $X \neq Y$, then the sets Ψ_X and Ψ_Y are incomparable with respect to inclusion and hence are contained in different complete types. This implies that the number of complete types of the theory in question equals $2^{\aleph_0}$. The only sets Ψ_X which are realized in the standard model are those for which the set X is finite. All other sets of the form Ψ_X are realized in the non-standard models of the theory under consideration.

5. Let T be the axiomatic Peano arithmetic and consider the set:

$$\Psi_x = \{x \neq 0, x \neq s(0), x \neq s(s(0)), \ldots\}.$$

The standard model of the theory omits this type. All non-standard models realize this type.

6. Let T be the theory of ordered fields and let Ψ_x be the set:

$$\{1 \leqslant x, 1+1 \leqslant x, 1+1+1 \leqslant x, \ldots\}.$$

An ordered field omits this type if and only if it is Archimedean. In particular, the fields of real numbers and of rational numbers omit this type.

4.6 Special models

Extremal axioms were formulated in order to characterize intended models of certain fundamental theories (arithmetic of natural and real numbers, Euclidean geometry, and set theory). The properties taken into account were those of minimality and maximality (of structure). In general model theory we also consider models which are, in a precise sense, minimal or maximal. It is not only the size of the universe which matters. Rather, we try to characterize models which are "poor" or "rich" with respect to their structure and their semantical properties. Without going into much formal detail, let us recall the characterization of such models:

1. A model $\mathfrak{A}$ is *atomic* if any complete type $tp_{\mathfrak{A}}(\vec{a})$ realized in $\mathfrak{A}$ is an isolated type of the theory $Th(\mathfrak{A})$, for each sequence $\vec{a}$ of elements of $dom(\mathfrak{A})$.

2. A model $\mathfrak{A}$ of a theory T is *prime* if it can be elementarily embedded in any model of T.

3. We say that a structure $\mathfrak{A}$ is $\aleph_0$-*homogeneous* if for any sequences $\vec{a}, \vec{b}$ of elements of $\mathfrak{A}$, if $(\mathfrak{A}, \vec{a}) \equiv (\mathfrak{A}, \vec{b})$, then for each $c \in dom(\mathfrak{A})$ there exists $d \in dom(\mathfrak{A})$ such that $(\mathfrak{A}, \vec{a}, c) \equiv (\mathfrak{A}, \vec{b}, d)$.

4. We say that a model $\mathfrak{A}$ of a complete theory T is *countably universal* if $\mathfrak{B}$ can be elementarily embedded in $\mathfrak{A}$, for any countable model $\mathfrak{B}$ of T.

5. We say that a model $\mathfrak{A}$ of a complete theory T is *weakly saturated* if each type in T is realized in $\mathfrak{A}$.

6. We say that a structure $\mathfrak{A}$ is $\aleph_0$-*saturated* if for any finite sequence $\vec{a}$ of elements of $\mathfrak{A}$ the structure $(\mathfrak{A}, \vec{a})$ is a weakly saturated model of the theory $Th((\mathfrak{A}, \vec{a}))$.

7. A model $\mathfrak{A}$ is called *κ-saturated* if for all subsets $A \subseteq dom(\mathfrak{A})$ of cardinality less than κ, $\mathfrak{A}$ realizes all complete types over A.

   A model $\mathfrak{A}$ is called *saturated* if it is κ-saturated where κ is the cardinality of $\mathfrak{A}$.

Thus atomic models are "poor" in the sense that they realize only those types which must be realized. On the other hand, saturated models are "rich" because they realize many types.

It should be stressed that being atomic or being saturated are not properties which could serve as defining properties of intended models. Obviously, intended models may have such properties, but being an intended model is determined by other factors. Such models should be distinguished in the class of all possible models and this requires taking into account relations between models.

Examples of models of the kinds defined above are:

1. The standard model of Peano arithmetic is its prime model.

2. The ordered field of real algebraic numbers is the unique atomic model of the theory of real closed fields.

3. Any prime model of a countable theory is atomic, which follows at once from the omitting types theorem.

4. The ordered set of rational numbers $(\mathbb{Q}, <)$ is saturated.

5. The set of real numbers $(\mathbb{R}, <)$ with their usual ordering is not saturated. For example, a type consisting of all formulas of the form $x < \frac{1}{n}$ (for all n) and the formula $x > 0$ is not realized in this structure.

6. A dense totally ordered set without endpoints is a η_α-set (meaning a linearly ordered set of power $\aleph_\alpha$) if and only if it is $\aleph_\alpha$-saturated.

7. For theories T with only countably many types, a model $\mathfrak{A}$ of T is saturated if and only if it is countably universal and $\aleph_0$-homogeneous.

Important properties of special kinds of models defined above are for instance the following:

1. A complete theory T has a countable $\aleph_0$-saturated model if and only if the set $S(T)$ is countable.

2. Every countable $\aleph_0$-saturated model of a complete theory is $\aleph_0$-homogeneous.

3. Every countable $\aleph_0$-saturated model of a complete theory is countably universal.

4. For any structure $\mathfrak{A}$ the following conditions are equivalent:

   (a) $\mathfrak{A}$ is $\aleph_0$-saturated.

   (b) $\mathfrak{A}$ is $\aleph_0$-homogeneous and each 1-type in $\mathfrak{A}$ is realized in $\mathfrak{A}$.

   (c) $\mathfrak{A}$ is countably universal and $\aleph_0$-homogeneous.

5. A theory is $\aleph_0$-categorical if and only if it has an atomic model and an $\aleph_0$-saturated model and these models are isomorphic.

6. For any complete theory T and any countable models $\mathfrak{A}$ and $\mathfrak{B}$ of T, if these models are $\aleph_0$-saturated, then they are also isomorphic.

4.7 Classifying theories

Saharon Shelah has developed a theory which provides criteria for classifying theories. It is impossible (for us at least) to summarize even the main trends of the theory, not to mention going into detail. In what follows we are going to say only a few words regarding certain concepts.

4.7.1 Spectra of theories

Let us introduce the general notation: for any theory T and any infinite cardinal number κ let $\mu(T,\kappa)$ denote the number of isomorphism equivalence classes of models of T in the power κ. Thus, if T is κ-categorical, then $\mu(T,\kappa) = 1$. The number $\mu(T,\kappa)$ is not greater than the number of classes of elementary equivalence of models of T. We remember that all models of a complete theory are elementarily equivalent. But a complete theory T may have the number $\mu(T,\kappa)$ greater than 1. For instance, for the theory of dense linear orderings and $\kappa = \aleph_0$ this number equals 4. The number $\mu(T,\kappa)$ describes the mathematical diversity of models of T (of power κ). On the other hand, the number of equivalence classes of the relation of elementary equivalence is related to the expressive power of first-order languages.

It follows from the downward Löwenheim-Skolem theorem that if T is any consistent theory, then $\mu(T, \aleph_0) \geqslant 1$. For any consistent theory T and any infinite cardinal number κ we have:

$$1 \leqslant \mu(T, \kappa) \leqslant 2^\kappa.$$

If T is a complete theory, then the Lindenbaum algebra of its sentences (that is, the algebra of T-equivalence classes) has two elements. For $n > 0$ (when we consider formulas with free variables), the corresponding algebra $\mathcal{L}_T^n$, as well as the set $S_n(T)$, have a much more complicated structure. Each complete n-type of T is an infinite set of formulas and has uncountably many subsets which themselves are types. The number of complete n-types of a theory T, meaning the cardinality of the set $S_n(T)$, is thus a measure of the semantic complexity of T. In general, for any complete theory T this cardinal number is not less than 1 and not greater than $2^{\aleph_0}$. One very interesting matter is the investigation of relations between the number of types of a theory and the number of their non-isomorphic models in different powers. For any consistent theory T and any $n > 0$ the cardinality of $S_n(T)$ is not greater than $\aleph_0$ times $\mu(T, \aleph_0)$. Thus if $\mu(T, \aleph_0)$ is at most countable, then also the cardinality of $S_n(T)$ is at most countable for all $n > 0$. And if for all $n > 0$ the cardinality of $S_n(T)$ is equal $2^{\aleph_0}$, then also $\mu(T, \aleph_0) = 2^{\aleph_0}$.

Still open is the following hypothesis formulated in 1963 by Vaught: for any consistent theory T, if $\mu(T, \aleph_0) < 2^{\aleph_0}$, then $\mu(T, \aleph_0) \leqslant \aleph_0$. Connected with this hypothesis is the theorem proved by Morley: for any consistent theory T, if $\mu(T, \aleph_0) < 2^{\aleph_0}$, then $\mu(T, \aleph_0) \leqslant \aleph_1$.

The investigations of spectra of theories, that is of the function $\mu(T, \kappa)$, can be divided into three cases: $\kappa = \aleph_0$, κ uncountable, and κ finite. In each of these cases different tools are used for characterizing spectra of theories. Uncountable spectra for countable theories have been investigated in details. There remain open problems for the values of $\mu(T, \aleph_0)$.

4.7.2 Stability

Different kinds of stability of theories are determined by the number of types (over sets of a given power) satisfiable in the models of theories.

Let T be a countable theory and κ an infinite cardinal number. Then there exist at most 2^κ pairwise non-isomorphic models of T of the cardinality κ. The function $\mu(T, \aleph_0)$ can thus reach its maximal value. However, some of the values are excluded: Vaught has proved that a complete theory cannot have exactly two non-isomorphic models.

Let T be a countable complete theory and κ an infinite cardinal number. We say that T is κ-stable if for any subset A of a domain of a model of T, if the cardinal number of A is not greater than κ, then the cardinal number of the set of all complete types over A also is not greater than κ. We say that a complete countable theory is:

1. *stable*, if it is κ-stable for arbitrarily large κ;

2. *superstable*, if it is κ-stable for sufficiently large κ;

3. *ω-stable*, if it is κ-stable for all infinite κ;

4. *strictly stable*, if it is stable, but not superstable;

5. *strictly superstable*, if it is superstable, but not ω-stable;

6. *nonstable*, if it is not stable.

Each theory belongs to one of the four classes: ω-stable, strictly superstable, strictly stable or nonstable. Stable theories are sometimes defined in terms of orderings definable in their models. We say that a theory T has the ordering property, if there exists a formula $\psi(\vec{x}, \vec{y})$ such that: $\mathfrak{A} \models \psi(\vec{x}, \vec{y})[\vec{a}_i, \vec{b}_j]$ if and only if $i < j$ for some model $\mathfrak{A}$ of T and sequences $(\vec{a}_i)_{i\in\omega}$ and $(\vec{b}_j)_{j\in\omega}$ of elements of $dom(\mathfrak{A})$.

One can prove that T is nonstable if and only if it has the ordering property. For instance, the theory of a linear ordering is nonstable. Furthermore, if T is not stable, then $\mu(T, \kappa) = 2^\kappa$ for all uncountable cardinals κ.

In model theory one uses the intuitive notions of tame and wild theories. A theory is tame, if it can be analyzed in terms of model theory. Theories which are not tame are called wild. Let us look at a few examples of theories with different degrees of stability (examples given in Hedman 2004):

1. The undecidable theory of arithmetic belongs to wild theories.

2. Examples of tame theories are: *o*-minimal theories and some theories described below.

3. Stable theories include superstable theories mentioned below and such strictly stable theories as, for example, the theory of countably many equivalence relations E_i nested in such a way that for all i:

   (a) $\forall x \forall y \ (E_{i+1}(x, y) \to E_i(x, y))$

   (b) each E_i-class contains countably many E_{i+1}-classes.

   This theory is stable, but it is not superstable.

4. Superstable theories, among which are the theories mentioned below, and such strictly superstable theories as, for example, the theory of countably many equivalence relations E_i such that each of them has exactly two equivalence classes and the intersection of E_i and E_j for $i \neq j$ has exactly four equivalence classes. This is an example of a superstable theory which is not ω-stable.

5. ω-*stable* theories which include all uncountably categorical theories and some non categorical theories, for example the theory of one equivalence relation with two infinite equivalence classes.

6. Uncountably categorical theories include all strongly minimal theories. Each uncountably categorical theory is related to some strongly minimal structure. Certain uncountably categorical theories which are not strongly minimal also belong to this level, for example the theory of integers with the successor function and one unary relation P such that: $\forall x \ (P(x) \equiv \neg P(s(x)))$.

7. Strongly minimal theories, for example the theory of integers with successor function, the theory of algebraically closed fields, or the theory of vector spaces.

All uncountably categorical theories are ω-stable, but countably categorical theories are dispersed among levels of the above hierarchy. There exist countably categorical theories which are strongly minimal (for example clique theory) or nonstable (theory of a random graph).

Part II

Mathematical aspects

Chapter 5

The axiom of completeness in geometry, algebra and analysis

At the very beginning a few explanations are in order. In this chapter we are going to discuss several interrelated issues since all of them deal with certain aspects of the continuity property, considered as a kind of completeness. There are also some important differences between them, at least from a methodological point of view.

The property of continuity may be intuitively ascribed to physical phenomena such as time, space, movement. In the realm of mathematics, continuity is also treated as a property which can be ascribed for example to spaces, orderings and functions. One should distinguish between continuity of an ordering and continuity of a function, although they can be expressed using similar mathematical tools.

Intuitively, the property of completeness understood as a kind of a maximality property corresponds to the fact that the universe under consideration does not contain "gaps". This intuitive explanation could be a little bit misleading. Similar confusion occurs if we intuitively characterize the completeness of a structure as based on the fact that it cannot be expanded to a larger one with exactly the same properties. However, both of the intuitive statements above can be given precise mathematical meaning, for instance in one or another form of the axiom of continuity.

Discussion concerning the structure of the continuum has a very long history, reaching back to antiquity, and there is a huge amount of literature on the subject. We are not going to report here on the history of that debate, because our main subject is that of extremal axioms and particularly – in this chapter – maximality axioms related primarily to characterizations of the Euclidean space and the set of real numbers. The reader interested in the history of the continuum debate may consult for example the classic book Boyer 2009, the extensive entry Bell 2005 in the Stanford Encyclopedia of Philosophy (as well as the bibliographical sources in it) or the concise booklet Buckley 2012.

The continuum is investigated in several areas of mathematics: geometry, algebra, analysis, topology, and set theory. The question to naturally arise is whether all these disciplines are tackling the same object. It seems reasonable to assume that particular mathematical domains consider continuum slightly differently, focusing on chosen aspects of it. A recent paper by Solomon Feferman contains a comparison of several approaches to the structure of continuum (Feferman 2009). In particular, Feferman distinguishes the following approaches:

1. *Euclid.* In Euclid's *Elements* it is assumed that any straight line (which, remember, is a *finite* object!) can be arbitrarily extended in a continuous way. The nature of this continuity is not further investigated.

2. *Hilbert.* The Hilbertian continuum presupposes the Archimedean axiom and axiom of completeness (or, in later versions, line completeness). Feferman observes that from the point of view of conceptual structuralism, Hilbertian continuum is a hybrid of geometrical and set-theoretical notions.

3. *Cantor.* The real numbers obtained in Cantor's construction form the completion of the ordered set of rational numbers with respect to fundamental sequences.

4. *Dedekind.* The family of all Dedekind cuts of the ordered set of all rational numbers is a representation of continuum which refers to the order properties of rationals. According to Feferman, Dedekind's and Cantor's approaches are hybrid, in the sense that they characterize continuum by geometric, arithmetic and set-theoretical notions.

5. *Full powerset* $\wp(\mathbb{N})$. The family of all subsets of the set of all natural numbers is a representation of the continuum without reference to geometric properties.

6. *Full binary tree.* The branches of this tree are all infinite $0 - 1$ sequences. As is well known, real numbers can be represented by such a tree. This representation of continuum is arithmetical in nature.

Feferman also discusses in brief other approaches to the problem of mathematical representation of continuum, including phenomenological, non-standard, foundational (intuitionistic and predicative), and physical.

5.1 Geometry

In the late 19th century (and at the beginning of the 20th century) many formal systems of geometry were proposed (for example by David Hilbert, Moritz Pasch, Giuseppe Veronese, Oswald Veblen, Edward Huntington, Mario Pieri, and Giuseppe Peano), and later in the 20th century influential systems were proposed, by, among others, Alfred Tarski and George Birkhoff. One of the goals of these systems was to improve the system of Euclid, in the sense that it was necessary to fill the gaps in several of Euclid's proofs. Another goal (at least in some of these systems) was to secure the uniqueness of such systems, that is, to prove that the proposed axioms had only one (up to isomorphism) interpretation.

5.1.1 Euclid

Euclid's *Elements* served as the main geometrical treatise for two thousand years. It contained not only geometrical investigations but also a theory of proportions which discussed magnitudes and their properties, and it is acknowledged as early Greek number theory.

Euclid's geometrical postulates are as follows (citing Heath's translation of the *Elements*, Heath 1968, 154–155):

> Let the following be postulated:
>
> 1. To draw a straight line from any point to any point.
> 2. To produce a finite straight line continuously in a straight line.
> 3. To describe a circle with any centre and distance.
> 4. That all right angles are equal to one another.
> 5. That, if a straight line falling on two straight lines make the interior angles on the same side less than two right angles,

> the two straight lines, if produced indefinitely, meet on that side on which are the angles less than the two right angles.

Thus, the postulates were supposed to warrant the possibilities of construction of geometric objects. It was also implicitly assumed that the constructions in question led to objects determined in a unique way.

Besides postulates, Euclid also introduced what he called *definitions*, which can be conceived of as intuitive explanations of the primitive terms as well as defined ones. He also has a group of assumptions which concern *common notions* related to operations on magnitudes.

As had been mentioned above, axiomatic systems of geometry developed in the 19th century were partly aimed at correcting Euclid's system. Some of the theorems given by Euclid were known not to follow from his postulates. For instance, the fact that two circles with centers at the distance of their radius will intersect at two points does not follow from the postulates – the existence of the points of intersection would require some form of continuity axiom which was not present in Euclid's system. From the modern point of view there are also further shortcomings in the system, for example: no notion of *betweenness*, which could be used for distinguishing the inside and outside of figures, or some unclarity concerning the meaning of arbitrary extendability of segments.

Moritz Pasch's *Vorlesungen über neuere Geometrie* was published in 1882 and was the first axiomatic system of geometry to propose essential corrections to Euclid's system (Pasch 1926). It was Pasch who introduced the term *primitive term* (Kernbegriff). He advocated the view that intuition in geometry should concern only the axioms and primitive terms. Among his primitive terms there were those of a *point* and a *segment*, but a *straight line* was a defined concept. Pasch is best remembered for his axiom saying, roughly speaking, that if a line, not passing through any vertex of a triangle, meets one side of the triangle then it meets another side. Pasch's treatise was very influential as the first modern axiomatic approach to geometry, which can be seen for example in the work of Hilbert discussed below.

5.1.2 Hilbert's *Grundlagen der Geometrie*

The first edition of Hilbert's *Grundlagen der Geometrie* was published in 1899 (Hilbert 1899). In 1900 Hilbert also published a small but very influential paper *Über den Zahlbegriff* (Hilbert 1900) in which he proposed an axiom system for real numbers. The text of *Grundlagen der Geome-*

trie originated from the course on the foundations of geometry which Hilbert presented in the academic year 1898-1899. The French translation of it was published very soon afterwards, and it was this text to which Hilbert added the axiom V.2 (axiom of completeness). The first English translation was provided by Edgar J. Townsend in 1902, and was authorized by Hilbert himself. Hilbert made some further revisions up to the 7th edition. The second English translation was published by Leo Unger in 1971, based on the 10th German edition.

The primitive terms of Hilbert's system are: point, line, plane, betweenness, lies on (containment), and congruence. The last three notions are relational.

The axioms of the system are divided into five groups, with the names corresponding to their content:

I *Incidence.* This group consists of eight axioms expressing the properties of the relation *lies on* as the relation between points and lines, lines and planes, and points and planes. The last axiom of this group is existential: it says that there exist at least four points not lying on the same plane.

II *Order.* Axioms of this group concern the relation of *betweenness.* In particular, they may be thought of as providing a precise meaning to the intuitive understanding of a straight line as an object "without width and depth". The last axiom of this group is Pasch's axiom formulated as follows: Let A, B, C be three points not lying in the same line and let a be a line lying in the plane ABC and not passing through any of the points A, B, C. Then, if the line a passes through a point of the segment AB, it will also pass through either a point of the segment BC or a point of the segment AC.

III *Congruence.* Axioms of this group concern the properties of the relation of *congruence.* This relation corresponds to *displacement* of geometrical objects preserving some of their properties. The axioms can be further used to compare magnitudes (length of segments, measure of angles).

IV *Parallels.* This is an axiom corresponding to the famous fifth axiom in Euclid's system, formulated in a form ascribed to Playfair: Let a be any line and A a point not on it. Then there is at most one line in the plane, determined by a and A, that passes through A and does not intersect a.

V *Continuity.* This group contains two axioms:

(1) *Axiom of Archimedes.* If AB and CD are any segments then there exists a number n such that n segments CD constructed contiguously from A, along the ray from A through B, will pass beyond the point B.

(2) *Axiom of line completeness.* An extension of a set of points on a line with its order and congruence relations that would preserve the relations existing among the original elements as well as the fundamental properties of line order and congruence that follows from Axioms I–III and from V-1 is impossible.

The formulation of the axiom of line completeness above comes from the 7th German edition of the text. In the English 1902 edition of the *Grundlagen der Geometrie* (as well as in the reprinted edition Hilbert 1950) the axiom of completeness had the following form:

> Remark. To the preceding five groups of axioms, we may add the following one, which, although not of a purely geometrical nature, merits particular attention from a theoretical point of view. It may be expressed in the following form:
>
> Axiom of Completeness. *(Vollständigkeit): To a system of points, straight lines, and planes, it is impossible to add other elements in such a manner that the system thus generalized shall form a new geometry obeying all of the five groups of axioms. In other words, the elements of geometry form a system which is not susceptible of extension, if we regard the five groups of axioms as valid.*
>
> This axiom gives us nothing directly concerning the existence of limiting points, or of the idea of convergence. Nevertheless, it enables us to demonstrate Bolzano's theorem by virtue of which, for all sets of points situated upon a straight line between two definite points of the same line, there exists necessarily a point of condensation, that is to say, a limiting point. From a theoretical point of view, the value of this axiom is that it leads indirectly to the introduction of limiting points, and, hence, renders it possible to establish a one-to-one correspondence between the points of a segment and the system of real numbers. However, in what is to follow, no use will be made of the "axiom of completeness". (Hilbert 1950, 15–16)

This remark explains the reason why the axiom of completeness was added to the main body of axioms. It is not used in the proofs within the system, but it is necessary for proving the consistency of the system:

it enables one to establish a $1-1$ correspondence between points lying on a straight line and the real numbers.

The Archimedean axiom can be formulated with explicit reference to the relations *lies on* and *betweenness*:

> V. Let A_1 be any point upon a straight line between the arbitrarily chosen points A and B. Take the points A_2, A_3, A_4,... so that A_1 lies between A and A_2, A_2 between A_1 and A_3, A_3 between A_2 and A_4 etc. Moreover, let the segments AA_1, A_1A_2, A_2A_3, A_3A_4, ... be equal to one another. Then, among this series of points, there always exists a certain point A_n such that B lies between A and A_n. (Hilbert 1950, 15)

The notion of *equality of segments* is involved in both of the above formulations of the axiom of Archimedes. As Hilbert himself writes, this notion can be defined either by the relation of congruence or by certain constructions (described later in his book) corresponding to the idea that "a segment is to be laid off from a point of a given straight line so that a new, definite segment is obtained 'equal' to it."

In order to show that his axiom system is consistent Hilbert explicitly gives an interpretation (a model) of it, presenting a domain Ω with suitable relations such that all the axioms are fulfilled in that domain. The domain consists of algebraic numbers generated by the number 1 by applying the arithmetical operations of addition, subtraction, multiplication, division as well as by an operation $\sqrt{1+\omega^2}$, where ω belongs to the considered domain. Then one obtains the following interpretation of the primitive terms:

1. Points are defined as pairs (x, y) of elements of the domain Ω.
2. The ratio $(u : v : w)$ of three elements of the domain Ω (where it is not the case that u and v are both equal zero) defines a straight line.
3. The fact that the point (x, y) lies on the straight line $(u : v : w)$ is expressed by the equation:

$$ux + vy + w = 0$$

Notice that the domain Ω is a countable set of real numbers. By using algebraic properties of these numbers, Hilbert shows that the axioms of his system hold in this interpretation, which means that the system is consistent (relative to the consistency of the system of real numbers).

Hilbert also proves the independence of each axiom from the rest of the body of axioms, by providing interpretations in which a given axiom is false while all the remaining axioms are true. In particular, he proves in this way the independence of the parallel axiom, thus distinguishing between Euclidean and non-Euclidean geometry.

The independence of the axiom of Archimedes is obtained by showing that the domain $\Omega(t)$ of all algebraic functions of t which may be obtained by arithmetical operations of addition, subtraction, multiplication, and division as well as by an operation $\sqrt{1+\omega^2}$, where ω represents any function from the domain $\Omega(t)$, satisfies all axioms of the system with the exception of the axiom of Archimedes. Thus it is an example of a *non-Archimedean geometry*.

Hilbert does not speak about the independence of the axiom V.2 in the early editions of the *Grundlagen der Geometrie*. Indeed, as the formulation of this extremal axiom is not presented in the object language of the theory, it does not seem possible to raise the question of its independence. The axiom V.2 is supposed to impose requirements on the interpretation of the system of geometry by expressing the property of its maximality. This is something different from the notions of completeness of mathematical theories discussed earlier in chapter 2. The following comment in Awodey and Reck 2002a best explains this issue:

> Hilbert's Axiom of Completeness asserts that (the whole space or) each line in space cannot be extended further – by adding additional points – while maintaining all the other axioms. It is worth being very precise and explicit here so as to prevent a common misinterpretation. Namely, the axiom does *not* say anything about the semantic, deductive or logical completeness of the system of axioms; *nor* does it say anything about categoricity, e.g. explicitly requiring the system of axioms to be categorical. It is true, of course, that the Axiom of Line Completeness together with the other axioms has as a consequence the categoricity of the whole system of axioms; and such categoricity has, in turn, as a consequence the semantic completeness of this system of axioms. Still, what the Axiom of Line Completeness itself mentions is points in geometric space, not formulas in the corresponding language. In other words, what it asserts is the 'completeness' (better: maximality) of the geometric space, not the completeness of the axiomatic system. This aspect comes out clearly if we reformulate Hilbert's axioms in formal logical terms. The Axiom of Line Completeness then shows itself to involve quantification over models of the axioms, not over sentences. (Awodey and Reck 2002a, 12)

Hilbert's Axiom of Completeness was critically discussed in Richard Baldus' paper *Zur Axiomatik der Geometrie. I. Über Hilberts Vollständigkeitsaxiom* (Baldus 1928). Baldus considers both versions of Hilbert's Axiom of Completeness: the one given above as well as the axiom of line completeness. He shows that the axiom of line completeness can be replaced by the axiom of continuity equivalent to it. His paper contains important results concerning the system of absolute geometry, which is to say a system like Hilbert's with the Archimedean axiom and the continuity axiom, but without the axiom of parallels. Baldus shows that:

> Theorem of completeness. It is not possible to add in thought any kind of things ("points", "lines", "planes") to an interpretation of the axioms of absolute geometry (...) so that the axioms again hold in the extended interpretation and that the axiomatic relations between the elements (points, lines, planes) of the original interpretation are preserved under the extension. (Baldus 1928, 324, citing the English translation in Schiemer 2010a, 138)

Thus the system of absolute geometry does have the Hilbertian property of maximality but that fact can be proved in metatheory and it is not necessary to assume it as an extra axiom of the theory itself.

Baldus shows also that in the case of absolute geometry the maximality property expressed above does not imply categoricity of the system. Absolute geometry can be extended either to Euclidean geometry or to hyperbolic geometry. Only after adding to absolute geometry either the Euclidean axiom of parallels or its hyperbolic alternative are categorical theories obtained.

It follows from Baldus' results that model maximality is not equivalent to categoricity. The latter implies the former, but not vice versa.

5.1.3 Tarski

Alfred Tarski devoted several of his works to geometrical investigations. He had some teaching practice of school geometry at the beginning of his academic career. He worked on formalizations of geometry for more than sixty years – from the nineteen-twenties to the nineteen-eighties. We present below his axioms for elementary geometry from the work Tarski 1959.

Tarski's system is formulated in the language of the first-order logic with identity. It is an elementary system; it uses quantification for individual variables only, where these variables range over points.

There are only three primitive terms of the system: *points* and the relations of *betweenness* and *congruence* (or *equidistance*). Points are elements of the domain.

Betweenness B is a triadic relation – the expression $Bxyz$ should be read: "y is between x and z", or, equivalently, y belongs to the segment xz. If $x = y$ or $y = z$, then the relation $Bxyz$ holds trivially.

Congruence $\equiv$ is a tetradic relation. The expression $xy \equiv uv$ should be read: "the segment xy is congruent with the segment uv", or, equivalently, "the length of the segment xy is the same as the length of the segment uv". In yet other words, $xy \equiv uv$ means that the distance from x to y is the same as the distance from u to v.

The axioms of the system are formulated with the use of primitive terms only (and with identity predicate) – no defined concept appears in the axioms. They can be grouped with respect to the primitive terms contained in them. Here are these groups (the axioms are understood to be universal closures of the formulas below):

Axioms of congruence

Reflexivity of congruence. $xy \equiv yx$.

Identity of congruence. $xy \equiv zz \to x = y$.

Transitivity of congruence. $(xy \equiv zu \wedge xy \equiv vw) \to zu \equiv vw$.

Axioms of betweenness

Identity of betweenness. $Bxyx \to x = y$.

Axiom of Pasch. $Bxuz \wedge Byvz \to \exists w(Buwy \wedge Bvwx)$.

Axiom schema of continuity. Let $\varphi(x)$ and $\psi(y)$ be first-order formulas such that: none of them contains any free occurrence of either a or b, $\varphi(x)$ does not contain any free occurrences of y and $\psi(y)$ does not contain any free occurrences of x. Then all the instances of the following schema are axioms of the system:

$$\exists a \forall x \forall y[(\varphi(x) \wedge \psi(y)) \to Baxy] \to$$
$$\to \exists b \forall x \forall y[(\varphi(x) \wedge \psi(y)) \to Bxby].$$

Lower dimension. $\exists a \exists b \exists c(\neg Babc \wedge \neg Bbca \wedge \neg Bcab)$.

Axioms of congruence and betweenness

Upper dimension.

$$(xu \equiv xv \wedge yu \equiv yv \wedge zu \equiv zv) \rightarrow (Bxyz \vee Byzx \vee Bzxy).$$

Axiom of Euclid.

$$(Bxuv \wedge Byuz \wedge x \neq y) \rightarrow \exists a \exists b(Bxya \wedge Bxzb \wedge Bavb).$$

Five segments.

$$(x \neq y \wedge Bxyz \wedge Bx'y'z' \wedge xy \equiv x'y' \wedge yz \equiv y'z' \wedge xu \equiv x'u' \wedge yu \equiv y'u') \rightarrow zu \equiv z'u'.$$

Segment construction. $\exists z(Bxyz \wedge yz \equiv ab)$.

Let us add a few comments concerning the intuitive meaning of these axioms and some consequences of them:

1. The basic objects in the system are points. Segments are determined by pairs of points. Straight lines are not represented as basic objects (unlike in Hilbert's approach); as the system is elementary, one cannot speak of straight lines as *sets* of points.

2. The affine aspects of Euclidean geometry are reflected by the properties of the relation of betweenness. In turn, the relation of congruence corresponds to the metric aspects of Euclidean geometry.

3. Formally, congruence is a tetradic relation between points, but one should think of it also as a binary relation between segments.

4. The axiom of Pasch expresses the idea that the diagonals of a quadrilateral should intersect at some point.

5. The axiom schema of continuity expresses the idea of Dedekind's cuts. If the formulas φ and ψ define sets of points on a ray with the endpoint a such that each point in the set defined by φ precedes (in the sense of the relation B) each point in the set defined by ψ, then there exists a point b on that ray separating these sets, following all the points in the set defined by φ and preceding all the points in the set defined by ψ.

6. The lower dimension axiom says that there exist three non-collinear points.

7. The upper dimension axiom says that three points equidistant from a given two distinct points are collinear.

8. The axiom of parallels can be stated in several forms in this system. Besides the formulation given above (which says that given any angle and a point in its interior there exists a line segment including this point and having endpoints on each side of the angle) one can express it also in the following forms:

   $$((Bxyw \wedge xy \equiv yw) \wedge (Bxuv \wedge xu \equiv uv) \wedge (Byuz \wedge yu \equiv zu)) \to \to yz \equiv vu.$$

   This is equivalent to the statement that the interior angles of any triangle sum up to two right angles.

   $$Bxyz \vee Byzx \vee Bzxy \vee \exists a(xa \equiv ya \wedge xa \equiv za).$$

   This is equivalent to the statement that for any given triangle there exists a circle circumscribed on it, in other words containing all vertices of this triangle.

9. The meaning of the five segments axiom can be illustrated by the following construction. Given two triangles xuz and $x'u'z'$ let us draw segments yu and $y'u'$ connecting the vertex of each triangle (that is, u and u', respectively) with the point on the side opposite to this vertex (so y and y', respectively). Then we get two triangles with five specified segments. The axiom says that if four of them are pairwise congruent, then also the remaining pair consists of congruent segments.

10. The axiom of segment construction says that for any point y it is possible to draw a segment congruent to any given segment ab.

Tarski's system is very elegant from the logical point of view. Among its most important metalogical properties are the following:

1. The system is consistent: there is no sentence ψ from the language of the system such that both ψ and $\neg\psi$ are provable from the axioms.

2. The system is deductively complete: for any sentence ψ in the language of the system, either ψ or $\neg\psi$ is provable from the axioms.

3. The system is decidable: there exists an effective procedure deciding whether any sentence ψ in the language of the system is a theorem of the system. This is obtained from the decidability result for real closed fields credited to Tarski and based on the elimination of quantifiers.

Tarski's system is described in full detail in the monograph Schwabhäuser, Szmielew and Tarski 1983.

5.2 Algebra and analysis

The theory of magnitudes has its roots in the Greek theory of proportions credited to Eudoxos and presented in book V of Euclid's *Elements*. But it took several hundred years to obtain a precise mathematical theory of the arithmetic continuum. Important steps toward this achievement were taken by such authors as (listed in alphabetical order): Georg Cantor, Richard Dedekind, Eduard Heine, David Hilbert, Otto Hölder, Edward Huntington, Charles Méray, Heinrich Weber, Karl Weierstrass and others. There are numerous works devoted to the history of these issues. In our opinion, one of the best expositions of the topic in question is the book Błaszczyk 2007.

We have already mentioned in chapter 2 the characterizations of the real numbers credited to Richard Dedekind, David Hilbert and Edward Huntington.

5.2.1 Constructions and representations of real numbers

The arithmetic continuum can be characterized either in a genetic way or in an axiomatic way. The first is based on the extension of an already known number universe to a new one, supplied by some new operations which were inaccessible in full generality in the initial universe. In this way one can go from natural numbers to the integers, then to the rational numbers, the real numbers and complex numbers. One should probably add that this logical path does not correspond exactly to the historical one, and for example (positive) rational numbers were familiarized before (negative) integers.

The second way starts from a list of primitive terms and axioms characterizing them, and can be applied to any sort of mathematical objects. Up to the 19th century this approach was presented only in Euclid's geometry. Since then, the axiomatic method has been the standard approach.

There are several *constructions* of the real numbers. The most commonly known are those given by Georg Cantor and Richard Dedekind. On the other hand, the approach to the real numbers popular in mathematical education is based on ideas like the one in the Hoborski system, using *decimal expansions*. There is also an approach based on *continued fractions*, which has many theoretical virtues but is not popular in contemporary mathematical education. Let us recall in brief the constructions mentioned above.

Cantor's construction

Consider all sequences $f : \mathbb{N} \to \mathbb{Q}$. We say that a sequence $(f(n))$ is a *Cauchy sequence*, if:

$$\forall \varepsilon \in \mathbb{Q}^+ \exists k \in \mathbb{N} \forall m \in \mathbb{N} \forall n \in \mathbb{N} (m \geqslant k \wedge n \geqslant k \to |f(n) - f(m)| < \varepsilon).$$

Let $\mathbb{F}$ be the set of all Cauchy sequences. Define a relation $\sim$ on $\mathbb{F}$ by: $\{f(n)\} \sim \{g(n)\}$ if and only if $\lim_{n\to\infty}(f(n)-g(n)) = 0$. It is an equivalence relation. Then the set of real numbers (in Cantor's sense) is defined to be the quotient set $\mathbb{F}/\sim$. One defines arithmetical operations on the real numbers defined in that way as well as their ordering. The fundamental role in this approach is played by the concept of limit.

Dedekind's construction

Let us recall the construction presented in chapter 2. Consider the ordered set of rational numbers $(\mathbb{Q}, <)$. By a *cut* we mean any pair (A, B) of disjoint subsets of $\mathbb{Q}$ such that $A \cup B = \mathbb{Q}$ and for all $a \in A$ and for all $b \in B$, $a < b$. There are only three possible types of cut in $(\mathbb{Q}, <)$:

1. There is the greatest element in A and there is no smallest element in B.

2. There is no greatest element in A and there is the smallest element in B.

3. There is no greatest element in A and there is no smallest element in B.

The fourth possibility (the greatest element in A and the smallest element in B) is of course excluded, because the ordering of the rational numbers is dense. In the first two of the above three cases the cut (A, B) corresponds to a rational number (the greatest element of A or the smallest element of B, respectively). In the third case we say that the cut (A, B) determines a *gap* (in $(\mathbb{Q}, <)$). For example, the cut

$$(\{x \in \mathbb{Q} : x^2 < 2\}, \{x \in \mathbb{Q} : x^2 > 2\})$$

determines a gap. Dedekind proposes to *define* the set of real numbers as the set of all cuts of $(\mathbb{Q}, <)$. Then the gaps correspond to irrational numbers. One defines arithmetical operations on the real numbers defined in that way as well as their ordering. In this approach the fundamental role is played by order properties.

Hoborski's construction

By a (positive) Hoborski real number we mean any sequence $\{a_n\}$ such that:

1. $a_0 \in \mathbb{N}$
2. $a_{n+1} = a_n + \frac{c_{n+1}}{10^n}$, where $c_{n+1} \in \{0, 1, 2, \ldots, 9\}$
3. $\neg\exists k \in \mathbb{N} \forall m > k(a_m = 9)$.

Thus each positive Hoborski real number can be represented in a decimal expansion written as usual as $a_0, a_1a_2a_3\ldots$, where a_0 is a natural number and all a_i (for $i > 0$) are the digits belonging to the set $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$. The third condition above means that we do not consider infinite expansions with an infinite "tail" consisting entirely of the digit 9.

Continued fractions

By a representation of the real number x in form of a (simple) continued fraction we mean the construction:

$$x = n_0 + \cfrac{1}{n_1 + \cfrac{1}{n_2 + \cfrac{1}{n_3 + \cfrac{1}{n_4 + \ddots}}}}.$$

Here $n_0, n_1, n_2, \ldots$ are natural numbers. Instead of such multi-level notation one often uses the linear notation: $x = [n_0; n_1, n_2, n_3, \ldots]$.

Continued fractions were investigated very early in the history of mathematics. They still are of great importance in many areas of mathematics, though they are rarely discussed in school. One can prove that each rational number has a finite representation as a continued fraction, while irrational numbers have infinite such representations (quadratic real irrationals and only quadratic real irrationals have periodic continued fractions representing them). The representation of real numbers by continued fractions is related to the Euclidean algorithm. It could be worth remarking that the representations in question reveal some regularities which are not visible in decimal expansions, for example:

1. The golden ratio has the decimal representation

   $$\varphi = \frac{1+\sqrt{5}}{2} \approx 1.61803398874989484 82\ldots$$

   showing no regularity. Its continued fraction is very regular:

   $$\frac{1+\sqrt{5}}{2} = [1; 1, 1, 1, 1, 1, \ldots]$$

   $$\varphi = 1 + \cfrac{1}{1 + \cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{1+\ddots}}}}$$

2. The decimal expansion of $\sqrt{2}$ begins with:

   $$\sqrt{2} \approx 1.4142135623730950488016887242096980\ldots$$

   The continued fraction corresponding to this number is very regular: $\sqrt{2} = [1; 2, 2, 2, 2, \ldots] = [1; \overline{2}]$. This can be easily seen from the following:

   $\sqrt{2} - 1 = \frac{(\sqrt{2}-1)\cdot(\sqrt{2}+1)}{\sqrt{2}+1} = \frac{2-1}{\sqrt{2}+1} = \frac{1}{\sqrt{2}+1}$

   $\sqrt{2} = 1 + \frac{1}{\sqrt{2}+1} = 1 + \frac{1}{1+\sqrt{2}} = 1 + \frac{1}{1+1+\frac{1}{\sqrt{2}+1}} = 1 + \frac{1}{2+\frac{1}{\sqrt{2}+1}} = 1 + \frac{1}{2+\frac{1}{2+\frac{1}{\sqrt{2}+1}}}$.

   Continuing this procedure we obtain the representation given above.

3. The irrational numbers π and e also have regular (generalized) continued fraction representations, e.g. the following ones:

$$\pi = \cfrac{4}{1 + \cfrac{1^2}{2 + \cfrac{3^2}{2 + \cfrac{5^2}{2 + \cfrac{7^2}{2 + \ddots}}}}} = \cfrac{4}{1 + \cfrac{1^2}{3 + \cfrac{2^2}{5 + \cfrac{3^2}{7 + \cfrac{4^2}{9 + \ddots}}}}} = 3 + \cfrac{1^2}{6 + \cfrac{3^2}{6 + \cfrac{5^2}{6 + \cfrac{7^2}{6 + \ddots}}}}$$

$$e = 2 + \cfrac{2}{2 + \cfrac{3}{3 + \cfrac{4}{4 + \cfrac{5}{5 + \ddots}}}}$$

5.2.2 Tarski's axiomatization

Two axiomatic systems for real numbers were proposed by Alfred Tarski. We shall recall one of them here, one which is very elegant from the logical point of view (Tarski 1994). The system has four primitive terms: $\mathbf{R}$ (the set of real numbers), $<$ (the relation of ordering), $+$ (the operation of addition) and 1 (a constant). There are eight axioms:

1. $\forall x \forall y (x < y \rightarrow \neg y < x)$
2. $\forall x \forall z (x < z \rightarrow \exists y (x < y \wedge y < z))$
3. $\forall X \subseteq \mathbf{R} \forall X \subseteq \mathbf{R} (\forall x \in X \forall y \in Y x < y \rightarrow \exists z (\forall x \in X \forall y \in Y (z \neq x \wedge z \neq y \rightarrow x < z \wedge z < y)))$
4. $\forall x \forall y \forall z (x + (y + z) = (x + z) + y)$
5. $\forall x \forall y \exists z (x + z = y)$
6. $\forall x \forall y \forall z \forall w (x + y < z + w \rightarrow x < z \vee y < w)$
7. $1 \in \mathbf{R}$
8. $1 < 1 + 1$

As one can see, this system is a second-order system: axiom 3 involves second-order quantifiers binding variables representing subsets of real numbers. Some of the most important properties of this system are as follows:

1. It can be proved that the relation $<$ is a linear ordering.
2. Axioms 4 and 5 imply that $(\mathbf{B}, +)$ is an abelian group.
3. The ordering $<$ is Dedekind-complete: if one set of reals precedes another set of reals, then there exists at least one real number separating the two sets.
4. Tarski proved that the above axioms are sufficient to define uniquely the operation of multiplication on $\mathbf{R}$.
5. $\mathbf{R}$ is thus a complete ordered field under addition and multiplication.
6. Tarski also proved that the above axioms are independent.

5.3 Axiom of continuity and its equivalents

The arithmetical continuum is considered the system of real numbers with the standard arithmetical operations, the natural ordering and topology. Its most salient property is that of continuity described by the completeness condition. The latter is also called the axiom of continuity.

The axiom of continuity can be formulated in several (mutually equivalent) ways. We add to the axioms of an ordered field $(R, +, \cdot, 0, 1, <)$ one of the following versions of this axiom (cf. Błaszczyk 2007, 306):

1. For any cut (A, B) in $(R, <)$ either in A there exists the greatest element or in B there exists the smallest element.
2. For any non-empty and bounded from above set $A \subseteq R$ there exists in R its least upper bound.
3. For any infinite and bounded set $A \subseteq R$ there exists in R its accumulation point (in order topology).
4. $(R, +, \cdot, 0, 1, <)$ is an Archimedean field and for any sequence $(a_n) \subseteq R$ there exists $a \in R$ such that $\lim_{n \to \infty} a_n = a$.
5. $(R, +, \cdot, 0, 1, <)$ is an Archimedean field and for every descending sequence of closed intervals (A_n) we have $\bigcap_n A_n \neq \emptyset$.

In such a setting, the axiom of continuity applies to Archimedean structures. A slightly separate topic is continuity considered in the presence of infinitesimals; see the results obtained by Ehrlich discussed in brief in one of the sections below.

One should remember about relations between several versions of completeness in connection to the Archimedean axiom. Let us first recall the necessary definitions (here in the case of real numbers; more general concepts are also considered):

1. Cauchy completeness is the statement that every Cauchy sequence converges.
2. Dedekind completeness is the property that there exist no gaps in the ordering of the real numbers – the ordering of the real numbers (understood as Dedekind's cuts of the ordered set of rational numbers) is a continuous ordering.
3. For an ordered field, the Archimedean axiom may be stated for example in the following ways:
   (a) For any element x of the field there exists a natural number n such that $x < n$.
   (b) For any element $\varepsilon > 0$ of the field there exists a natural number n such that $\frac{1}{n} < \varepsilon$.
4. Observe that the above formulations of the Archimedean axiom express the idea that there are no infinitely large or infinitely small elements in the field.

Then there are the following dependencies between these properties:

1. In an ordered field Cauchy completeness is equivalent to Dedekind completeness together with the Archimedean axiom.
2. There are non-Archimedean fields which are ordered and Cauchy complete.

Cauchy completeness can be generalized to a notion of completeness applicable to metric topological spaces. Also Dedekind completeness can be generalized, most notably to linearly ordered structures of several sorts.

Several theorems of analysis are equivalent to the continuity axiom, for instance:

1. *Monotone convergence theorem.* Every nondecreasing, bounded sequence of real numbers converges.

2. *Bolzano–Weierstrass theorem.* Every bounded sequence of real numbers contains a convergent subsequence.

3. *The intermediate value theorem.* If f is a continuous function defined on the interval $[a, b]$, then it takes any value between $f(a)$ and $f(b)$ at some point within the interval. In particular, if $f(a)$ is positive and $f(b)$ is negative (or vice versa), then there exists $c \in [a, b]$ such that $f(c) = 0$.

The axiom of continuity is necessary for obtaining precise definitions of some operations on real numbers, e.g. the operation of exponentiation (raising a real number to a real power). This axiom is also essential in the proofs of many theorems in analysis, for example:

1. *Extreme value theorem.* Let $f : [a, b] \to \mathbb{R}$ be a continuous function and let $M = \sup f[a, b]$. Then M is finite and there exists $c \in [a, b]$ such that $f(c) = M$.

2. *Heine–Borel theorem.* If $\{U_\alpha\}$ is any family of open sets which covers the closed interval $[a, b]$, then there exists a finite subfamily of this family which covers $[a, b]$ as well.

5.4 Generalizations and isomorphism theorems

Abstract algebra deals with arbitrary structures – thus it might seem that the notion of extremal axiom is of no use there. However, certain algebraic structures play a special role in the sense that they are fundamental in many fields of mathematics and they used to serve as intended models of the theories developed later. This concerns primarily the structure of real numbers $\mathbb{R}$, as a completely ordered Archimedean field and the structure of complex numbers $\mathbb{C}$ as an algebraically closed field, though of course not ordered.

5.4.1 Real, complex and hypercomplex structures

Among the most important properties of the structure $\mathbb{R}$ (as an algebraic structure endowed with ordering and topology) one may list:

1. $\mathbb{R}$ is the only completely ordered field (up to isomorphism).
2. $\mathbb{R}$ is not algebraically closed.
3. $\mathbb{R}$ is the maximal Archimedean field.

Among the most important properties of the structure $\mathbb{C}$ (as an algebraic structure endowed with topology) one may list:

1. $\mathbb{C}$ is algebraically closed.
2. $\mathbb{C}$ has characteristic 0.
3. The transcendence degree of $\mathbb{C}$ over $\mathbb{Q}$ equals continuum.
4. One can show that the above three properties characterize the structure $\mathbb{C}$ uniquely (up to isomorphism).

The structures $\mathbb{R}$ and $\mathbb{C}$ are thus very special and they possess many useful properties (from the point of view of applications): arithmetical, algebraical, topological and connected with ordering. There is a question that arises naturally here concerning generalizations of these structures. For instance, could we obtain structures with nice arithmetical and algebraic properties in which addition and multiplication of triples, quadruples, n-tuples of real numbers (or complex numbers) are definable and have useful properties? One may think intuitively about real numbers as one-dimensional and of complex numbers as two-dimensional and then ask about the possibility of three-dimensional, and four-dimensional numbers, and so on, satisfying suitable arithmetical and algebraic properties. The solution to this problem is positive only in some special cases and with some restrictions. Let us mention only one example: that of quaternions.

The system of *quaternions* was introduced by William Rowan Hamilton in 1843. Initially he wanted to develop an arithmetical structure for n-tuples of real numbers, in particular for triples of real numbers, corresponding to points in three-dimensional space (similarly to complex numbers which can be represented by points on the plane). He had already succeeded in the case of arithmetic operations for pairs of real numbers:

$$(a_1, b_1) + (a_2, b_2) = (a_1 + a_2, b_1 + b_2)$$

$$(a_1, b_1) \cdot (a_2, b_2) = (a_1 a_2 - b_1 b_2, a_1 b_2 + b_1 a_2)$$

These rules are exactly the rules of addition and multiplication, respectively, for the complex numbers of the form $a + bi$, understood as the ordered pairs (a, b). The absolute value is then defined as $|(a, b)| = \sqrt{a^2 + b^2}$ and it has the multiplicative property:

$$|(a_1, b_1)| \cdot |(a_2, b_2)| = |(a_1 a_2 - b_1 b_2, a_1 b_2 + b_1 a_2)|.$$

The rules for addition of *triples* of real numbers can be defined in an analogous way, however it is not possible to define multiplication of triples in such a way that the usual algebraic laws hold and the absolute value is multiplicative. The argument is very simple. Suppose that it were possible to define multiplication for triples of real numbers $x_1 = (a_1, b_1, c_1)$ and $x_2 = (a_2, b_2, c_2)$ in such a way that $|x_1| \cdot |x_2| = |x_1 x_2|$, where $|x_1| = \sqrt{a_1^2 + b_1^2 + c_1^2}$ and $|x_2| = \sqrt{a_2^2 + b_2^2 + c_2^2}$. Then:

$$|x_1|^2 \cdot |x_2|^2 = |x_1 x_2|^2,$$

which implies that $(a_1^2 + b_1^2 + c_1^2)(a_2^2 + b_2^2 + c_2^2) = a^2 + b^2 + c^2$ for some numbers a, b and c which should be obtained by arithmetical operations on the numbers a_1, b_1, c_1 and a_2, b_2, c_2. But this is not possible in general, because for instance

$$(1^2 + 1^2 + 1^2)(0^2 + 1^2 + 2^2) = 15,$$

and 15 cannot be represented as the sum $a^2 + b^2 + c^2$ for any integers $a, b, c < 15$. There are also other arguments which show the impossibility of multiplication of triples satisfying the multiplicative property of norms.

However, it is possible to define such multiplication for quadruples of real numbers, which gives rise to the algebra $\mathbb{H}$ of quaternions discovered by Hamilton (Hamilton 1853).

We may think of $\mathbb{H}$ as a four-dimensional vector space over the field of real numbers. The elements of the base of this vector space are usually denoted by 1, i, j, k.

The addition of quaternions and multiplication by scalars is defined as is usually done in vector spaces, and their multiplication is fully determined by the equalities: $i^2 = j^2 = k^2 = ijk = -1$. It can be also represented by the following table:

	1	i	j	k
1	1	i	j	k
i	i	-1	k	$-j$
j	j	$-k$	-1	i
k	k	j	$-i$	-1

It can be seen from this table that multiplication of quaternions is not commutative. The structure $\mathbb{H}$ satisfies all axioms for fields with the exception of the commutativity law for multiplication. Quaternions are usually denoted by (a, bi, cj, dk) or $(a + bi + cj + dk)$, where a, b, c, d are real numbers. By a conjugate of the quaternion $q = a + bi + cj + dk$ we mean the quaternion $q^* = a - bi - cj - dk$. The norm of the quaternion $q = a + bi + cj + dk$ is defined to be $||q|| = \sqrt{qq^*}$. Hence

$$||a + bi + cj + dk|| = \sqrt{a^2 + b^2 + c^2 + d^2}$$

and it follows that the norm is multiplicative: $||pq|| = ||p|| \cdot ||q||$. One can introduce a distance determined by this norm: $d(p, q) = ||p - q||$. Unit quaternions are those with norm equal 1. The unit quaternions form the three-dimensional sphere S^3 in the four dimensional Euclidean space. The algebra of quaternions is very useful as a mathematical representation of rotations in three-dimensional space – it serves as an alternative for the known Euler's description of rotations. The cross-product of vectors taught in schools is related to the multiplication of quaternions.

It is also possible to define multiplication for octaves of real numbers satisfying reasonable algebraic conditions, however in this case the multiplication is not only non-commutative but also non-associative. The algebra in question is that of *octonions*. The octonions $\mathbb{O}$ were discovered in 1843 independently by John T. Graves and Arthur Cayley. Recently, both quaternions and octonions have found many interesting applications in theoretical physics. There exist many further types of hypercomplex numbers, including sedenions, tessarines, dual numbers, double numbers, dual quaternions, and hyperbolic quaternions.

5.4.2 Isomorphism theorems

Several *isomorphism theorems* in algebra characterize the most fundamental structures up to isomorphism. Such theorems are thus counterparts, in a sense, of the extremal axioms in algebra, for example:

1. *Frobenius Theorem.* Each associative algebra with division over $\mathbb{R}$ is isomorphic either with $\mathbb{R}$, or $\mathbb{C}$, or $\mathbb{H}$.

2. *Hurwitz Theorem.* Any normed algebra with division is isomorphic either with $\mathbb{R}$, or $\mathbb{C}$, or $\mathbb{H}$ or $\mathbb{O}$.

3. *Ostrowski Theorem.* Any field complete with respect to an Archimedean norm is isomorphic with either $\mathbb{R}$ or $\mathbb{C}$ and the norm is equivalent to the usual norm determined by the absolute value.

4. *Pontriagin Theorem.* Any connected locally compact topological field is isomorphic with either $\mathbb{R}$, or $\mathbb{C}$ or $\mathbb{H}$.

There are structural distinctions between the structures $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$ and $\mathbb{O}$ mentioned in the above theorems. For instance $\mathbb{R}$ is an ordered field and $\mathbb{C}$ cannot be ordered (in such a way that the ordering would be compatible with arithmetic operations), multiplication in $\mathbb{R}$ and $\mathbb{C}$ is commutative but it is not commutative in $\mathbb{H}$ and $\mathbb{O}$ (in $\mathbb{O}$ it is not even associative), and so on. One may informally say that along with more generality of structure one looses more and more properties which appear as intuitive features of "more domesticated" arithmetical operations.

5.4.3 Extended reals

Still other generalizations (or rather extensions) of the structure $\mathbb{R}$ have been considered and applied in many areas of mathematics, for instance:

Projectively extended real line $\hat{\mathbb{R}}$. This structure is obtained from $\mathbb{R}$ by adding to it the element ∞ (point at infinity). This point at infinity may be thought of as the limit of every sequence of real numbers whose absolute values are increasing and unbounded. Arithmetical operations in $\hat{\mathbb{R}}$ can be defined, though with some restrictions. For instance, $a-\infty = \infty - a = \infty$ for $a \neq \infty$, $\frac{a}{0} = a \cdot \infty = \infty$ for $a \neq \infty$. Some expressions remain undefined, for example $\infty - \infty$, $0 \cdot \infty$, $\frac{\infty}{\infty}$, $\frac{0}{0}$. The structure $\hat{\mathbb{R}}$ is homeomorphic to a circle. Obviously, the ordering of $\mathbb{R}$ cannot be extended to a total (linear) ordering of $\hat{\mathbb{R}}$. It is possible to talk about a cyclic ordering of $\hat{\mathbb{R}}$, understood as a ternary relation.

Affinely extended real number line $\overline{\mathbb{R}}$. This structure is obtained from $\mathbb{R}$ by adding to it two elements: $+\infty$ (positive infinity, often written simply as ∞) and $-\infty$ (negative infinity). The ordering of $\mathbb{R}$ is extended to the ordering on $\overline{\mathbb{R}}$ by the conditions that $+\infty$ is the biggest element and $-\infty$ is the smallest element. Then we can make use of order topology and, in particular, talk about neighbourhoods of the points $+\infty$ and $-\infty$. The arithmetical operations involving infinities can be defined with some restrictions. For instance, $a \cdot (+\infty) = +\infty$ for $a > 0$, $|\frac{a}{0}| = +\infty$ for

$a \neq 0$, $\frac{a}{\pm\infty} = 0$ for $a \neq \pm\infty$, but for example $\frac{0}{0}$, $+\infty + -\infty$, $0 \cdot (+\infty)$ remain undefined. Thus, $\overline{\mathbb{R}}$ is not a field, though for defined expressions all the arithmetic laws hold. Some functions can be extended to $\overline{\mathbb{R}}$ which simplifies formulations of many theorems.

Long line. This is a topological space which, intuitively speaking, is "longer" than the real line. As everyone remembers, the real line consists of only countably many half-open intervals $[0, 1)$ laid end-to-end and the long line consists of uncountably many such intervals. Formally, the closed long ray is defined to be the cartesian product $\omega_1 \times [0, 1)$ (that is, the Cartesian product of the first uncountable ordinal by the half-open interval $[0, 1)$). It is equipped with order topology (determined by the lexicographical order on $\omega_1 \times [0, 1)$). The open long ray is obtained by removing the first element from the closed long ray. The reversed long ray is the long ray with its order reversed. The long line is obtained by joining the reversed long ray with the long ray (we omit some technicalities) and is thus equipped with a total order.

In the generalizations mentioned above the absolute value played the fundamental role (as a norm and thus characterizing the concept of *closeness*). There exist yet other possibilities of extending the structure $\mathbb{Q}$ of rational numbers using a different norm instead. These extensions are systems of p-adic numbers, where p is any prime number. In these systems closeness is related to divisibility by powers of p: two p-adic numbers are considered to be close when their difference is divisible by a high power of p and the higher this power, the closer they are. Recently p-adic numbers have found numerous applications in many domains of mathematics.

5.4.4 Real closed fields

We already know that the completely ordered field $\mathbb{R}$ of real numbers is unique up to isomorphism. The isomorphism theorems recalled above characterize also other algebraic structures up to isomorphism. One can also ask about a weaker type of indistinguishability, namely that based on elementary equivalence. We remember from chapter 4 that two structures are elementarily equivalent (in the first-order language) if they satisfy exactly the same sentences from that language. Such structures are thus semantically indistinguishable, elementary equivalence is an equivalence relation. If two structures are isomorphic, then they are elementarily equivalent, but not vice versa. Hence classes of elementary equivalence may in general contain many classes of isomorphisms.

The structure $\mathbb{R}$ is not characterized uniquely with respect to elementary equivalence. This means that there are ordered fields which satisfy the same first-order sentences as $\mathbb{R}$. These are exactly *real closed fields* and they include, besides $\mathbb{R}$, also the hyperreal field (see below) and the field of real algebraic numbers.

We recall that a field $\mathbb{F}$ is *algebraically closed* if any polynomial with coefficients in the field has a root in this field (equivalently, this means that $\mathbb{F}$ has no bigger algebraic extensions). A field $\mathbb{F}$ is an algebraic closure of a field $\mathbb{E}$ if it is algebraic over $\mathbb{E}$ and is algebraically closed.

We also recall (see the previous chapter) that by a *formally real field* we mean any field $\mathbb{F}$ such that for any elements $x_1, \ldots, x_n \in \mathbb{F}$, if

$$x_1^2 + \ldots + x_n^2 = 0,$$

then all elements $x_1, \ldots, x_n$ are equal to 0. On the basis of the Artin-Schreier theorem a field can be ordered (meaning endowed with a linear order compatible with the field operations) if and only if it is formally real.

There are several possibilities for defining a *real closed field*. For instance, each of the following conditions may serve as a definition of a field $\mathbb{F}$ as real closed:

1. $\mathbb{F}$ is a formally real field such that every polynomial of odd degree with coefficients in $\mathbb{F}$ has at least one root in $\mathbb{F}$ and for every element a of $\mathbb{F}$ there exists b in $\mathbb{F}$ such that either $a = b^2$ or $a = -b^2$.

2. There exists a linear ordering in $\mathbb{F}$ which is compatible with the field operations (which means that $\mathbb{F}$ is an ordered field) and each positive element of $\mathbb{F}$ has a square root in $\mathbb{F}$ and any polynomial of odd degree with coefficients in $\mathbb{F}$ has at least one root in $\mathbb{F}$.

3. There exists a linear ordering of $\mathbb{F}$ which cannot be extended to any proper algebraic extension of $\mathbb{F}$.

4. $\mathbb{F}$ is a formally real field such that no proper algebraic extension of F is formally real.

5. There exists a linear ordering of $\mathbb{F}$ such that the intermediate value theorem holds for all polynomials over $\mathbb{F}$.

6. $\mathbb{F}$ is elementarily equivalent to the field $\mathbb{R}$.

The last condition is semantical, all the remaining conditions have algebraic formulations. All conditions involve the notion of ordering (compatible with field operations).

Artin and Schreier proved that any ordered field has some special algebraic extension $\mathbb{K}$ (called its *real closure*) such that $\mathbb{K}$ is real closed and its ordering is the unique extension of the ordering of $\mathbb{F}$ (Artin and Schreier 1927).

The theory of real closed fields can be formulated as a first-order theory; one adds to the axioms of an ordered field the axiom stating that every positive element of the field has a root and an axiom schema saying that every polynomial of odd degree has at least one root in the field. This theory has some nice metalogical properties, as shown by Tarski (Tarski 1951). It admits elimination of quantifiers and hence is complete and decidable.

Examples of real closed fields include algebraic, computable, definable, real and hyperreal numbers. Observe that real closed fields may be Archimedean (for example the field of real numbers) as well as non-Archimedean (see the field of hyperreal numbers described below).

The real closed fields can be characterized in terms of certain *invariants*: cardinality, cofinality and weight. We say that a set X contained in an ordered set Y ordered by $<$ is *cofinal* in Y, if for any element $y \in Y$ there exists an element $x \in X$ such that $y < x$, which is equivalent to saying that X contains a sequence unbounded in Y. The *cofinality* of a set Y is the cardinality of the smallest set X cofinal in Y. Since the natural numbers form a set cofinal in the set of real numbers, the latter has cofinality $\aleph_0$.

By the weight of a set Y we understand the minimum cardinality of a dense subset of Y. The weight of the set of real numbers equals $\aleph_0$ (the rational numbers are dense in the set of real numbers).

We say that an ordered field $\mathbb{F}$ has the η_α property, if for any sets A and B of cardinality $\aleph_\alpha$ such that every element of A is less than every element of B there exists an element x in $\mathbb{F}$ such that x is larger than every element of A and smaller than every element of B. The notion of sets with η_α property was introduced by Felix Hausdorff.

For all α there exist real closed fields of cardinality $\aleph_\alpha$ with the η_α property. This property is related to the notion of a *saturated model*, which was discussed in chapter 4. It can be proved that two real closed fields have η_α property if and only if they are $\aleph_\alpha$-saturated. Furthermore, any real closed fields of cardinality $\aleph_\alpha$ and with η_α property are order isomorphic.

5.5 Degrees of infinity, pantachies and gaps

Georg Cantor proved that all countable dense orderings without endpoints are isomorphic. Hence the order type η of rational numbers is unique, up to isomorphisms. Density condition in the case of countable sets is, so to speak, the smallest possible density property. One can reasonably ask about analogical density properties in the case of uncountable sets, sets of arbitrarily large cardinality.

Paul Du Bois-Reymond was interested in the rates of growth of real valued functions. He attempted to characterize "degrees of infinity" corresponding to the rates of growth of such functions (Du Bois-Reymond 1877, 1882, cf. also Hardy 1954). Could we reasonably speak about something like degrees of divergence of real functions which grow faster and faster when their arguments grow? Is it possible to somehow define a cut between convergence and divergence?

Paul Du Bois-Reymond introduced the term *pantachie* for a system $(G, \prec)$, where:

1. G is the set of all positive real functions, meaning functions from $\mathbb{R}$ to $(0, \infty)$.
2. $f \prec g$ if and only if $\lim\limits_{x \to +\infty} \frac{f(x)}{g(x)} = 0$
3. $f \succ g$ if and only if $\lim\limits_{x \to +\infty} \frac{f(x)}{g(x)} = +\infty$
4. $f \sim g$ if and only if $\lim\limits_{x \to +\infty} \frac{f(x)}{g(x)} \in (0, +\infty)$.

Du Bois-Reymond asserted that if $f \prec g$, then g represented larger infinity than f. He proved in 1873 the following theorem:

> For any sequence of $(f_n(x))$ of elements of G which is strictly increasing with respect to $\prec$ there exists a function f in G which $\prec$-dominates all functions f_n, i.e. $f_n(x) \prec f(x)$ for all n.

The proof of this theorem contains the first application of the diagonal method about 15 years before Cantor's famous proof using this method.

Du Bois-Reymond believed his *pantachie* to be an appropriate scala for the ordering of infinities represented by strictly monotonic unbounded real functions. He claimed that $\mathcal{P} = \{[f] : f \in G\}$ (where $[f]$ is the $\sim$-equivalence class of f) formed a non-Archimedean linear continuum. However, that was not the case. The ordering induced on $\mathcal{P}$ by $\prec$ is only a partial ordering and there are functions for which none of the conditions $f \prec g$, $f \succ g$ and $f \sim g$ holds (this is because the limits of some of the quotients may not exist).

The study of the degrees of infinity was further developed by several authors, most notably by Felix Hausdorff and Godffrey Hardy (Hausdorff 1907, Hardy 1954). Below we refer in brief to the ideas investigated by Hausdorff, some of them still of much interest in the foundations of mathematics. We shall formulate them in a modern notation, following the presentation in Scheepers 1993, 1994.

Let ω be the set of all positive integers and ${}^{\omega}\omega$ the set of all functions from ω to ω. For f, g in ${}^{\omega}\omega$ let $f \prec g$ means that $\lim_{n\to\infty}(g(n) - f(n)) = \infty$ and $f =^* g$ means that f and g differ only at a finite number of arguments. Then $({}^{\omega}\omega, \prec)$ is a partially ordered set and $=^*$ is an equivalence relation preserving $\prec$.

For any ordinal numbers κ and λ (not both equal 0) and any subsets $\{f_\alpha : \alpha < \kappa\}$, $\{g_\beta : \beta < \lambda\}$ of the set ${}^{\omega}\omega$ we say that the pair of sets $(\{f_\alpha : \alpha < \kappa\}, \{g_\beta : \beta < \lambda\})$ is a (κ, λ)-pregap, if

$$f_{\alpha_1} \prec f_{\alpha_2} \prec g_{\beta_2} \prec g_{\beta_1}$$

for all $\alpha_1 < \alpha_2 < \kappa$ and all $\beta_1 < \beta_2 < \lambda$.

A (κ, λ)-pregap $(\{f_\alpha : \alpha < \kappa\}, \{g_\beta : \beta < \lambda\})$ is called a (κ, λ)-gap, if there does not exist h in ${}^{\omega}\omega$ such that for all $\alpha < \kappa$ and all $\beta < \lambda$: $f_\alpha \prec h \prec g_\beta$.

These definitions are applicable to any partially ordered set, for instance also to $(\wp(\omega), \subset^*)$, where the relation $\subset^*$ is defined as follows: $A \subset^* B$ if $A - B$ is finite and $B - A$ is infinite.

The notion of a gap played a very important role in Hausdorff's analysis of densely ordered sets. Hausdorff investigated also the problem of a possible linear scala for rates of growth of real valued functions diverging to infinity (or converging to zero). He discussed several possibilities of such ranking and showed under which conditions they worked. Hausdorff focused his attention on subsets of the quotient space ${}^{\omega}\omega / =^*$ which were maximal linearly ordered. These he called *pantachies.*

Felix Hausdorff first showed that maximal linearly ordered subsets of the ordered set $({}^{\omega}\omega, \prec)$ have cardinality $2^{\aleph_0}$ and that if they have no

(ω_1, ω_1)-gaps, then they are of cardinality $2^{\aleph_1}$. But in 1909 Hausdorff proved the existence of (ω_1, ω_1)-gaps in $({}^\omega\omega, \prec)$ and the proof can be carried on in ZFC.

Some types of gap cannot occur in $({}^\omega\omega, \prec)$. If κ and λ are ordinal numbers less than $\omega + 1$, then there are no (κ, λ)-gaps in $({}^\omega\omega, \prec)$.

The gap structure of $({}^\omega\omega, \prec)$ (and of similar systems, for instance $(\wp(\omega), \subset^*)$) is very complicated and it is intensively studied in the contemporary foundations of set theory. The properties in question are related to certain independent statements in set theory. We have mentioned the notion of a gap here because it is of course related to some form of completeness. As the reader surely remembers, real numbers may be represented by arbitrary sequences of natural numbers and these sequences are elements of ${}^\omega\omega$. Hence properties of gaps in the structure $({}^\omega\omega, \prec)$ are related to properties of real numbers.

5.6 Infinitesimals and non-Archimedean structures

Infinitely small magnitudes were already known in antiquity – for instance angles between a circle and a straight line tangent to it are such. Infinitely small quantities also appeared in the form of *indivisibles*, for instance in Archimedes' heuristic arguments used in calculation of areas and volumes and later as well, for instance in the works of Cavalieri. The use of infinitesimals in the origin of Calculus is described in detail in numerous treatises and we see no reason to replicate the known views here. We are going to recall only two constructions of non-Archimedean structures, namely those of a hyperreal field and a surreal field.

In popular opinion, the infinitesimal quantities were abandoned from mathematics as a result of the *arithmetization of analysis* credited to Cauchy, Dedekind, Weierstrass and others in the 19th century. According to such views, analysis can be built on arithmetical concepts alone without any references to intuitive geometrical or kinematic notions. The concepts of limit and continuity can be given definitions in the $\varepsilon - \delta$ language and the fundamental number structure for analysis is that of the field of real numbers, which is Archimedean.

However this story is not entirely true. Obviously, the $\varepsilon - \delta$ style of talking has become a commonplace and was extremely useful in precise formulations and proofs of theorems. But at the same time non-Archimedean structures (and infinitesimals connected with them) did not fall into complete oblivion – they were intensively investigated. A comprehensive discussion of the origins of a mathematical theory of non-Archimedean structures can be found in an article Ehrlich 2006, where the works of Hans Hahn, Otto Stolz, Giuseppe Veronese and others are mentioned.

A simple example of a non-Archimedean structure is the field of rational functions, meaning functions which are fractions of polynomials with coefficients in some field, for instance the field of real numbers. In the ordering of this field, the function $f(x) = \frac{1}{x}$ is an infinitesimal.

5.6.1 Hyperreal field

The hyperreal field can be obtained by an application of the *ultraproduct* construction. The concept of ultraproduct was investigated by Edwin Hewitt in 1948 and by Jerzy Łoś in 1949. The ultraproduct construction was also anticipated by Thoralf Skolem when he introduced his non-standard model of arithmetic in 1933.

We shall present in brief the construction of the hyperreal field as an ultrapower of the real numbers. To this goal we need to recall some algebraic notions which were introduced in chapter 4.

Let F be an ultrafilter on $\mathbb{N}$ containing the Fréchet filter, the filter consisting of all cofinite sets of natural numbers. The set $\mathbb{R}^{\mathbb{N}}$ is the set of all sequences of real numbers. We define the relation $\sim_F$ on $\mathbb{R}^{\mathbb{N}}$ as follows:

$$(a_n) \sim_F (b_n) \quad \text{if and only if} \quad \{n \in \mathbb{N} : a_n = b_n\} \in F.$$

Thus two sequences from $\mathbb{R}^{\mathbb{N}}$ are within the relation $\sim_F$ if and only if they coincide on a set of indices belonging to the ultrafilter F. It can be shown that $\sim_F$ is an equivalence relation. Let $\mathbb{R}^*$ be the quotient set $\mathbb{R}^{\mathbb{N}}/\sim_F$ and let:

1. $\mathbb{N}^* = \{[(n_j)]_{\sim_F} : (n_j) \in \mathbb{N}^{\mathbb{N}}\}$
2. $\mathbb{Q}^* = \{[(q_j)]_{\sim_F} : (q_j) \in \mathbb{Q}^{\mathbb{N}}\}$
3. $r^* = [(r, r, r, \ldots)]_{\sim_F}$ for $r \in \mathbb{R}$.

One can define arithmetical operations and an ordering in $\mathbb{R}^*$:

1. $[(a_n)]_{\sim_F} +_F [(a_n)]_{\sim_F} = [(a_n + b_n)]_{\sim_F}$

2. $[(a_n)]_{\sim_F} \cdot_F [(a_n)]_{\sim_F} = [(a_n \cdot b_n)]_{\sim_F}$
3. $[(a_n)]_{\sim_F} <_F [(a_n)]_{\sim_F}$ if and only if $\{n \in \mathbb{N} : a_n < b_n\} \in F$.

One can prove that the structure $\mathfrak{R}^* = (\mathbb{R}^*, +_F, \cdot_F, <_F, 0^*, 1^*)$ is an ordered field. It is called the *hyperreal field* and its elements are called *hyperreal numbers.* One defines the norm (the absolute value) in the standard way:

1. $|a| = a$, if $a \geqslant_F 0^*$,
2. $|a| = -a$, if $a <_F 0^*$.

The following sets play a special role in work with hyperreal numbers:

1. $O = \{x \in \mathbb{R}^* : \exists\lambda \in \mathbb{R}_+ \; |x| <_F \lambda^*\}$
2. $\Omega = \{\varepsilon \in \mathbb{R}^* : \forall\lambda \in \mathbb{R}_+ \; |x| <_F \lambda^*\}\}$.

Elements of the set O are *bounded numbers* and elements of the set Ω are *infinitely small numbers* (*infinitesimals*). Any $\sim_F$-equivalence class of a sequence of real numbers converging to 0 is an example of an infinitesimal in $\mathbb{R}^*$.

The set Ω is closed with respect to addition and multiplication. There is neither the greatest nor the smallest element in this set. Because the structure $\mathfrak{R}^*$ contains infinitely small and infinitely large numbers, it is not an Archimedean field.

The structure $(\{r^* : r \in \mathbb{R}\}, +_F, \cdot_F, <_F, 0^*, 1^*)$ is isomorphic to the standard field of real numbers and hence the latter may be considered as a subfield of the hyperreal field.

We say that two hyperreal numbers r and s are *infinitely close*, in symbols $r \approx s$, if $r - s \in \Omega$, that is if they differ by an infinitesimal. This relation is an equivalence relation. For any bounded hyperreal number $a \in O$ by its *monad* we mean the set $\mu(a) = \{x \in O : a \approx x\}$, which is the set of all hyperreal numbers which are infinitely close to a.

For any bounded hyperreal number $a \in O$ there exists exactly one standard number $r \in \mathbb{R}$ such that $a \approx r^*$. This number is called the *standard part* of a.

One can prove that the set of all cuts of the ordered set $(\mathbb{R}^*, <_F)$ does not contain any gap, in other words it is Dedekind complete (if we define the ordering of those cuts similarly as in Dedekind's original definition). However, if we repeat Dedekind's definitions of addition and multiplication of those cuts, then the family of such cuts does not form a field. If we consider the cut (X, Y) where $X = \mathbb{R}^*_- \cup \Omega$ and $Y = \mathbb{R}^* - Y$, then (X, Y) added to itself equals again (X, Y) and (X, Y) multiplied by itself equals again (X, Y), too. Hence the set of all cuts of the ordered set $(\mathbb{R}^*, <_F)$ with arithmetical operations on those cuts is neither an additive nor a multiplicative group.

Observe also that neither the set of all cuts of the ordered set $(\mathbb{R}^*, <_F)$ nor the set $(\mathbb{R}^*, <_F)$ itself contains a dense countable subset (meaning they do not form a separable space in order topology). Hence none of these two sets is isomorphic to the standard real numbers.

The field $\mathfrak{R}^*$ is Cauchy complete (but not Archimedean, as mentioned above). It plays a fundamental role in the *non-standard analysis* originated by Abraham Robinson, see Robinson 1996.

The construction of the hyperreal field given above depends on the choice of an ultrafilter. It can be proved, under assumption of the generalized continuum hypothesis, that the field obtained by the ultrapower construction from the space of all sequences of real numbers is unique (up to isomorphism). Moreover, the paper Kanovei and Shelah 2003 presents an explicit construction of a definable (in ZFC) countably saturated elementary extension of the real numbers.

5.6.2 Surreal field

The surreal field can be described either axiomatically or by an explicit construction. The former was chosen by Norman Alling (Alling 1962), the latter by John Conway (Conway 1976, Knuth 1974).

In the axiomatic approach we characterize the system $(No, <, b)$ of *surreal numbers* by the following conditions:

1. $<$ is a linear ordering on No.

2. b is a function from No to the class of all ordinal numbers. It is called the *birthday function*.

3. For any subclasses A and B of the class No such that $x < y$ for all $x \in A$ and all $y \in B$ there exists exactly one $z \in No$ such that $b(z)$ is minimal (in the ordering of ordinal numbers) and for all $x \in A$ and all $y \in B$: $x < z$ and $z < y$. In addition, if an ordinal number α is greater than $b(x)$ for all $x \in A$, then $b(z) \leqslant \alpha$ (in the ordering of ordinal numbers).

No is a proper class and thus one needs to work in a set theory with classes, for example von Neumann-Bernays-Gödel set theory. The above conditions characterize the system $(No, <, b)$ in a unique way (up to isomorphism) – see the discussion in the next subsection.

Conway's construction of surreal numbers is based on set theory. The surreal numbers are created in stages via transfinite induction. The same concerns the ordering of these numbers and their birthday function. The initial object is the empty set $\emptyset$, which is identified with the number 0. Each surreal number is a pair (L, R) (also written $(L|R)$) of sets of surreal numbers created at earlier stages and such that every number in L is less than every number in R and the value of the birthday function for (L, R) is greater than the value of this function for every element of L as well as every element of R. This mode of creation resembles Dedekind's cuts, however there are some differences between the two constructions.

The ordering is defined inductively. Suppose that $x = (X_L, X_R)$ and $y = (Y_L, Y_R)$ are two surreal numbers obtained at some stage of the construction. Then $x \leqslant y$ if and only if:

1. $\forall u \in X_L \forall v \in Y_L\, u \leqslant v$ and
2. $\forall u \in X_R \forall v \in Y_R\, u \leqslant v$.

If $x \leqslant y$ and $y \leqslant x$, where $x = (X_L, X_R)$ and $y = (Y_L, Y_R)$, then we say that x and y represent the same surreal number. Thus surreal numbers are actually equivalence classes of pairs of sets. The abuse of terminology can be avoided if we call the pairs (X_L, X_R) *numeric forms* and their equivalence classes (under the relation $\leqslant \cap \geqslant$) *surreal numbers.*

The construction makes use of transfinite induction. At the first stage we construct a form $\{|\}$ which consists of the empty set on the left as well as on the right. Its equivalence class is the surreal number 0. The first stage (or generation) is thus $S_0 = \{0\}$. For any ordinal number α the generation S_α is defined to be the set of all surreal numbers that are generated by the construction rule from subsets of $\bigcup_{\beta<\alpha} S_\beta$.

Thus $S_1 = \{\{|0\}, \{|\}, \{0|\}\}$, which corresponds to the numbers, -1, 0 and 1, respectively and in this order. The set S_2, in turn, corresponds to the numbers (ordered from left to right):

$$-2 < -1 < -\frac{1}{2} < 0 < \frac{1}{2} < 1 < 2.$$

The arithmetical operations on surreal numbers are defined as follows:

1. *Additive inverse.* $-x = -(X_L, X_R) = (-X_R, -X_L)$, where $-X = \{-x : x \in X\}$.

2. *Addition.* $x+y = (X_L, X_R)+(Y_L, Y_R) = (X_L+y \cup x+Y_L, X_R+y \cup x+Y_R)$, where $X+y = \{x+y : x \in X\}$ and $x+Y = \{x+y : y \in Y\}$.

3. *Multiplication.* $xy = (X_L, X_R)(Y_L, Y_R) =$

   $((X_Ly+xY_L-X_LY_L) \cup (X_Ry+xY_R-X_RY_R), (X_Ly+xY_R-X_LY_R) \cup (X_Ry + xY_L - X_RY_L))$,

   where $XY = \{xy : x \in X \wedge y \in Y\}$, $xY = \{x\}Y$, $Xy = X\{y\}$.

4. *Division.* Division is defined in terms of multiplication and reciprocal: $\frac{x}{y} = x\frac{1}{y}$.

Here are a few examples of (numeric forms of) surreal numbers:

1. All numbers in the set S_n, where n is a natural number are *dyadic fractions.*

2. Let $S^* = \bigcup_{n \in \mathbb{N}} S_n$. Then S^* is closed under addition and multiplication.

3. $\omega = (S^*|) = (1, 2, 3, 4, 5, \dots |)$

4. $\varepsilon = (0|1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \dots)$

5. $\omega + 1 = (1, 2, 3, 4, 5, \dots |) + (0|) = (1, 2, 3, 4, 5, \dots, \omega|)$

6. The set S_ω has the same cardinality as the set of real numbers but it is not a field.

7. $\frac{1}{5} = (\{x \in S^* : 5x < 1\}|\{x \in S^* : 5x > 1\})$

8. $\frac{1}{\varepsilon} = \omega$

9. $\frac{\varepsilon}{2} = (0|\varepsilon)$

10. $\omega - 1 = (\{1, 2, 3, 4, 5, \dots\}|\omega)$

11. $2\omega = \omega + \omega = (\{\omega + 1, \omega + 2, \omega + 3, \dots\}|)$

A proper subset of the class of surreal numbers might not have a least upper bound (or a greatest lower bound). This happens when the set does not have a maximal (respectively, minimal) element. By a *gap* in the class No one means a pair (L, R) such that $L < R$ and $L \cup R = No$. Observe that if (L, R) is a gap, then at least one of L and R is a proper class.

The surreal numbers form a field, but the domain is not a set but a proper class. It is the maximal ordered field and all other ordered fields can be embedded in it. Thus, the ordered field of surreal numbers contains as substructures the following structures, among others: the field of rational numbers $\mathbb{Q}$ as well as the field of real numbers $\mathbb{R}$, the field of hyperreal numbers, the field of rational functions, the Levi-Civitta field, and the class of transfinite ordinal numbers, where the arithmetical operations of addition and multiplication are those proposed by Hessenberg (the so called *natural operations*). There also exist further representations of the field of surreal numbers, for example in terms of games like Hackenbush.

5.6.3 Ehrlich on continuity axiom

Quite recently Philipp Ehrlich published several papers (Ehrlich 1988, 1992, 1995, 2006, 2012) on continuity and infinitesimals which stress the importance of studies of non-Archimedean structures. In the short paper *Universally extending arithmetic continua* (Ehrlich 1992) Ehrlich characterizes his program as follows in the introductory remarks (we omit references to the items in the bibliography of the paper):

> Novel (categorical) axiomatizations of the ordered field of reals are provided making use of the Archimedean axiom in conjunction with heretofore unnoticed *continuity axioms* based on concepts adapted from the theory of *homogeneous universal models* (...), and the closely related theory of *real-closed fields that are* η_α*-orderings* (...). We believe we are justified in referring to these axioms as continuity axioms since in the context of these axiomatizations they are equivalent to any of the more familiar continuity conditions including, for example, those due to Cantor, Dedekind and Hilbert. (Ehrlich 1992, 168)

The main results obtained by Ehrlich can be summarized as follows:

1. Working in a rather strong metatheory (set theory NBG with global choice) Ehrlich obtains categorical axiomatizations of Conway's ordered field No (the Archimedean axiom is in this case deleted from the list of axioms).

2. Using a natural generalization of the notion of a *gap* enables him to prove, among other things that:

   (a) $\mathbb{R}$ is the unique (up to isomorphism) real-closed ordered field having no gaps modulo the Archimedean condition.

   (b) No is the unique (up to isomorphism) real-closed ordered field having no gaps that are definable in terms of sets of NBG.

Ehrlich claims that these results support the following theses:

1. The real number system should be regarded as constituting *an Archimedean arithmetic continuum.*

2. The field No may be regarded as an *absolute arithmetic continuum* (modulo NBG).

Let us add that some of the above results were obtained by applying notions from classical model theory (such as the homogeneous and universal models mentioned in chapter 4). However, Ehrlich's results can also be expressed in the primitive language of elementary algebra, though supplemented with some basic set-theoretic notions. For example, it was shown in Ehrlich 1988 that: No is (up to isomorphism) the unique real closed field F that is an η_{On}-ordering, meaning that for all subsets X and Y of F where $X < Y$, there exists a $y \in F$ such that $X < \{y\} < Y$. Ehrlich claims that this property may be regarded as *an axiom of absolute density* and that No is thus the unique absolutely dense ordered field which admits no order preserving algebraic extension.

5.7 Continua in topology

Continua are investigated in general topology. Here by a *continuum* one means any non-empty compact connected metric space. Sometimes the term *continuum* is reserved for a compact connected Hausdorff space. Simple examples of continua are:

1. An arc, that is a space homeomorphic to the closed interval $[0, 1]$.

2. An n-cell, that is a space homeomorphic to the closed ball in the Euclidean space $\mathbb{R}^n$.

3. An n-sphere, that is a space homeomorphic to the standard Euclidean n-sphere in the $(n + 1)$-dimensional Euclidean space.

4. The Hilbert cube, that is the topological product of the intervals $[0, \frac{1}{n}]$, for all $n \in \mathbb{N}$.

However, continua in topology also include many much more complicated structures, sometimes called *pathological*. It must be stressed that these pathologies were invented *on purpose*, partly in order to show that our intuitions must be modified when confronted with logical consequences of the accepted initial assumptions. Some examples of such "bizarre" continua are:

1. An indecomposable continuum is a continuum that is not the union of any two of its proper subcontinua. The pseudo arc is an example of a hereditarily indecomposable continuum.

2. The lakes of Wada are three disjoint connected open sets of the plane or open unit square such that they all have the same boundary. Thus, a common boundary of three sets may be an indecomposable continuum.

3. The Sierpiński carpet, described by him as early as in 1916 and also known as a universal curve on the plane. A geometrical description of this object runs as follows. We start with a square. The square is cut into 9 congruent subsquares in a 3-by-3 grid, and the central subsquare is removed. The same procedure is then applied recursively to the remaining 8 subsquares, and so on. The Sierpiński carpet is the intersection of all sets obtained at these stages.

Chapter 6

The axiom of induction in arithmetic

The axiom of induction is a minimal extremal axiom. It is supposed to characterize the natural numbers sequence as a minimal structure ordered in a specific way. We discuss its origin and its use in arithmetic. We have already presented the original axioms for natural numbers given by Giuseppe Peano and the construction of simply infinite systems credited to Richard Dedekind in chapter 2. Below we present in brief the axiomatic systems of arithmetic: first-order Peano arithmetic, Robinson arithmetic and second-order arithmetic. We add brief remarks about non-standard models of arithmetic and recall the role of Tennenbaum's theorem in the characterization of the intended model of arithmetic.

6.1 A few historical remarks

Some forms of inductive reasoning can already be found in the early history of mathematics. This primarily concerns argumentations which involve the impossibility of infinite regress. Such argumentations were presented as early as in Euclid's work (for example Proposition 31 of Book VII, in which Euclid proves that every composite integer is measurable by some prime number). A simple example is the proof of the irrationality of $\sqrt{2}$. The most popular proof is by contradiction and refers to infinite descent. Suppose that $\sqrt{2}$ is rational, say $\sqrt{2} = \frac{a}{b}$, where a and b are natural numbers. Then:

1. By squaring the equation $\sqrt{2} = \frac{a}{b}$ we get: $2 = \frac{a^2}{b^2}$ and, consequently $2b^2 = a^2$.
2. The left-hand side of the equation $2b^2 = a^2$ is even, and hence the right-hand side is also even.
3. If a^2 is even, then a is also even, that is $a = 2c$ for some natural number c.
4. Hence we have: $2b^2 = (2c)^2$, that is $b^2 = 2c^2$.
5. Since $2c^2$ is even, b^2 is even, too.
6. This implies that b is even, that is $b = 2d$ for some natural number d.
7. Observe that we can repeat the above argumentation indefinitely: since $(2d)^2 = 2c^2$, we have $2d^2 = c^2$ and we proceed as before.
8. But notice that we have: $\frac{a}{b} = \frac{2c}{2d} = \frac{c}{d}$.
9. If it were the case that $\sqrt{2} = \frac{a}{b}$, then also $\sqrt{2} = \frac{c}{d}$ for $c < a$ and $d < b$.
10. But such an infinite descent is impossible in the ordered set of all natural numbers and hence $\sqrt{2} = \frac{a}{b}$ holds for no natural numbers a and b, which means that $\sqrt{2}$ is irrational.

Thus reasoning by infinite descent was known already in ancient Greece. In order to show that no number has the property ψ it suffices to show that for any number n which has the property ψ there exists a number $m < n$ which has the property ψ. If there existed a number with the property ψ, then we could obtain smaller and smaller numbers with that property, which was considered absurd.

Blaise Pascal used the method of inductive proof in 1654, in justification of the arithmetical dependencies in what is commonly called the Pascal triangle (although this triangle was already known to the Chinese more than 600 years before Pascal).

Pierre Fermat used the method of infinite descent to prove several theorems in number theory. For example, he proved Girard's claim that every prime number of the form $4n+1$ can be written in one and only one way as the sum of two squares. Fermat conjectured (though admitted that he could not prove) that all numbers of the form $F_n = 2^{2^n} + 1$

(now known as the Fermat numbers) are prime. This claim is not true, which was shown by Euler who noticed that $F_5 = 2^{2^n} + 1 = 2^{32} + 1 = 4294967297$ is a composite number: it is the product of 641 and 6700417. Fermat's little theorem (if p is a prime number, then for any integer a, the number $a^p - a$ is an integer multiple of p), however, can be proved by mathematical induction:

> Fermat's lesser theorem fared better than his conjecture on prime Fermat numbers. A proof of the theorem was left in manuscript by Leibniz, and another elegant and elementary demonstration was published by Euler in 1736. The proof by Euler makes ingenious use of mathematical induction, a device with which Fermat, as well as Pascal, was quite familiar. In fact, mathematical induction, or reasoning by recurrence, is sometimes referred to as "Fermatian induction," to distinguish it from scientific, or "Baconian," induction. (Merzbach and Boyer 2010, 328)

In *Arithmetica Infinitorum* published in 1655, John Wallis used a form of incomplete induction to obtain that:

$$\int_0^1 x^n dx = \frac{1}{n+1}.$$

Jacques Bernoulli wrote a treatise *Ars Conjectandi* (published in 1713, after his death) which is considered to be "the earliest substantial volume on the theory of probability" as Merzbach and Boyer write:

> The second part of the *Ars Conjectandi* includes a general theory of permutations and combinations, facilitated by the binomial and multinomial theorems. Here we find the first adequate proof of the binomial theorem for positive integral powers. The proof is by mathematical induction, a method of approach that Bernoulli had rediscovered while reading the *Arithmetica Infinitorum* of Wallis and that he had published in the *Acta Eruditorum* in 1686. (Merzbach and Boyer 2010, 393)

6.1.1 Hermann Grassmann

The axiomatic approach to number systems originated in the 19th century. The first axiomatization – of integers – was proposed by Hermann Grassmann in 1861. The essential property of the structure described in his system was characterized as follows:

> 6. *Erklärung. Die Arithmetik* ($\alpha\rho\iota\theta\mu\eta\tau\iota\kappa\eta$) behandelt diejenigen Grössen, welche aus einer einzigen Grösse e durch Verknüpfung hervorgehen.
>
> 7. *Erklärung.* Man bilde aus einer Grösse e eine Reihe von Grössen durch folgendes Verfahren: Man Setze e als ein Glied der Reihe, setze $e+e$ (gelesen e plus e) als das nächstfolgend Glied der Reihe, und so fahre man fort, aus dem jedesmal letzten Gliede das nächstfolgende dadurch abzuleiten, dass man ze jenem $+e$ hinzufügt. Ebenso setze man $e + -e$ (gelesen e plus minus e) als dem e zunächst vorhergehende Glied der Reihe, und so fahre man fort, aus dem jedesmal ersten Gliede der Reihe das nächst vorhergehend dadurch abzuleiten, dass man zu jenem Gliede $+-e$ hinzufügt, so erhält man eine nach beiden Seite unendliche Reihe
>
> $$\ldots e + -e + -e + -e, e + -e + -e, e + -e, e, e + e, e + e + e \ldots$$
>
> Wenn man in dieser Reihe jedes Glied von allen übrigen Gliedern der Reihe als verschieden annimmt, so nennt man diese Reihe die *Grundreihe*, e die *positive* Einheit, $-e$ die *negative* Einheit. (Grassmann 1861, 2–3)

Grassmann's system was formally reconstructed by Hao Wang – see Wang 1957; compare also Hanusek 2015. One should remember that at that time Grassmann did not have at his disposal any system of formal logic. However, his presentation is clear enough to formulate it as a system of definitions and axioms. According to the reconstruction proposed in Hanusek 2015, the language of the system uses individual variables, an individual constant 1, two two-argument function symbols $+$ and $\cdot$ (for addition and multiplication, respectively), two one-argument function symbols $^\bullet$ and $\sim$ and the symbol Pos denoting a subset of the domain of interpretation. The terms of the system can be defined inductively: 1 and $1^\bullet$ are terms, individual variables are terms and if s and t are terms, then also $s + t$ and $s \cdot t$ are terms. Some defined terms are:

1. $0 = 1 + 1^\bullet$
2. For any a and b, $a - b$ is the number c such that $b + c = a$
3. $\sim a = 0 - a$
4. $a > b$ if and only if $a - b \in Pos$

The axioms of this system are as follows:

1. $a = (a + 1) + 1^{\bullet}$
2. $a = (a + 1^{\bullet}) + 1$
3. $a + (b + 1) = (a + b) + 1$
4. $a \cdot 0 = 0$
5. $1 \in Pos$
6. $a \in Pos \rightarrow a + 1 \in Pos$
7. $(b = 0 \vee b \in Pos) \rightarrow a \cdot (b + 1) = a \cdot b + a$
8. $b \in Pos \rightarrow a \cdot \sim b = \sim (a \cdot b)$
9. For any set A, if $1 \in A$ and for every a, if $a \in A$, then $a + 1 \in A$ and $a + 1^{\bullet} \in A$, then for every a, $a \in A$.
10. For any set A, if $1 \in A$ and for every a, if $a \in A$, then $a + 1 \in A$, then for every a if $a \in Pos$, then $a \in A$.

Grassmann's work went long undervalued in the history of mathematics. Yet it provided essential influence for such as the later work of Peano, which became more well known. Grassmann was the first who proposed the recursive definitions of addition and multiplication. He was also the first who formulated the (non-elementary) axiom of induction (see the last two items on the list above). What is more important, he made it clear that the principle of induction should be considered as a *method of proof* in arithmetic.

The paper Hanusek 2015 contains some improvements on Grassmann's axiomatics. In particular, the author is able to show that the improved system is categorical (which is not true of Grassmann's original system).

6.1.2 Gottlob Frege

Gottlob Frege considered the problem of definition of natural numbers in his *Die Grundlagen der Arithmetik: eine logisch-mathematische Untersuchung über den Begriff der Zahl* (Frege 1884) and *Grundgesetze der Arithmetik, begriffsschriftlich abgeleitet* (Frege 1893–1903). Important roles in these systems were played, among others, by: ancestor relations (hereditary properties), definition of cardinal numbers (in Frege's sense), the Hume principle, as well as Frege's earlier ideas from his *Begriffsschrift: eine der arithmetischen nachgebildete Formelsprache des reinen Denkens* (Frege 1879).

As is well known, Frege advocated logicism, in other words the view that all of mathematics could be reduced to logic. Øysten Linnebo gives a concise presentation of Frege's standpoint as follows:

> What exactly is the arithmetical theory that Frege wished to reduce to pure logic? The target theory is now known as *second-order Dedekind-Peano arithmetic*. This is an arithmetical theory that is formulated in a second-order language with three (for now) non-logical symbols: a constant '0' for the number zero and predicates 'Pxy' and '$\mathbb{N}x$', stating that x immediately precedes y and that x is a natural number, respectively. The theory has the following axioms:
>
> 1. $\mathbb{N}0$
> 2. $\mathbb{N}x \wedge Pxy \to \mathbb{N}y$
> 3. $\mathbb{N}x \wedge Pxy \wedge Pxy' \to y = y'$
> 4. $\mathbb{N}x \wedge Pxy \wedge Px'y \to x = x'$
> 5. $\mathbb{N}x \to \exists y Pxy$
> 6. $\forall F(F0 \wedge \forall x \forall y(\mathbb{N}x \wedge Fx \wedge Pxy \to Fy)) \to \forall x(\mathbb{N}x \to Fx)$
>
> To reduce this theory to pure logic, Frege's first task is to provide a logical analysis of the three symbols still left unanalyzed. (Linnebo 2017, 33)

Let $\#Fx$ denote the number of x which have the property F. The Hume principle, accepted by Frege, has the following form:

$$\#Fx = \#FGx \;\equiv\; F \approx G,$$

where F and G are concepts and $F \approx G$ means that there is a one-one correspondence between the extensions of F and G. Frege's reduction of the terms 0, Pxy and $\mathbb{N}x$ to purely logical terms can then be presented as follows:

1. $0 =_{df} \#x(x \neq x)$
2. $Pxy \equiv_{df} \exists F \exists a(Fa \wedge x = \#u(Fu \wedge u \neq a) \wedge y = \#uFu)$

3. Let us call a concept F *hereditary* (in symbols $Her(F)$), if it is inherited under the predecessor relation, that is if:

$$\forall x \forall y (Fx \wedge Pxy \rightarrow Fy).$$

Then we can define the concept of natural number in the following way:

$$\mathbb{N}x \equiv_{df} \forall F(F0 \wedge Her(F) \rightarrow Fx).$$

Linnebo writes that if one admits that abstraction principles such as the Hume principle are purely logical, then one should evaluate Frege's program as successful. However, there remains a serious problem with Frege's famous Basic Law V which concerns the extensions of concepts and their identity:

$$\{x : Fx\} = \{x : Gx\} \ \equiv \ \forall x(Fx \equiv Gx).$$

The difficulty lies in the fact that not every property determines a set – compare Russell's paradox. The difficulty has been eliminated in axiomatic set theory by assuming suitable restrictions in the formulation of the separation axiom. Frege's program, in turn, has been reconsidered, as for example in Wright 1983.

6.1.3 Witold Wilkosz

An interesting, though poorly-known axiom system for natural numbers was proposed in 1932 by Witold Wilkosz (Wilkosz 1932). There are two primitive terms of the system: the set of all natural numbers $\mathbb{N}$ and the relation $<$ (less than). The axioms are not elementary, because they involve set variables. Here are the axioms (set variables are capital letters):

1. $\exists x \; x \in \mathbb{N}$
2. $\forall x \exists y \; x < y$
3. $\forall x \forall y \; (x < y \rightarrow \neg(y < x))$
4. $\exists z \; z \in Z \rightarrow \exists x \; (x \in Z \wedge \forall y \; (y \in Z \rightarrow x \leqslant y))$
5. $(\exists z \; z \in Z \wedge \exists y \forall x \; (x \in Z \rightarrow x \leqslant y)) \rightarrow \exists x \; (x \in Z \wedge \forall y \; (y \in Z \rightarrow y \leqslant x))$

Here the relation $\leqslant$ is defined in the usual way (it is assumed that the system uses the identity predicate). The meaning of the first three axioms is obvious. The fourth axiom expresses a minimum principle, while the fifth expresses the maximum principle:

1. *The minimum principle.* In any non-empty set of natural numbers there exists the smallest number.

2. *The maximum principle.* In any non-empty set of natural numbers bounded from above there exists the greatest number.

One can prove that this system is equivalent to the system proposed by Peano and presented in chapter 2.

6.1.4 Ernst Zermelo

Around the time when Ernst Zermelo proposed the first axiomatization of set theory, he also published two papers devoted to arithmetic: Zermelo 1909a, 1909b. He writes at the beginning of Zermelo 1909a:

> Is the principle of mathematical induction provable or not? This question has occupied many minds during the past few years. In several articles in the *Revue de Métaphysique et de Morale*, *Mr. Poincaré* has defended the thesis that this principle is a *synthetic a priori judgment*. Other autors, such as *Mr. Couturat, Mr. Russell, and Mr. Whitehead*, have supported the contrary position and have presented proofs of the principle in question. (Citing Zermelo 2010, 237)

Zermelo further argues that the answer to the above question can be obtained by providing a definition of a finite set. By a *shifting* of a set M he means a one-one mapping f of a non-empty proper subset P of M properly into M such that for any partition of M into two non-empty subsets M_1 and M_2 there exists $p \in M$ such that either $p \in M_1$ and $f(p) \in M_2$ or vice versa, $p \in M_2$ and $f(p) \in M_1$. If a set M can be mapped into itself by such a shifting, then it is called a *finite chain.*

Zermelo proves that any finite chain contains exactly one initial element and exactly one final element. What is the initial and what is the final element depends on the mapping f (and does not depend on the partitions of the set in question). Then Zermelo proves that being a finite chain is equivalent to having a double (forwards and backwards) well ordering. He makes use of this fact in his proof of the induction principle for finite sets:

> **Theorem III** (Theorem on mathematical induction for finite cardinal numbers). – If a property E holds for every set consisting of a single element, and if, furthermore, it holds for a set M whenever it holds for a set M_1 which results when a single element m_1 is removed from M, then it holds generally for all finite sets. (Zermelo 1909b, citing Zermelo 2010, 257)

Let us add that a German mathematician, Paul Stäckel, also proposed calling a set finite if it can be totally ordered in such a way that the ordering is a well ordering forwards and backwards (Stäckel 1907). Then a finite set in this sense has the smallest as well as the greatest element. Heinrich Weber defines a finite set as one which has an ordering such that every non-empty subset has a first and a last element (Weber 1906).

6.2 Definitions of finiteness

Natural numbers are finite objects which are characterized by the axioms of arithmetic. However, one can pose a more general question: what are finite objects? Can we characterize the notion of finiteness in general formal terms, not necessarily depending on the notion of natural number but such that natural numbers would also be considered finite according to these general criteria?

The proposals by Dedekind and Zermelo mentioned above are attempts at providing such criteria. Several authors have given yet other characterizations and comparisons between different definitions of finiteness. The most famous are those proposed by Kazimierz Kuratowski and Alfred Tarski. Comparisons between different notions of finiteness were discussed for example in: Lindenbaum and Mostowski 1938, Tarski 1924, Mostowski 1938, Levy 1958, and Howard and Yorke 1989.

Let us say that a set X is Num-finite (numerically finite) if there exists a natural number n such that X has n elements. This (very intuitive) definition presupposes that natural numbers are somehow characterized a priori.

We remember that a set X is infinite in Dedekind's sense (D-infinite) if X is equinumerous with some of its proper subsets, that is if there exists an injection from X into X. Otherwise, it is finite in Dedekind's sense (D-finite).

Kuratowski's definition (Kuratowski 1920) of a finite set uses the fact that for any set X, its powerset $\wp(X)$ together with the operation of set-theoretical union forms a semi-lattice. Let $K(X)$ be the sub-semi-lattice of this semi-lattice generated by the empty set and all singletons of elements from X. Then $K(X)$ is the intersection of all sub-semi-lattices containing the empty set and all singletons.

We say that a set X is K-finite (Kuratowski-finite), if $X \in K(X)$. Notice that this definition refers only to the set-theoretical notions (inclusion, union, empty set, and singleton) but makes no use of the concept of natural number.

The system $K(X)$ can also be defined in the following way. Let M_X be the family of all sets $Y \subseteq \wp(X)$ such that:

1. $\emptyset \in Y$
2. For every $Z \in \wp(X)$, if $Z \subseteq Y$, then $Z \cup \{x\} \subseteq Y$ for any $x \in X$.

Then one can put $K(X) = \bigcap M_X$. A set X is K-finite if $\wp(X) - \{\emptyset\}$ is equal to:

$$\bigcap \{Y \in \wp(\wp(X)) - \{\emptyset\} : \forall y \in X \ \{y\} \in Y \wedge \forall A, B \in Y \ A \cup B \in Y\}.$$

Kuratowski showed that K-finiteness is equivalent to Num-finiteness. It can be also shown that in set theory ZF (without the axiom of choice), K-finiteness implies D-finiteness but not vice versa.

An interesting definition of the notion of a finite set is the one provided by Tarski: a set X is finite (T-finite) if every non-empty family of its subsets has a maximal element (with respect to inclusion), see Tarski 1924. According to this definition the set $\mathbb{N}$ of all natural numbers is not finite, because the following family of its subsets does not contain a maximal element:

$$\{\{k \in \mathbb{N} : k \leqslant n\} : n \in \mathbb{N}\}.$$

Azriel Levy gives the following list of definitions of the notion of finiteness:

> Definitions of finiteness
>
> I. [This is one of the definitions of Tarski 1924.] A is finite if every non-void family [By "family" we mean a set of sets.] of subsets of A has a maximal element (an element which is not a proper subset of any element of A).
>
> We get an equivalent definition if we replace "maximal" by "minimal" (cf. Tarski 1924).

> Ia. A is finite if it is not the union of two disjoint sets neither of which is finite according to definition I.
>
> II. [This is one of the definitions of Tarski 1924.] A is finite if every non-void monotonic family (a family which is completely ordered by inclusion) of subsets of A has a maximal element.
>
> We get an equivalent definition if we replace "maximal" by "minimal" (the proof is the same as in the case of definition I).
>
> III. [This is one of the definitions of Tarski 1924.] A is finite if the power-set of A is irreflexive [A set is reflexive if it is equivalent to one of its proper subsets.]
>
> IV. [This is one of the definitions of Tarski 1924.] A is finite if it is irreflexive.
>
> V. [This is one of the definitions of Tarski 1924.] A is finite if $\overline{\overline{A}} = 0$ or $2\overline{\overline{A}} > \overline{\overline{A}}$.
>
> VI. [This is a definition of Tarski 1924 mentioned in Mostowski 1938.] A is finite if $\overline{\overline{A}} = 0, 1$ or $\overline{\overline{A^2}} > \overline{\overline{A}}$
>
> VII. A is finite if $\overline{\overline{A}}$ is not an aleph greater then $\aleph_0$ or equal to it.
>
> All these definitions are equivalent if one assumes the axiom of choice. In fact, definition VII is the most inclusive definition that becomes equivalent to definition I if the choice axiom is assumed. What we usually mean by "finite" is finite according to I; therefore sets finite according to I will be simply called finite. (Levy 1958, 2–3; in square brackets are the footnotes from that paper, $\overline{\overline{A}}$ denotes cardinality of A)

Levy also proves a theorem holding in set theories with the axiom of foundation and admitting no *Urelements* which says that if a set is finite according to any of the above definitions it is finite also according to any definition which follows it.

6.3 First-order arithmetic

Formalizations of arithmetic in a first-order language have advantages and disadvantages. First-order theories have well-developed metatheory and the most important (limitative) metatheorems were obtained for first-order Peano arithmetic and its extensions. But instead of a single induction axiom one has to accept an infinite axiom schema of induction. In a second-order arithmetic addition and multiplication can be defined from the successor function, but in the first-order arithmetic one has to introduce these operations as primitive and characterize them by suitable axioms providing their recursive definitions.

First-order axiomatizations alternative to Peano arithmetic may also be considered, for example axioms for discrete ordered semiring (models of these axioms have a standard part isomorphic to the standard natural numbers).

6.3.1 First-order Peano arithmetic

The first-order Peano arithmetic is based on the following axioms (here s is the successor function, $+$ and $\cdot$ are addition and multiplication, respectively, and 0 stands for zero; we omit the axioms concerning the identity predicate $=$):

1. $\forall x \neg(0 = s(x))$
2. $\forall x \forall y (s(x) = s(y) \to x = y)$
3. $\forall x (0 + x = 0)$
4. $\forall x \forall y (x + s(y) = s(x + y))$
5. $\forall x (s(0) \cdot x = x)$
6. $\forall x \forall y (x \cdot s(y) = x \cdot y + x)$
7. $(\varphi(0) \wedge \forall x(\varphi(x) \to \varphi(s(x)))) \to \forall x \varphi(x)$,

   where φ is a formula of the language of arithmetic with one free variable.

Mathematical induction is represented here by an axiom schema, involving all properties of natural numbers which are expressible in the language of the theory.

Among the important incompleteness results concerning the first-order system of Peano arithmetic (PA) are the following:

1. *Gödel.* If PA is consistent, then it is not complete; it contains undecidable statements. Gödel proved this fact using his arithmetization of syntax and under the assumption of ω-consistency of arithmetic. Rosser showed that this assumption could be replaced by the assumption of consistency.

2. *Gödel.* If PA is consistent, then this fact cannot be proved in the system PA.

3. *Tarski.* The property of being a true sentence (meaning that it is true in the standard model) of first-order arithmetic is not definable in the language of this system of arithmetic.

These incompleteness phenomena hold true also for other important mathematical theories – notably the Zermelo-Fraenkel set theory. This should be evident, because arithmetic may be interpreted in set theory.

Among other metamathematical results concerning the system PA one may list:

1. *Non-existence of finite axiomatizability.* The system PA is recursively axiomatizable, but it is not finitely axiomatizable. It is also not finitely axiomatizable, when one considers only formulas with restricted complexity in the induction axiom schema.

2. *Decidability of some subsystems of PA.* The system PA (with successor, addition, and multiplication) is not decidable. However, arithmetic with successor and addition only is decidable (Presburger) and also arithmetic with successor and multiplication only is decidable (Skolem).

6.3.2 Robinson arithmetic

An important arithmetical theory is the system **Q** of *Robinson arithmetic.* It is formulated in the same first-order language as PA. The system **Q** is based on the following *finite* set of axioms (we omit the axioms concerning the identity predicate $=$):

1. $\forall x \neg(0 = s(x))$

2. $\forall x \forall y (s(x) = s(y) \rightarrow x = y)$

3. $\forall x (x = 0 \vee \exists y (x = s(y)))$

4. $\forall x (0 + x = 0)$

5. $\forall x \forall y (x + s(y) = s(x + y))$

6. $\forall x (s(0) \cdot x = x)$

7. $\forall x \forall y (x \cdot s(y) = x \cdot y + x)$

Notice the differences between the systems PA and **Q**:

1. **Q** is finitely axiomatizable, while PA is based on an infinite (though recursive) set of axioms.

2. The third axiom of **Q** can be proved in PA, with an essential use of induction axiom.

3. **Q** does not include induction axiom schema and for this reason some general theorems of PA are not theorems of **Q** – for instance, commutativity of addition and multiplication is not provable in **Q**. But one can prove in **Q** *concrete* mathematical truths, such as $2 + 3 = 3 + 2$.

4. In **Q** one cannot prove that $x \neq s(x)$.

5. Though **Q** is weaker then PA, it is also – like PA – incomplete and essentially undecidable.

Some fundamental facts about the system **Q** are the following:

1. The intended model of **Q** is the same as the intended model of PA and it is the standard model. But **Q**, like PA, has non-standard models as well.

2. Due to the Tennenbaum theorem (see below), the standard model of PA is its only recursive model (up to isomorphism). The situation with **Q** is different: it has recursive non-standard models.

3. The system PA can be interpreted in set theory ZF. In turn, the system **Q** can be interpreted in a weak subsystem of ZF, whose axioms are extensionality, the existence of the empty set, and the axiom of adjunction, in the following form:

$$\forall x \forall y \exists u \forall v\ (v \in u \equiv (v \in x \vee v = y)).$$

   The meaning of this axiom becomes clear if one observes that $u = x \cup \{y\}$.

6.4 Second-order arithmetic

By *second-order arithmetic* one means the system based on the following axioms:

1. Axioms for successor operation (the successor of a natural number is never zero, the successor function is injective, every natural number is zero or a successor).

2. Axioms providing recursive definitions of addition and multiplication.

3. Axioms for the order relation $<$:

   (a) $\forall n \neg(n < 0)$
   (b) $\forall n \forall m((m < s(n)) \equiv (m < n \vee m = n))$
   (c) $\forall n(0 = n \vee 0 < n)$
   (d) $\forall n \forall m((s(m) < n \vee s(m) = n) \equiv m < n)$

4. Second-order induction axiom.

5. The comprehension axiom stating the existence of a set of all numbers satisfying a given formula of the language of the system (with some obvious restrictions concerning these formulas).

This system is obviously much stronger than the system PA. In a sense, it is sufficient for developing analysis, as it makes it possible to talk about real numbers (real numbers correspond to sets of natural numbers).

Some fundamental facts about this system are as follows:

1. Models of second-order arithmetic have interpretation of successor, zero, addition, multiplication and less-than as in the case of first-order arithmetic. In addition, each model is equipped with a family of subsets of the universe, over which the second order quantification is allowed (this family may coincide with the full powerset of the universe – in this case we talk about the full model).

2. The axioms of second-order arithmetic have only one full model.

3. An ω-model is one in which the universe consists of all (standard) natural numbers and interpretation of all the primitive symbols is standard.

4. The unique full ω-model is called the intended (or standard) model of second-order arithmetic.

5. In *reverse mathematics* (see for instance Stillwell 2018) several subsystems of second-order arithmetic are considered, including the following:

   (a) *The system ACA_0*. In this system the comprehension axiom schema is assumed for each arithmetical formula of the language. It is a conservative extension of first-order arithmetic.

   (b) *The system RCA_0*. In this system the comprehension axiom schema is assumed for each Δ^0_1 formula and the induction axiom is assumed for each Σ^0_1 formula. The set of first-order consequences of RCA_0 is the same as those of the subsystem $I\Sigma^0_1$ of Peano arithmetic in which induction is restricted to Σ^0_1 formulas. This subsystem of PA is conservative over *primitive recursive arithmetic* for Π^0_1 sentences. The family of all recursive sets of natural numbers determines an ω-model of RCA_0.

6. In the subsystem RCA_0 one can formalize several fundamental mathematical structures, for example integers, rational numbers, real numbers, as well as complete separable metric spaces and continuous functions between them.

Systems of arithmetic with *restricted induction principle* are also considered. Here restriction concerns the degree of complexity of the formulas in the induction schema. For instance, in order to make possible the construction of recursive functions and apply the schema of recursion one needs the schema of mathematical induction for Σ^0_1-formulas of the language of arithmetic.

6.5 Non-standard models of arithmetic

The intended model of arithmetic can be characterized if we turn to ω-logic, a system of logic with an infinitary rule. Suppose that we consider a theory T in a first-order countable language with names 0, $\overline{1}$, $\overline{2}$, and so on (individual constants) for all natural numbers. Let N be a unary predicate whose intended meaning is "to be a (standard) natural number". The system of ω-logic is obtained by adding to the usual axioms for first-order logic the ω-rule (here φ stands for any formula with one free variable from the language):

$$\frac{\varphi(0), \varphi(\overline{1}), \varphi(\overline{2}), \ldots}{\forall x(N(x) \rightarrow \varphi(x))}.$$

By an ω-model for T one means a model whose domain includes all (standard) natural numbers, with obvious interpretations for 0, $\overline{1}$, $\overline{2}$, and so on, as well as for N. One can prove that T has ω-model if and only if it is consistent in ω-logic. A theory consistent in ω-logic is also ω-consistent, but the converse implication does not hold.

6.5.1 Methods of constructing non-standard models

The existence of non-standard models of the first-order system PA can be proved in several ways, e.g.:

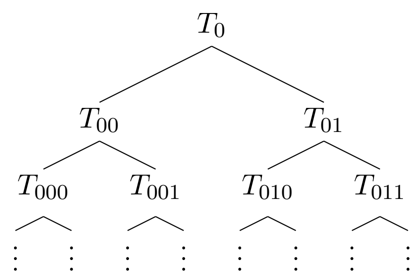
1. *Compactness theorem.* Expand the language of PA by a new constant c and let
$$\Gamma = \{\neg c = \overline{n} : n \in \omega\},$$
where $\overline{n}$ is the numeral for the natural number $n \in \omega$. Then each finite subset of Γ has a model – it suffices to choose an appropriately large natural number as interpretation of c. Due to the compactness theorem Γ itself has a model. But in that model, the interpretation of c differs from each natural number and hence that model of Γ is a non-standard model.

2. *Ultraproduct construction.* An example of a non-standard model of arithmetic was given by Thoralf Skolem back in 1933 (Skolem 1933, 1934). His construction anticipated, in a sense, the general construction of an ultraproduct presented later by Edwin Hewitt and, independently, by Jerzy Łoś. If F is a non-principal ultrafilter on $\wp(\mathbb{N})$ (say, an ultrafilter extending the filter of cofinite subsets of $\mathbb{N}$), then the ultrapower $\mathbb{N}/F$ is a model of PA, with operations of successor, addition, and multiplication defined in the standard way used in such constructions (as well as the relation of less-than). In this model the element which is the equivalence class of the identity function is non-standard.

3. *Tree of extensions of PA.* Continuum many extensions of the system PA are constructed in the following way. Let $T_0 = PA$ and let ψ_0 be any undecidable statement in T_0. We put: $T_{00} = PA \cup \{\psi_0\}$ and $T_{01} = PA \cup \{\neg\psi_0\}$. For any finite $0-1$ sequence σ let:

   (a) $T_{\sigma 0} = T_\sigma \cup \{\psi_\sigma\}$
   (b) $T_{\sigma 1} = T_\sigma \cup \{\neg\psi_\sigma\}$.

   We obtain in this way the following full binary tree of extensions of PA:



   This tree has continuum many ($2^{\aleph_0}$) branches. Due to the compactness theorem the union of theories from each branch is consistent (under the assumption of consistency of PA). Further, due to the downward Löwenheim-Skolem theorem each such union has a countable model. No two such models are elementarily equivalent which follows from the construction of the above tree.

   If, for instance, we start with ψ_0 identical with $Con(PA)$ (that is, the sentence expressing the fact that PA is consistent) and the subsequent ψ_α expresses the consistency of T_α, then the model of the leftmost branch of the above tree is the standard model of PA and all other branches will have countable non-standard models. Each sentence of the form $\neg Con(T_\alpha)$ will have a Gödel number which is a non-standard natural number in the respective model.

6.5.2 Overspill

By a *cut* in a non-standard model M of PA one means a subset I of M which has the following closure properties:

1. If $x < y$ and $y \in I$, then $x \in I$.

2. If $x \in I$, then $s(x) \in I$ (meaning that I is closed with respect to the successor operation).

A *proper cut* in a non-standard model M of PA is a cut which is a proper subset of M. Obviously, the standard part of any non-standard model of PA is a proper cut. However, each non-standard model has many other proper cuts. Abraham Robinson proved that proper cuts are not definable and this fact is a consequence of the induction axiom schema. More exactly, the following theorem holds:

> Overspill Lemma. Let M be a non-standard model of PA and let I be a proper cut of M. For any formula $\psi(x, \vec{y})$ with $n+1$ free variables and any n-tuple $\vec{a}$ of parameters (elements of the domain of M), if $M \models \psi[b, \vec{a}]$ for any $b \in I$, then there exists c which is greater than all elements in I and such that $M \models \psi[c, \vec{a}]$.

6.5.3 Tennenbaum's theorem

The following theorem is sometimes used as supporting the claim that the intended model of arithmetic is nothing other than its standard model:

> Tennenbaum's Theorem. *Let $\mathfrak{M}$ be a non-standard model of the first-order Peano arithmetic PA. Then $\mathfrak{M}$ is not recursive.*

The paper by Quinon and Zdanowski uses this kind of argumentation (Quinon and Zdanowski 2006). Let us summarize their views as follows:

1. By a *standard* model of the first-order Peano arithmetic PA they mean a model with ordering of the type ω.

2. A model that reflects our intuitions about numbers adequately is what they call an *intended* model.

3. Now, a natural question that arises here is: are standard models intended models?

4. They make some *cognitive* assumptions:

    (a) The basic feature of natural numbers is that we can count with them.

(b) *The psychological version of Church's thesis*: Any property that humans can compute can be also computed by Turing machines. They write:

> This thesis gives an upper bound on what we may compute. Notice also that from this assumption it follows that in order to distinguish a class of intended models we cannot identify any two isomorphic models, but at most recursively isomorphic models. Indeed, if two models are isomorphic but not recursively isomorphic then their equivalence can be regarded as beyond our cognitive capacities. (Quinon and Zdanowski 2006, 3–4)

5. Now they formulate two postulates:

   (a) Postulate 1. The intended model of arithmetic has to be recursive.

   (b) Postulate 2. The intended model of arithmetic has to satisfy induction (at least first-order induction).

6. Finally, they apply Tennenbaum's theorem claiming that induction together with this theorem describes exactly the class of intended models of arithmetic.

7. In conclusion they write:

> Our three assumptions: computability of basic arithmetical operations, the psychological version of Church's thesis and the principle of induction together with Tennenbaum's theorem result in our main conclusion: the intended model of arithmetic is a recursive model with ω-type ordering. Thus, ω-type ordering is now just a corollary of some more fundamental principles, not a basic requirement for being an intended model of arithmetic. The conditions which we postulate define a proper subclass of the standard models for arithmetic that we call intended models. (Quinon and Zdanowski 2006, 6)

One should also mention that some writers do not accept argumentation like that presented above. Identification of the intended model with the standard model *via* Tennenbaum's theorem faces the problem of definition of recursivity (in either of many equivalent forms, for example Turing machines, primitive recursive functions, Markov's algorithms, Post's systems, or Church λ-calculus). The very idea of *finititude* is inseparably connected with all representations of recursivity. For a more detailed discussion see for example Button and Smith 2012.

One can also add that the standard model of arithmetic is a prime model of PA: it is elementarily embeddable in any other model of PA. From a purely formal point of view the standard model of PA is an exception in the class of all its models. Still, it is the standard model which is identified with the intended model of arithmetic of natural numbers.

6.6 Transfinite induction in set theory

Inductive arguments play a fundamental role not only in arithmetic of natural numbers but also in other domains, notably in set theory. Modern axiomatic set theory uses several forms of induction. The most important are:

1. *Transfinite induction.* Suppose that $\varphi(x)$ is a property defined for all ordinal numbers. Suppose also that whenever $\varphi(\beta)$ holds for all $\beta < \alpha$, then $\varphi(\alpha)$ holds as well. Then $\varphi(x)$ holds for all ordinals x. Proofs by transfinite induction are usually split into three cases:

   (a) proof that $\varphi(0)$ holds;

   (b) proof that $\varphi(\alpha)$ implies $\varphi(\alpha + 1)$

   (c) proof that $\varphi(\lambda)$ holds (for limit ordinals λ) whenever $\varphi(\alpha)$ holds for all $\alpha < \lambda$.

2. *Well-founded induction (Noetherian induction).* Let us suppose that $(X, \prec)$ is a well-founded system and $\varphi(x)$ is a property of sets. In order to prove that $\varphi(x)$ holds for all elements of X it suffices to prove that for any $x \in X$: if $\varphi(y)$ holds for all $y \prec x$, then also $\varphi(x)$ holds. In symbolic form:

   $$\forall x \in X((\forall y \in X(y \prec x \rightarrow \varphi(y))) \rightarrow \varphi(x)) \rightarrow \forall x \in X \varphi(x).$$

   When $\prec$ is taken to be the well ordering of the ordinal numbers, then we obtain the standard transfinite induction mentioned above. The usual mathematical induction on natural numbers is also a special case of well-founded induction. Several forms of *structural induction* are special cases of the well-founded induction as well.

3. $\in$-*induction.* This is a variant of transfinite induction and takes the following form:

$$\forall x(\forall y(y \in x \to \varphi(y)) \to \varphi(x)) \to \forall x \varphi(x).$$

It is equivalent to the axiom of regularity (in the presence of other axioms of ZF):

$$\forall x(x \neq \emptyset \to \exists y(y \in x \wedge x \cap y = \emptyset)).$$

The procedure of *transfinite recursion* that results in the construction of an object (a function) in a transfinite number of steps is closely connected to transfinite induction. The corresponding theorem about transfinite recursion may be formulated in several forms, for instance:

1. Let $g : V \to V$ be a class function from the universe of all sets into itself. Then there exists a unique transfinite $f : Ord \to V$ from the class of all ordinal numbers Ord into V such that: $f(\alpha) = g(f \restriction \alpha)$ for all ordinal numbers α (here $\restriction \alpha$ denotes the restriction of f to the set of all ordinal numbers less than α).

2. Given a set a and class functions f and g there exists a unique function $F : Ord \to V$ such that:

   (a) $F(0) = a$

   (b) $F(\alpha + 1) = f(F(\alpha))$ for all ordinal numbers α

   (c) $F(\lambda) = g(F \restriction \lambda)$ for all limit ordinal numbers $\lambda \neq 0$.

Applications of transfinite induction and transfinite recursion are ubiquitous in set theory. In the next chapter we will meet the cumulative hierarchy of sets as well as the hierarchy of constructible sets, both defined by transfinite induction. Another example of an hierarchy of sets defined in that way is the hierarchy of Borel sets.

The Borel hierarchy of subsets of a set X endowed with some topology is the smallest σ-algebra containing all the open subsets of X. This means that Borel sets can be obtained by taking complements, as well as countable unions and intersections, starting from open sets of the space (alternatively: from closed sets). The family of all Borel sets of the space X can be characterized as follows by transfinite induction up to ω_1, that is up to the first uncountable ordinal:

1. Let $A \subseteq \wp(X)$. We define the following two operations:
   (a) Let A_σ be the family of all countable unions of elements of A.
   (b) Let A_δ be the family of all countable intersections of elements of A.
   (c) Let $A_{\delta\sigma} = (A_\delta)_\sigma$.
2. Now we define a sequence B^α, where α is an ordinal number:
   (a) B^0 is the family of all open subsets of X.
   (b) $B^{\alpha+1} = (B^\alpha)_{\delta\sigma}$
   (c) $B^\lambda = \bigcup_{\alpha<\lambda} B^\alpha$
3. One can prove that the family of all Borel sets of the space X equals B^{ω_1}.

Chapter 7

Two types of extremal axioms in set theory

The axioms of restriction in set theory are minimal extremal axioms. The two most commonly known examples of such axioms are: Fraenkel's *Beschränkheitsaxiom* (*Axiom of Restriction*) and Gödel's *Axiom of Constructibility*. Perhaps less known is Suszko's *Axiom of Canonicity* and Myhill's ideas from Myhill 1952. Also the *Axiom of Limitation of Size* (John von Neumann) expresses a minimality condition, in a sense. The axioms of the existence of large cardinals are maximal extremal axioms. They express the idea that the universe of sets should be as large as possible. Due to the existence of a large amount of sentences independent from the axioms of set theory one may reasonably ask whether some of them could be considered new axioms, characterizing the universe of all sets.

7.1 Introductory remarks

According to a widespread view, modern set theory may serve as the foundations of mathematics. There exist several axiomatic versions of set theory, the most commonly known of which are: ZF – Zermelo-Fraenkel set theory (ZFC is ZF with the axiom of choice) and NBG – von Neumann-Bernays-Gödel set theory. ZF in its standard form is a first-order theory, with two primitive terms: identity $=$ and set membership $\in$. The theory NBG considers sets and classes.

7.1.1 A few words about history

A few words about early set theory are perhaps in order before we turn to extremal axioms in this theory. There are of course numerous elaborate expositions on the history of set theory, for example Ferreirós 1999, Hallet 1984, Moore 2013, or Kanamori 1994. For our present purposes it will be sufficient to mention the following events:

1. It is justified to assume that *mathematical* set theory originated from the works of Cantor and Dedekind. Earlier reflections on sets and infinite collections were not systematic in their character, though they might of course have influenced the views of Cantor and Dedekind. This concerns for example the ideas of Bolzano and Riemann.

2. In 1871 Dedekind in fact used a set-theoretical approach in his work on algebraic number theory. He called certain subsets of the ring of integers in a field of algebraic numbers *ideals*, and treated them as a new kind of objects.

3. In 1872 Dedekind and Cantor, independently, proposed constructions of the set of real numbers. They both started with the set of rational numbers and made use of infinite subsets of it (either in the form of cuts or in the form of Cauchy sequences).

4. In 1872 Cantor introduced the operation of derivative of a point set, admitting that this operation could be iterated in the transfinite (that is, beyond any finite number of steps).

5. In 1873, due to the efforts of Dedekind and Cantor, it was proved that the set of algebraic numbers is countable and that the set of real numbers is not countable.

6. From 1878 to 1885 Cantor has published several works concerning main concepts of set theory. He developed ideas concerning transfinite numbers, formulated the continuum hypothesis, defined certain important topological concepts, investigated well-ordered sets. A little bit later, in 1892, he presented his famous diagonalization method.

7. At the turn of the 20th century set theory experienced a crisis; several antinomies had been discovered, showing that the very concept of set required clarification.

8. Even in the pre-axiomatic stage of set theory there was serious debate about the admissibility of some of its fundamental assumptions, notably the axiom of choice.

9. Around 1900, several mathematicians contributed to the development of descriptive set theory, among others Borel, Baire, Bernstein, Lebesgue, and Vitali. A little later Suslin, Luzin, Sierpiński and some Polish mathematicians were also active in this domain.

10. Hausdorff's *Grundzüge der Mengenlehre* (Hausdorff 1914) was an influential textbook in set theory and general topology.

11. In the first decades of the 20th century, important results concerning cardinal arithmetic were obtained by König, Hausdorff, Sierpiński, Ulam, and Tarski.

Many results obtained in the pre-axiomatic stage of set theory served as guidelines for its axiomatization.

7.1.2 Axiomatic Zermelo-Fraenkel set theory ZFC

The axioms of ZFC determine the meaning of the primitive notions and characterize operations which can be applied to sets. Let us recall these axioms, together with intuitive comments.

Extensionality

$\forall x \forall y\ (\forall z\ (z \in x \equiv z \in y) \to x = y)$

This axiom says that each set is uniquely determined by its elements. Two sets are equal if and only if they consist of exactly the same elements.

Pair

$\forall x \forall y \exists z \forall u\ (u \in z \equiv (u = x \vee u = y))$

This axioms implies the existence of an unordered pair of sets.

Union

$\forall x \exists y \forall z\ (z \in y \equiv \exists u\ (z \in u \wedge u \in x))$

This axioms implies the existence of the union of a family of sets.

Powerset

$\forall x \exists y \forall z\ (z \in y \equiv \forall u (u \in z \to u \in x))$

According to this axiom, for any set there exists the family of all its subsets.

Separation schema

$\forall x_1 \forall x_2 \ldots \forall x_n \forall y \exists z \forall u\ (u \in z \equiv (u \in y \wedge \varphi(u, x_1, x_2, \ldots, x_n)))$
where φ is a formula of the language of ZF theory and such that z is not a free variable in φ and $x_1, x_2, \ldots, x_n$ are free variables of φ different from the variable u.

According to the schema of separation, given any set one can form a subset of it consisting of exactly those elements which have a given property expressible in the language of ZF. This schema is a countable list of axioms.

Infinity

$\exists x\ (\exists y\ (y \in x \wedge \neg \exists z\ (z \in y)) \wedge \forall y\ (y \in x \rightarrow \forall z (\forall u\ (u \in z \equiv u = y) \rightarrow z \in x)))$

The axiom of infinity states the existence of at least one infinite set.

Replacement schema

$\forall u (\forall x \forall y \forall z\ (x \in u \wedge \varphi(x, y) \wedge \varphi(x, z) \rightarrow y = z) \rightarrow \exists w \forall v\ (v \in w \equiv \exists x\ (x \in u \wedge \varphi(x, v))))$

Intuitively speaking, according to the schema of replacement the image of a given set with respect to a function (described by a formula of the language of ZF) is again a set. This schema is a countable list of axioms.

Foundation

$\forall x (\exists u\ (u \in x) \rightarrow \exists y (y \in x \wedge \forall z\ (z \in y \rightarrow \neg z \in x)))$

The axiom of foundation excludes the existence of infinite $\in$-descending chains of sets, meaning sequences $(x_1, x_2, x_3, x_4, \ldots)$, such that:

$$x_2 \in x_1,\ x_3 \in x_2,\ x_4 \in x_3, \ldots$$

Choice

$\forall x ((\forall y\ (y \in x \rightarrow \exists z\ (z \in y)) \wedge \forall y \forall u\ ((y \in x \wedge u \in x) \rightarrow y = u \vee \neg \exists v\ (v \in y \wedge v \in u))) \rightarrow \exists w (\forall y\ (y \in x \rightarrow \exists z\ ((z \in y \wedge z \in w) \wedge \forall v\ ((v \in y \wedge v \in w) \rightarrow v = z)))))$

According to the axiom of choice, for any non-empty family of non-empty pairwise disjoint sets there exists a set which has exactly one member in common with each set of this family.

Identity axioms

1. $\forall x\ (x = x)$
2. $\forall x \forall y\ (x = y \rightarrow y = x)$

3. $\forall x \forall y \forall z\, ((x = y \wedge y = z) \rightarrow x = z)$;
4. $\forall x \forall y \forall z\, ((x = y \wedge x \in z) \rightarrow y \in z)$;
5. $\forall x \forall y \forall z\, ((x = y \wedge z \in x) \rightarrow z \in y)$.

The terms *infinite* and *countable* used in the comments above belong to the metalanguage.

The cumulative hierarchy. The full universe of sets is defined by transfinite induction involving the powerset operation at successor levels and union operation at limit levels:

1. $V_0 = \emptyset$
2. $V_{\alpha+1} = \wp(V_\alpha)$
3. $V_\lambda = \bigcup_{\alpha<\lambda} V_\alpha$ for limit ordinals λ
4. $V = \bigcup_{\alpha \in Ord} V_\alpha$

If set theory ZFC is consistent, then it has a model. According to the Löwenheim-Skolem theorem if it has a model, then it also has a countable model. We have recalled some results concerning the metatheory of set theory in chapter 3. In particular, ZFC is undecidable and it cannot prove its own consistency.

There are many statements which are known to be independent from the axioms of set theory. Is it reasonable in this situation to ask what the intended model of set theory is? This question was not asked at the very beginning of set theory, in the works of Cantor, Hausdorff, or Sierpiński, or by other mathematicians who developed the pre-axiomatic set theory. The first metatheoretical reflections concerning models of set theory were formulated much later by John von Neumann (von Neumann 1925), Abraham Fraenkel (Fraenkel 1928), Ernst Zermelo (Zermelo 1930), and Kurt Gödel (Gödel 1940).

As far as extremal axioms in set theory are concerned, there are differences and similarities between them and the axioms proposed for natural numbers (the axiom of induction – cf. chapter 6) and for real numbers (the continuity axiom – cf. chapter 5):

1. The axiom of induction is an extremal axiom expressing the idea that the universe of natural numbers should be minimal – these numbers should constitute the smallest infinity.

2. The axiom of continuity is an extremal axiom expressing the idea that the universe of real numbers should be maximal: the real numbers should constitute the maximal (Archimedean) completely ordered field. Similar maximality condition can be formulate in the case of the (non-Archimedean) field of surreal numbers.

3. In the case of set theory we notice an interesting change of perspective. Initially, extremal axioms for set theory were supposed to express the idea that the universe of all sets should be as minimal as possible (see Fraenkel's axiom of restriction and Gödel's axiom of constructibility). But then the attitude changed and since then logicians and mathematicians working in logic and the foundations of mathematics have taken the standpoint according to which the universe of all sets should be as rich as possible. One way of achieving this goal is to assume axioms stating the existence of large cardinal numbers.

7.2 Zermelo: two axiomatizations of set theory

After the first period of development of set theory (sometimes called *naive set theory*) the time arrived for its axiomatic settlement. This was done for the first time by Ernst Zermelo in Zermelo 1908. As is well known, Zermelo gave the first proof of the theorem about the possibility of well ordering of any set back in 1904. In his work from 1908 he gave a new proof, this time with explicit use of the axiom of choice. Over the next few years this first axiom system underwent a few improvements, credited to Abraham Fraenkel and Thoralf Skolem. An independent axiomatization was proposed by John von Neumann in 1925.

7.2.1 First axiomatization: Zermelo 1908

Let us write down the axioms for set theory proposed by Zermelo in 1908. It will be interesting to compare this list with the second axiomatization, proposed by Zermelo in 1930. We cite the English translation proposed in Zermelo 2010, 193–201:

> 1. *Axiom of extensionality.* If every element of a set M is also an element of N and vice versa, if, therefore $M \subseteq N$ and $N \subseteq M$, then always $M = N$; or, more briefly: Every set is determined by its elements.

2. *Axiom of elementary sets.* There exists a (fictitious) set, the "*null set*" 0, that contains no element at all. If a is any object of the domain, there exists a set $\{a\}$ containing a and only a as element; if a and b are any two objects of the domain, there always exists a set $\{a, b\}$ containing as elements a and b but no object x distinct from both.
3. *Axiom of separation.* Whenever the propositional function $\mathfrak{C}(x)$ is definite for all elements of a set M, M possesses a subset $M_{\mathfrak{C}}$ containing as elements precisely those elements x of M for which $\mathfrak{C}(x)$ is true.
4. *Axiom of the power set.* To every set T there corresponds another set $\mathfrak{U}T$, the "*power set*" of T, that contains as elements precisely all subsets of T.
5. *Axiom of the union.* To every set T there corresponds another set $\mathfrak{S}T$, the "*union*" of T, that contains as elements precisely all elements of the elements of T.
6. *Axiom of choice.* If T is a set whose elements all are sets that are different from 0 and mutually disjoint, its union $\mathfrak{S}T$ includes at least one subset S_1 having one and only one element in common with each element of T.
7. *Axiom of infinity.* There exists in the domain at least one set Z that contains the null set as an element and is so constituted that to each of its elements a there corresponds a further element of the form $\{a\}$, in other words, that with each of its elements a it also contains the corresponding set $\{a\}$ as an element.

Observe that the axiom of separation refers to "definite" propositional functions. Zermelo characterized this notion as follows:

> A question or assertion $\mathfrak{C}$ is said to be "*definite*" if the fundamental relations of the domain, by means of the axioms and the universally valid laws of logic, determine without arbitrariness whether it holds or not. Likewise a "propositional function" $\mathfrak{C}(x)$, in which the variable term x ranges over all individuals of a class $\mathfrak{K}$, is said to be "definite" if it is definite for *each single* individual x of the class $\mathfrak{K}$. Thus the question whether $a \varepsilon b$ or not is always definite, as is the question whether $M \subseteq N$ or not. (Zermelo 2010, 193)

7.2.2 Second axiomatization: Zermelo 1930

For our purposes the second axiomatization given by Zermelo is of greatest interest, and namely that proposed in his paper *Über Grenzzahlen und*

Mengenbereiche: Neue Untersuchungen über die Grundlagen der Mengenlehre, published in 1930 in *Fundamenta Mathematicae.*

In this second axiomatization Zermelo introduced his normal domains (*Normalbereiche*), proved isomorphism theorems characterizing these domains, and postulated the necessity of considering a transfinite hierarchy of (strongly) inaccessible cardinals. Zermelo knew the work Skolem 1922, and as such was aware of the effects of the Löwenheim-Skolem theorem. He also knew von Neumann's axiomatization of set theory from 1925. He accepted the necessity of addition of the replacement axiom due to the earlier remarks by Fraenkel. However, his separation axiom is not formulated in the style of Skolem, meaning that it is not restricted to first-order formulas.

The system from 1930 has the following axioms (citing the translation of Zermelo 1930 in Zermelo 2010, 403):

> B) *Axiom of extensionality*: Every set is determined by its elements, provided that it has any elements at all.
>
> A) *Axiom of separation*: Every propositional function $\mathfrak{f}(x)$ separates from every set m a subset $m_{\mathfrak{f}}$ containing all those elements x for which $\mathfrak{f}(x)$ is true. Or: To each part of a set there in turn corresponds a set containing all elements of this part.
>
> P) *Axiom of pairing*: If a and b are any two elements, then there is a set that contains both of them as its elements.
>
> U) *Axiom of the power set*: To every set m there corresponds a set $\mathfrak{U}m$ that contains as elements all subsets of m, including the null set and m itself. Here, an arbitrarily chosen "urelement" $\mathfrak{u}_0$ takes the place of the "null set".
>
> V) *Axiom of the union*: To every set m there corresponds a set $\mathfrak{S}m$ that contains the elements of its elements.
>
> E) *Axiom of replacement*: If the elements x of a set m are replaced in a unique way by arbitrary elements x' of the domain, then the domain contains also a set m' that has as its elements all these elements x'.
>
> F) *Axiom of foundation*: Every (decreasing) chain of elements, in which each term is an element of the preceding one, terminates with finite index at an urelement. Or, what amounts to the same thing: Every partial domain T contains at least one element t_0 that has no element in T.

Zermelo added the following footnote to the formulation of the axiom of separation:

> Like the replacement function in E), the propositional function $\mathfrak{f}(x)$ can be completely *arbitrary* here, all consequences of restricting it to a particular class of functions cease to apply from the present point of view. I shall consider elsewhere more thoroughly "the question of definiteness" in connection with my last contribution to this journal (*Fund. Math.* vol. XIV, pp. 339–344) and with the critical "remarks" by Mr. Th. Skolem (*Fund. Math.* vol. XV, pp. 337–341). (Zermelo 1930, 31, citing the translation in Zermelo 2010, 403)

Zermelo accepted the existence of urelements or atoms (*Urelemente*), that is objects which are not sets: they do not have elements but may be elements of other sets. Let the totality of urelements be denoted by Q.

The separation axiom (schema) is formulated – from the modern point of view – in a second-order language. Zermelo does not specify the language of the theory explicitly; he writes that an arbitrary propositional function may determine a subset of a given set. Similarly, no linguistic restrictions are imposed on functions in the axiom schema of replacement.

The axiom of foundation is a new axiom of this system and Zermelo writes that is is necessary in order to exclude "circular", "self-membered" and "rootless" sets ("zirkelhaften", "sich selbst enthaltenden", "wurzellosen" Mengen).

The axiom of infinity is not assumed in this system – according to Zermelo it does not belong to the general set theory (*"allgemeine" Mengenlehre*). Zermelo shows that the normal domain consisting of all finite sets satisfies the axioms of the system.

The axiom of choice is assumed as a *logical principle* and it is not included as an axiom specific for the theory. According to Zermelo, this axiom should not be used for the restriction of the size of the investigated domains (*Abgrenzung der Bereiche*). The fact that each set can be well-ordered is fundamental to all of his further considerations.

Systems consisting of sets and atoms, with the fundamental relation $a \epsilon b$ (a is an element b) which satisfy the axioms BAPUVEF Zermelo calls normal domains (*Normalbereiche*). One can take sums and intersections of such domains and investigate the relations between them (for example the relation of being a subdomain). Particular normal domains are not sets in any absolute sense: if P_1 *is an element* of P_2, then P_1 is a set in P_2. Domains which are sets in this sense are called closed. By a *basis* of a normal domain one understands the set of its atoms. *Characteristic* of a normal domain is the smallest ordinal number which is not a set in this domain.

Zermelo introduces the *boundary numbers* (*Grenzzahlen*) as the fixed points of some function defined by transfinite induction. They correspond to the (strongly) inaccessible numbers in modern terminology.

The set-theoretical world is represented in the form of a cumulative hierarchy of domains, similar to the one investigated by von Neumann, but starting from the level of urelements. The internal structure of normal domains is characterized by three "development theorems" – let us give here the first of them, with Zermelo's own commentary:

> *First development theorem. Each normal domain P of characteristic π can be decomposed into a well-ordered [[sequence]] of type π of non-empty and disjoint "layers" Q_α, so that each layer Q_α includes all elements of P which occur in no earlier layer and whose elements belong to the corresponding "segment" P_α, that is, to the sum of preceding layers. The first layer Q_0 includes all the urelements.*
>
> For the partial domains, or "segments", P_α are defined by transfinite induction by virtue of the following stipulations:
>
> 1. $P_1 = Q_0 = Q$ shall include the whole basis, the totality of urelements.
> 2. $P_{\alpha+1} = P_\alpha + Q_\alpha$ shall contain all sets of P that are "rooted" in P_α, that is all those sets whose elements lie in P_α.
> 3. If α is a limit number, then P_α shall be the sum or union of all preceding P_β with smaller indices $\beta < \alpha$.
>
> (Zermelo 1930, 36; citing the translation of Zermelo 1930 in Zermelo 2010, 413)

If α is a "boundary number", or in other words a strongly inaccessible cardinal, then P_α satisfies all the axioms BAPUVEF.

Zermelo gives theorems characterizing the normal domains up to isomorphism, taking into account two parameters: the power of the basis and the characteristic of the domain. The statements of these theorems are as follows:

> *First isomorphism theorem. Two normal domains with the same characteristic and with equivalent bases are isomorphic. In fact, the isomorphic mapping of the two domains onto one another is* uniquely *determined by the mapping of their bases.*

> *Second isomorphism theorem. Given two normal domains with equivalent bases and different boundary numbers π, π', one is always isomorphic to a canonical segment of the other.*
>
> *Third isomorphism theorem. Given two normal domains with the same characteristic, one is always isomorphic to a (proper or improper) subdomain of the other.* (Zermelo 1930, 36–39, citing the translation in Zermelo 2010, 421–423)

At the end of this article Zermelo expresses a suggestion concerning a possible extension of his axiom system:

> Let us now put forth the general hypothesis that *every categorically determined domain can also be conceived as a "set" in one way or another*; that is that it can occur as an element of a (suitably chosen) normal domain. It then follows that there corresponds to any normal domain a higher one with the same basis, to any unit domain a higher unit domain, and therefore also to any "boundary number" π a greater boundary number π'. Likewise, a categorically determined domain of sets arises through union and fusion from every infinite sequence of different normal domains with common basis, where one always contains the other as a canonical segment. This categorically determined domain of sets can then again be supplemented so as to become a normal domain of higher characteristic. Thus, to every categorically determined totality of "boundary numbers" there follows a greater one, and the sequence of "all" boundary numbers is as unlimited as the number series itself, allowing for the possibility of associating to every transfinite index a particular boundary number in one-to-one fashion. Once again, this is of course *not* "provable" on the basis of the ZF' axioms, since the asserted behavior leads us beyond any individual normal domain. Rather, we must postulate the *existence of an unlimited sequence of boundary numbers* as a new axiom for the "meta-theory of sets", where the question of "consistency" still requires closer examination. (Zermelo 1930, 46, citing the translation in Zermelo 2010, 429)

As we can see from this quotation, Zermelo's suggestion is one of the first proposals to postulate the necessity of considering large cardinal axioms in set theory.

Zermelo also stresses that one should not think of set theory as a theory describing one intended model. He points to the fact that the normal domains corresponding to different strongly inaccessible levels of the cumulative hierarchy are not isomorphic. This, in turn, he considers as a virtue of his system of set theory, reflecting the intuitions behind the concept of a set.

7.3 Axioms of restriction

Axioms of restriction in set theory express the idea that the universe of all sets is, in a specified sense, “small” or “narrow”. How may such conditions be formulated in a reasonable way? One may think for instance about universe of sets as closed under specified a priori operations. Or one may assume that the only subsets of a given set are those which are definable (with parameters) in the language of the theory.

7.3.1 Fraenkel’s axiom of restriction

Fraenkel’s axiom of restriction (Axiom der Beschränkheit) says, roughly speaking, that there are no more sets than those whose existence follows from the axioms of set theory. Its first formulation can be found in Fraenkel’s article *Zu den Grundlagen der Cantor-Zermeloschen Mengenlehre* (Fraenkel 1922, 234). Fraenkel notices that Zermelo’s system of set theory from 1908 does not exclude two types of sets which in his words are irrelevant for mathematical purposes. The first type is that of sets consisting of physical elements and second corresponds to the non-well-founded sets. Fraenkel also remarks that this fact is responsible for the non-categorical character of the set theory in question: the theory does not determine the totality of sets completely.

Fraenkel was of course fully aware of the fact that his axiom of restriction was not a “normal” axiom belonging to the object language. He expressed doubts whether its exact reformulation was at all possible. Fraenkel also stressed the role of the *inductive* component implicitly contained in the axiom: it refers to a number of operations applicable to sets.

Fraenkel returned to the idea of restriction of the set-theoretical universe several times in his later works (see Fraenkel 1928, Fraenkel and Bar-Hillel 1958). The informal formulation of the axiom can be improved in several ways (see for example Carnap 1954, 154). The idea of minimality can be for instance expressed in terms of the algebraic closure with respect to set-theoretical operations. Georg Schiemer proposed a very detailed analysis of Fraenkel’s results concerning the axiom of restriction in Schiemer 2010a, 2010b.

Fraenkel formulated three possible ways of understanding *completeness* of set theory (Fraenkel 1928, 347–354):

1. completeness in the sense of Hilbert from his *Grundlagen of Geometrie*;

2. completeness understood as a property that given any sentence ψ of the language of set theory either ψ or its negation $\neg\psi$ is a theorem of set theory (the deductive completeness in the modern sense of the term);

3. completeness understood as categoricity (monomorphy).

Fraenkel's ideas concerning the notions of completeness and the axiom of restriction were developed by Carnap (see Carnap 1930, Carnap 2000, Carnap and Bachmann 1936). We have discussed these works in chapter 2.

As far as the exclusion of not well-founded sets from the set-theoretical universe is concerned, observe that the axiom of restriction is not necessary for that, it suffices to assume the axiom of regularity (axiom of foundation), as in the proposals in von Neumann 1925 and Zermelo 1930.

Georg Schiemer has pointed out that it is likely that Fraenkel drew inspiration for his axiom of restriction from Dedekind's considerations about natural numbers, and especially from his *Kettentheorie* (Schiemer 2010a, 2010b). The set of all natural numbers is a minimal chain; similarly the universe of all sets should be in some sense *minimal*.

Fraenkel's axiom of restriction was already criticized at the time, shortly after its formulation, for example in von Neumann 1925 and Zermelo 1930. A sharp critique of the axioms of restriction in general is presented in Fraenkel, Bar-Hillel and Levy 1973 (see below).

7.3.2 Axiom of the limitation of size

It was observed already in naive set theory that there is no greatest cardinal number and that the totality of all cardinal numbers cannot form a set. A simple argument based on Cantor's theorem (no set is equinumerous with the family of all its subsets) shows that if κ is the greatest cardinal, then 2^κ is greater than κ. Furthermore, if X is a set containing sets of all cardinalities, then every element of X is a subset of $\bigcup X$ and hence its cardinality is less or equal to that of $\bigcup X$. Again, according to the Cantor's theorem, $2^{|\bigcup X|}$ is a cardinal greater than the cardinal $|\bigcup X|$ which shows that X cannot contain sets of all cardinalities. Thus the totality of all cardinal numbers is "too big" to be a set. The same concerns the totality of all ordinal numbers.

The idea that some collections are too big to be sets was therefore well known at the very beginning of set theory, and the *limitation of size principle* was a heuristic principle which recommended the exclusion of such collections from the universe of all sets. But there remained the problem of how to express this informal idea in a precise form.

The axiom of the limitation of size is an axiom added to the main body of axioms in systems of set theory which deal with sets and classes – like the von Neumann-Bernays-Gödel (NBG) system or the Morse-Kelley system. It captures the idea that any object in the domain of the theory which is "too big" to be a set is similar to the class of all sets, in the sense that there exists a bijection between it and the class of all sets.

The axiom of the limitation of size was proposed by John von Neumann (von Neumann 1925). Some of the relationships between this axiom and other axioms of the system of sets and classes are as follows (these results were proved by von Neumann himself and by Azriel Levy):

1. The axiom of the limitation of size implies the axiom of replacement (and consequently also the axiom of separation).

2. The axiom of the limitation of size implies that the class V of all sets can be well ordered.

3. On the basis of NBG the axiom of the limitation of size is implied by the conjunction of the axioms of replacement, global choice and union.

John von Neumann wrote in 1923 a letter to Zermelo in which he stated the first version of his axiom of the limitation of size: a class is a proper class if and only if there is a one-to-one correspondence between it and the class V of all sets. In the system of set theory proposed in Zermelo 1930 (see above) the author defines the transfinite hierarchy (similar to the cumulative hierarchy V) and his *normal domains* correspond to those P_κ for which κ is a strongly inaccessible cardinal. For Zermelo, sets in a normal domain P_κ are those X for which $X \in P_\kappa$ and classes are those X for which $X \subseteq P_\kappa$ but $X \notin P_\kappa$.

All axioms of ZFC are in accordance with the limitation of size principle, with the exception of the axiom of infinity and the powerset axiom. The principle of the limitation of size is discussed in full detail in the monograph Hallet 1984.

7.3.3 Gödel's axiom of constructibility

The axiom of constructibility was not conceived as an axiom of restriction, although it has the form of the axiom of minimality in set theory.

Gödel's work on the constructible universe aimed primarily to prove the consistency of the axiom of choice and the continuum hypothesis with the remaining axioms of Zermelo-Fraenkel set theory (ZF). That is to say, the inner model of all constructible sets was devised in order to prove that if ZF set theory is consistent, then ZF plus the axiom of choice and the continuum hypothesis is also consistent.

We recall that at successor stages in building the constructible universe one makes use of the poorest powerset operation possible: the powerset of X contains only *definable* subsets of X. At limit stages we take of course unions of all stages constructed so far. The class of all constructible sets is a minimal countable transitive model of set theory containing all ordinal numbers. Precise definitions are as follows:

$$Def(X) = \{\{x \in X : (X, \in) \models \varphi(x, y_1, \ldots, y_n)\} : \varphi \text{ is a formula of the language of set theory and } y_1, \ldots, y_n \in X\}.$$

Thus $Def(X)$ is the family of all subsets of X definable by a formula from the language of set theory with parameters belonging to X. The constructible hierarchy L is then defined by transfinite induction in the following way:

1. $L_0 = \emptyset$
2. $L_{\alpha+1} = Def(L_\alpha)$
3. $L_\lambda = \bigcup_{\alpha<\lambda} L_\alpha$ for limit ordinals λ
4. $L = \bigcup_{\alpha \in Ord} L_\alpha$

Elements of the class L are called *constructible sets*. The successor levels of the hierarchy are defined with the help of a semantic notion (definability). However, due to the normal form theorem proved by Gödel there is an algebraic counterpart of this notion, based on so-called *Gödel's operations*. Gödel himself used eight such operations, and in modern presentations of the constructible hierarchy (for example Jech 2003) the following operations are used:

1. $G_1(X, Y) = \{X, Y\}$
2. $G_2(X, Y) = X \times Y$
3. $G_3(X, Y) = \{(x, y) : x \in X \wedge y \in Y \wedge x \in y\}$
4. $G_4(X, Y) = X - Y$

5. $G_5(X,Y) = X \cap Y$
6. $G_6(X) = \bigcup X$
7. $G_7(X) = dom(X)$
8. $G_8(X) = \{(x,y) : (y,x) \in X\}$
9. $G_9(X) = \{(x,y,z) : (x,z,y) \in X\}$
10. $G_{10}(X) = \{(x,y,z) : (y,z,x) \in X\}$

The normal form theorem states that for any formula $\varphi(x_1,\ldots,x_n)$ in which all quantifiers are bounded the set

$$\{(x_1,\ldots,x_n) \in X_1 \times \ldots \times X_n : \varphi(x_1,\ldots,x_n)\}$$

can be obtained as a value of a function being a composition of the above functions G_1–G_{10} applied to the arguments $X_1,\ldots,X_n$. Some properties of the hierarchy of constructible sets are as follows:

1. $L_n = V_n$ for all finite ordinal numbers n.
2. For any ordinal number α, $L_\alpha = \bigcup_{\beta<\alpha} L_\beta$.
3. For any infinite ordinal number α there exists a (constructible) bijection between α and L_α.
4. Each successor level $L_{\alpha+1}$ of the constructible hierarchy can be obtained as the intersection of the powerset of L_α and the closure of $L_\alpha \cup \{L_\alpha\}$ under Gödel's operations mentioned above.
5. All arithmetical subsets of ω and relations on ω belong to $L_{\omega+1}$ and any subset of ω which belongs to $L_{\omega+1}$ is arithmetical.
6. The set of codes (for example Gödel numbers) of sentences true in the standard model of arithmetic (which is not an arithmetical subset of ω) belongs to $L_{\omega+2}$.
7. All hyperarithmetical subsets of ω and relations on ω belong to $L_{\omega_1^{CK}}$. Any subset of ω which belongs to $L_{\omega_1^{CK}}$ is hyperarithmetical. Here ω_1^{CK} is the Church-Kleene ordinal number, that is the smallest non-recursive ordinal.

The axiom of constructibility is the sentence $\forall x \exists \alpha\ x \in L_\alpha$ which is often abbreviated as $V = L$. Among the consequences of ZF plus $V = L$ are the following facts:

1. $V = L$ implies in ZF the axiom of choice.

2. $V = L$ implies in ZF the generalized continuum hypothesis.

3. $V = L$ implies in ZF the negation of the Suslin hypothesis.

4. $V = L$ implies in ZF the existence of a Δ^1_2 non-measurable set of real numbers.

Model theoretic properties of the class of all constructible sets include the following:

1. L is a *standard* model of ZF, which is to say that the class L is transitive and the relation $\in$ in it is the real membership relation. As a consequence, L is well-founded.

2. L contains all ordinal numbers from V. Actually, L is the smallest standard model of ZF containing all ordinal numbers.

3. If there exists in V a set W which is a standard model of ZF such that the ordinal number κ is the set of all ordinal numbers occurring in W, then L_κ coincides with the class of all sets constructible in W. If there exists a set which is a standard model of ZF, then the smallest such set is exactly such L_κ and is called the *minimal* model of ZF.

The method of inner models has its own limitations, as shown in Shepherdson 1951–1953. However, it is a very convenient point of departure for some more subtle constructions, including the celebrated method of forcing, credited to Paul Cohen.

Kurt Gödel himself was later opposed to axioms of restriction in set theory and he openly expressed his view in favor of axioms of maximality:

> On the other hand, from an axiom in some sense opposite to this one [meaning to the axiom of constructibility – JP], the negation of Cantor's conjecture could perhaps be derived. I am thinking of an axiom which (similar to Hilbert's completeness axiom in geometry) would state some maximum property of the system of all sets, whereas axiom A [the axiom of constructibility – JP] states

> a minimum property. Note that only a maximum property would seem to harmonize with the concept of set explained in footnote 14. (Gödel 1964, citing Gödel's *Collected Works* II, 262–263)

It seems that nobody in the community of set theoreticians has ever seriously taken into account the possibility of adjoining the axiom of constructibility to the body of fundamental axioms of set theory. "Normal" mathematicians may have a different opinion in this respect; compare Friedman's judgment:

> The set theorist is looking for deep theoretic phenomena, and so $V = L$ is anathema since it restricts the set theoretic universe so drastically that all sorts of phenomena are demonstrably not present. Furthermore, for the set theorist, any advantage that $V = L$ has in terms of power can be obtained with more powerful axioms of the same rough type that accommodate measurable cardinals and the like – e.g., $V = L(\mu)$, or the universe is a canonical inner model of a large cardinal.
>
> However, for the normal mathematician, since set theory is merely a vehicle for interpreting mathematics as to establish rigor, and not mathematically interesting in its own right, the less set theoretic difficulties and phenomena the better.
>
> I.e., less is more and more is less. So if mathematicians were concerned with the set theoretic independence results – and they generally are not – then $V = L$ is by far the most attractive solution for them.
>
> This is because it appears to solve all set theoretic problems (except for those asserting the existence of sets of unrestricted cardinality), and is also demonstrably relatively consistent.
>
> Set theorists also say that $V = L$ has implausible consequences – e.g., there is a PCA well ordering of the reals, or there are non-measurable PCA sets.
>
> The set theorists claim to have a direct intuition which allows them to view these as so implausible that this provides "evidence" against $V = L$.
>
> However, mathematicians disclaim such direct intuition about complicated sets of reals. Some say they have no direct intuition about all multivariate functions from $\mathbb{N}$ into $\mathbb{N}$! (Feferman, Maddy, Steel and Friedman 2000, 436–437)

Nevertheless, the axiom of constructibility, taken as a working assumption, has many consequences of considerable interest, in combinatorics, algebra, model theory, and theory of recursive functions, and so on. However, the axiom of constructibility implies for example the nonexistence of measurable cardinals as well as the negation of the Suslin hypothesis. The price to be paid, if one accepts this axiom seems, too high in relation to its alleged naturalness and evident economy. We prefer to stay in *Cantor's Paradise.*

Further investigations of the constructible universe include its *fine structure* (see Jensen 1972).

7.3.4 Suszko's axiom of canonicity

The two axioms of restriction discussed above are very well known. The next one is relatively less known, it seems. Roman Suszko wrote his *Canonic axiomatic systems* (Suszko 1951) with the goal of explicating Skolem's (alleged) paradox. He also stressed that his explication did not refer to the Löwenheim-Skolem theorem itself. Suszko described his *axiom of canonicity* as a formal counterpart of Fraenkel'a axiom of restriction.

In literature, Skolem's paradox is usually linked to the (downward) Löwenheim-Skolem theorem, which states that – in modern formulation – any theory in a first-order language which has at least one (infinite) model also has a countable model. In the case of first-order set theory and in the presence of Cantor's theorem this fact may seem paradoxical: how could it be possible that uncountable sets exist in a countable domain? The resolution of this alleged paradox is quite simple and shows that the notion of countability is not absolute in first-order logic. A countable model of set theory satisfies Cantor's theorem stating the existence of an uncountable set X, but the model itself does not contain a bijection between X and the set of all natural numbers. Thus X is countable from the outside point of view (in metatheory) but is uncountable from the point of view of the model itself. This explication can be found in many places in literature. The second aspect of the paradox is connected with the fact that in standard first-order theories we have only countable means of expression – in particular, there are only countably many closed terms which can name objects in the domain of a model. This second aspect is investigated in the work in question by Suszko.

Suszko quotes the following passage from Carnap's *The logical syntax of language* (Carnap 1937) concerning Skolem's paradox (Carnap translates Fraenkel's Axiom der Beschränkheit as Axiom of Limitation):

> Let us take as object-language S the system of axioms used in Fraenkel's Theory of Aggregates supplemented by a sentential and

> functional calculus (in the word-language). The theorem that more than one transfinite cardinal exists depends upon the theorem that the aggregate $U(M)$ of the sub-aggregates of an aggregate M has a higher cardinal number than has M; this theorem is based upon what is known as Cantor's theorem, which maintains that M and $U(M)$ cannot have the same cardinal number. Fraenkel has given a proof of this theorem which remains valid for his system S even though it contains the so-called Axiom of Limitation. On the other hand, however, we arrive at the contrary result as a consequence of the following argument. The Axiom of Limitation means that in the aggregate-domain which is treated in S – let us call it B – only those aggregates occur of which the existence is required by the other axioms. Therefore, only the following aggregates are existent in B: in the first place, two initial aggregates, namely, the null-aggregate and the denumerably infinite aggregate, Z, required by Axiom VII; and secondly, those aggregates which can be constructed on the basis of these initial aggregates by applying an arbitrary but finite number of times certain constructional procedures. There are only six kinds of these constructional steps (namely, the formation of the pair-aggregate, of the sum-aggregate, of the aggregate of sub-aggregates, of the aggregate of Aussonderung, of the aggregate of selection, and of the aggregate of replacement). Since only a denumerable multiplicity of aggregates can be constructed in this way, there is in B, according to the Axiom of Limitation, only a denumerable multiplicity of aggregates, and consequently, at the most, only a denumerable multiplicity of sub-aggregates of Z. Therefore $U(Z)$ cannot have a higher cardinal number than Z. (Carnap 1937, 267–268, Suszko 1951, 302–303)

Suszko uses logical syntax described in his doctoral dissertation Suszko 1949. Each mathematical theory deals with objects from some domain and those objects can be either nameable or not nameable in the language of the theory. *Categorematic* names (k-names) are those expressions which can be built from individual constants with the use of functor symbols – in modern terminology they are called closed terms. The relation of k-designation is the relation of designation by a k-name. A *constructible object* (in a given system X) is any object from the universe of the system X which is k-designated by a k-name. If the universe of a system X consists of exactly those objects which are constructible in X, then X is called a *canonic system.* Obviously, the universe of any canonic system is at most countable.

Suszko begins with an axiomatic system M of set theory with classes, resembling the system proposed by Bernays (see Bernays 1937–1948). He also uses some ideas from the works Gödel 1940 and Quine 1941, 1946. The system M has two primitive terms: $\in$ and El. The expression $a \in b$ is to be read: a is an element of b and $\mathrm{El}(a)$ is to be read: a is an element. Thus, we talk in M about classes, some of which can be elements of other classes – in this case they are sets.

The author's attention is focused on the concepts of denumerability and non-denumerability. An expression of the form $\Phi_x[a, b, \psi(x)]$ is constructed whose meaning is: the relation a establishes a bi-unique correspondence between the set of all those non-empty subsets x of the set b which are fulfilling the condition $\psi(x)$. Thus, the expression $\neg\exists y\ \Phi_x[y, b, \psi(x)]$ is to be read: the set of those non-empty subsets x of the set b which are fulfilling the condition $\psi(x)$ is non-denumerable. In particular, the following are theorems of the system M (here V stands for the class of all sets and ω_0 for the set of all finite ordinal numbers):

(†) $\neg\exists x\ \Phi_t[x, \omega_0, t = t]$.

(‡) $\neg\exists x\ \Phi_t[x, V, t = t]$.

The sentence (†) states that the class of all non-empty sets of finite ordinal numbers is uncountable and (‡) states that the class of all non-empty sets is uncountable.

We will present the proof of the first of these sentences expanding the original sketchy proof given by Suszko. The proof makes use of two operations on classes defined as follows (in Suszko's notation $\langle x, y\rangle$ is the ordered pair with x as the first and y as the second element):

1. $b > c = \{a : \ \mathrm{El}(a) \wedge \langle a, c\rangle \in b\}$
2. $\varphi(b, c) = \{a : \ \mathrm{El}(a) \wedge a \in b \wedge a \notin c > a\}$.

Let us notice that if b is a relation, then for any c in its domain $b > c$ is the class of all sets a for which $\langle a, c\rangle \in b$, meaning that it is the b-image of c. Further, if c is a relation, then $\varphi(b, c)$ is the class of these sets a which belong to the class b and do not belong to the c-image of a.

Suszko's proof is by contradiction. Let us suppose that the sentence $\exists x\ \Phi_t[x, \omega_0, t = t]$ is a theorem in M. Let us consider a relation a such that $\Phi_t[a, \omega_0, t = t]$.

We will show that the two conditions being parts of the definition of the relation a imply a contradiction with the initial supposition; these conditions are:

A. $\mathrm{dom}(a) \subset \omega_0$

D. $\forall z\ (z \subset \omega_0 \wedge z \neq \emptyset) \rightarrow \exists w\ (w \in \mathrm{dom}(a) \wedge a > w = z)$,

where the existential quantifier is relativized to sets. Remember that a satisfies the conditions in the definition of the sentence $\Phi_t[a, \omega_0, t = t]$ which means that a is a one-one relation with domain included in ω_0 and counterdomain equal to the set of all sets of finite ordinal numbers. Due to that and the condition D the set $\mathrm{dom}(a)$ contains at least two distinct elements, say x_1 and x_2.

Let us now consider three sets: $\{x_1\}$, $\{x_2\}$ and $\{x_1, x_2\}$. Because a is a one-one relation and its range includes the set of all non-empty subsets of the set ω_0, there exist three different elements $y_1, y_2, y_3 \in \omega_0$ such that:

1. $\langle \{x_1\}, y_1 \rangle \in a$
2. $\langle \{x_2\}, y_2 \rangle \in a$
3. $\langle \{x_1, x_2\}, y_3 \rangle \in a$.

The following cases exclude each other:

1. $\{x_1, x_2\} \cap \{y_1, y_2, y_3\} = \emptyset$
2. $\{x_1, x_2\} \cap \{y_1, y_2, y_3\} \neq \emptyset$

Hence there exists at least one element $t \in \mathrm{dom}(a)$ such that $t \notin a > t$ which means that t does not belong to the a-image of t. In the first of the above two cases we take for t any element of the set $\{y_1, y_2, y_3\}$. In the second case we take for t any element belonging to $\{y_1, y_2, y_3\} - \{x_1, x_2\}$. This is necessary for the non-emptiness of the diagonal set to be defined in a moment.

Due to the condition A we have $t \in \omega_0$. We construct the diagonal set:

$$\varphi(\omega_0, a) = \{x :\ \mathrm{El}(x) \wedge x \in \omega_0 \wedge x \notin a > x\}.$$

This set is non-empty and due to its definition it is a subset of ω_0. One can apply the condition D to it: there exists an element w such that $w \in \mathrm{dom}(a)$ and

$$(*) \qquad \varphi(\omega_0, a) = a > w = \{z :\ \mathrm{El}(z) \wedge \langle z, w \rangle \in a\}.$$

But due to the definition of the diagonal set we have for each z:

$$(**) \qquad z \in \varphi(\omega_0, a) \text{ if and only if } z \notin a > z.$$

We have already shown that the diagonal set is non-empty. Let us ask whether $w \in \varphi(\omega_0, a)$:

1. Due to $(*)$, $w \in \varphi(\omega_0, a)$ if and only if $\langle w, w\rangle \in a$.
2. Due to $(**)$, $w \in \varphi(\omega_0, a)$, if and only if $\langle w, w\rangle \notin a$.

We thus get a contradiction and hence the supposition of the existence of a relation a with the properties given above should be rejected. Which means in turn that $\neg\exists x\ \Phi_t[x, \omega_0, t = t]$ is a theorem in the system.

In general, the sentence $\exists x\ \Phi_y[x, b, \Psi(y)]$ states the existence of a one-one correspondence x between a subset of the set ω_0 and the class of all subsets y of the set b which satisfy the condition $\Psi(y)$. The negation of this sentence states thus the uncountability of the class of all subsets y of the set b which satisfy the condition $\Psi(y)$. It is evident now that Suszko's theorem is a generalization of Cantor's theorem: for $\Psi(y)$ we can take any property expressible in the language of the theory in question; in particular the property of being a constructive set.

The next step is a construction of a metasystem μM along the lines proposed by Tarski. This is obtained by adding to M its "morphology" (structural descriptions of expressions), together with appropriate axioms and rules of inference.

Within μM Suszko defines the concept of a *constructible* object (constructible set) of M. Auxiliary terms used in this definition are those of *categorematic* names (abbreviated: k-names; corresponding to closed terms, without any description operator) and of *designation* relation. An object is constructible in M if it is designated by a k-name. An axiomatic system is called *canonic* if all elements of its universe of discourse are constructible in this sense (that is, designated in a unique way by k-names).

There are of course only denumerably many k-names and hence there may also be only denumerably many constructible objects in the universe of discourse of any axiomatic system like M considered in this approach. However, it cannot be proven *within* M that there exist only denumerably many constructible objects.

Suszko now builds a system M^* obtained from M by adding a new unary predicate (corresponding to the property of being a k-set, meaning constructible object) together with the appropriate axioms (saying which operations on constructible sets again give constructible sets) and a suitable rule of complete induction. It is shown that M^* is a formal counterpart of the metasystem μM. The sentence saying that all objects are constructible is called the *axiom of canonicity*. The system M^* with the addition of this axiom is called $\overline{M}^*$. Obviously, $\overline{M}^*$ is a canonic system. Suszko proves several metatheorems concerning all these systems. In particular, it is shown that if M^* is consistent, then also $\overline{M}^*$ is consistent.

Since (‡) is a theorem of M, it is also a theorem of M^*. This fact, together with the axiom of canonicity provides an interesting reformulation of the Skolem paradox.

Suszko never did return to the study of his canonic axiomatic systems. Several authors have discussed his dissertation (including Fraenkel and Bernays 1958, 23; Fraenkel and Bar-Hillel 1958, 116; Fraenkel, Bar-Hillel and Levy 1973, 116; Mostowski 1955, 38–39; Wang 1955, 64–65). Independently, certain similar ideas were suggested by John Myhill (cf. Myhill 1952).

7.3.5 Critique of restriction axioms

Restriction axioms in set theory were criticized from the very beginning. They were rejected by von Neumann as early as in 1925 and – though on different grounds – by Zermelo in 1930.

We recall that the main construction in Zermelo 1930 is that of a hierarchy of *normal domains*. Those normal domains whose ordinal characteristic is a strongly inaccessible cardinal satisfy all the axioms of the Zermelo system. Zermelo insisted that we should assume the existence of an unbounded sequence of strongly inaccessible cardinals. Mainly due to that assumption Zermelo's views on set theory may be regarded as opposing any axiom of restriction. Zermelo also claimed that set theory should be concerned with the whole transfinite hierarchy of normal domains and not restricted to the investigation of any particular model of the axioms.

In the modern approach, those V_κ of the cumulative hierarchy where κ is a strongly inaccessible cardinal are natural models of ZFC set theory. We know today that the existence of strongly inaccessible cardinals cannot be proven in ZFC. Moreover, it cannot be proven that if ZFC is consistent then ZFC plus the statement "there exist strongly inaccessible cardinals" is consistent, which is a consequence of Gödel's second incompleteness theorem.

The most destructive critique of restriction axioms is presented in Fraenkel, Bar-Hillel and Levy 1973. We will recall a few of its main points here. The authors formulate two axioms of restriction. The main idea captured by the first of them is the following.

- THE FIRST AXIOM OF RESTRICTION. *If Q is a property such that each set whose existence follows from the axioms has this property, then every set has the property Q.*

Now, the property Q should be closed with respect to the set-forming operations described in the axioms. Thus for example if x and y have Q, then $\{x, y\}$ has Q, if x has Q, then $\bigcup x$ has Q, and so on. Also the axiom schemas of separation and replacement can be translated into the suitable closure conditions with respect to Q. One can then show the following facts, among others:

- The First Axiom of Restriction is equivalent to the conjunction of the axiom of regularity (the axiom of foundation) and the sentence saying that there are no strongly inaccessible cardinals. Obviously, all the consequences of nonexistence of strongly inaccessible cardinals are also provable.

- If we consider set theory in a second-order language with a suitable version of the First Axiom of Restriction, then we can prove categoricity of such a theory.

- No consequences concerning the continuum hypothesis can be drawn from the First Axiom of Restriction.

The Second Axiom of Restriction is the conjunction of the following sentences:

1. All sets are constructible (in Gödel's sense).

2. There are no transitive sets which are models of ZF.

It follows from 1) that all sets are well founded. As is known from Gödel's work, 1) implies the generalized continuum hypothesis. The sentence 2), in turn, implies that there are no strongly inaccessible cardinals. Thus, the Second Axiom of Restriction implies the first one. The two Axioms of Restriction share some common features:

- Each of them states that some "big" ordinals or sets with high rank do not exist:

  1. The First Axiom of Restriction implies the non-existence of inaccessible ordinal numbers;
  2. The Second Axiom of Restriction implies the non-existence of transitive sets which are models of ZF.

- Certain complicated sets do not exist:

  1. The First Axiom of Restriction implies the non-existence of non-well-founded sets.
  2. The Second Axiom of Restriction implies the non-existence of non-constructible sets.

The arguments against axioms of restriction presented in Fraenkel, Bar-Hillel and Levy 1973 may be summarized as follows.

- *Analogy.* Observe that the following argument is mostly pragmatic in character:

  > In the case of the axiom of induction in arithmetic and the axiom of completeness in geometry, we adopt these axioms not because they make the axiom systems categorical or because of some metamathematical properties of these axioms, but because, once these axioms are added, we obtain axiomatic systems which perfectly fit our intuitive ideas about arithmetic and geometry. In analogy, we shall have to judge the axioms of restriction in set theory on the basis of how the set theory obtained after adding these axioms fits our intuitive ideas about sets. (Fraenkel, Bar-Hillel and Levy 1973, 117)

- *Faith.* One could restrict the notion of a set to the narrowest possible only if one could have absolute faith in the axioms of ZF, which does not seem to be the case.

- *Faith, again.* Even if one had such faith, it is more likely that one would look for *maximality* axioms (as in geometry) rather than for restriction axioms.

Two more arguments against axioms of restriction are correlated with the author's attitude to the axiom of constructibility:

> *Mathematical elegance.* Axioms of restriction do not improve the mathematical elegance of set theory, in the sense that one can prove more powerful theorems based on them. Rather, they may be involved only in proofs that some sets do not exist. (Fraenkel, Bar-Hillel and Levy 1973, 108–109)

> *Platonistic point of view.* Axioms of restriction are unnatural also when we consider the universe of all sets as an entity capable of growing, in the sense that we can always produce more and more new sets. If an axiom of restriction forces us to accept that the universe of all sets is a fixed entity, then why could we not consider it as a new set in a still bigger universe? In other words, there is no property expressible in the language of set theory which distinguishes the universe from some "temporary universes". These ideas are embodied in the *principles of reflection*, which are, mostly, strong axioms of strong infinity. (Fraenkel, Bar-Hillel and Levy 1973, 118)

7.4 Large cardinal axioms

As we have seen above, minimal axioms in set theory have been generally rejected. This does not mean that we do not consider extremal axioms in set theory anymore. Quite the contrary: maximal axioms, in the form of the axioms of existence of very large cardinal numbers are just one of the central topics in contemporary set theory.

The beginnings of Cantor's set theory saw the introduction of the transfinite scale of *alephs* $\aleph_\alpha$ (where α is an ordinal number). When cardinal arithmetic was developed, then mathematicians discovered (or created) certain large cardinal numbers with special properties, including strongly inaccessible cardinals, Mahlo cardinals, and measurable cardinals. Hausdorff expressed the opinion that strongly inaccessible cardinals would be of no use in mathematics. Zermelo, in his second axiomatization of set theory, postulated the existence of a transfinite hierarchy of strongly inaccessible cardinals. The investigation of very large cardinal numbers began following the discovery that their existence is related to the consistency strength of theories.

Many very large cardinals, defined with respect to different mathematical properties, have recently been under investigation. Let us consider only a few examples:

1. *Strongly inaccessible cardinals.* An uncountable cardinal κ is strongly inaccessible, if κ is not a sum of fewer than κ cardinals which are less than κ and for any $\alpha < \kappa$, $2^\kappa < \kappa$.

   An uncountable cardinal is weakly inaccessible if it is a regular weak limit cardinal. Let us recall that a cardinal κ is regular, if it is equal to its own cofinality (which means that κ cannot be a union of fewer than κ sets of cardinality less than κ). A cardinal number κ is a weak limit cardinal if κ is neither a successor cardinal nor zero. The second condition means that κ cannot be obtained from smaller cardinal by the operation of successor, that is if $\alpha < \kappa$, then also $\alpha^+ < \kappa$.

2. *Mahlo cardinals.* A cardinal κ is weakly Mahlo if it is weakly inaccessible and the set of all weakly inaccessible cardinals less than κ is stationary in κ. A cardinal κ is strongly Mahlo if it is strongly inaccessible and the set of all strongly inaccessible cardinals less than κ is stationary in κ.

   Let us recall that if S is a subset of an uncountable cardinal κ, then S is called stationary if it has a non-empty intersection with any club set (a closed unbounded set in the order topology) in κ.

3. *Measurable cardinals.* An uncountable cardinal κ is measurable if there exists a non-trivial $0-1$ valued κ-additive measure on κ. Let us recall that a measure μ is κ-additive, if the measure of the union of any family of $\lambda < \kappa$ pairwise disjoint sets of cardinality λ equals the sum of all the sets of this family. This is a generalization of the property of σ-additivity of measure: a measure μ is σ-additive (or countably additive) if for any countable sequence of pairwise disjoint μ-measurable sets the measure of their union is the sum of the measures of these sets.

   Measurable cardinals can also be equivalently defined as uncountable cardinal numbers κ such that there exists a non-principal κ-complete ultrafilter on κ. Here κ-completeness of the ultrafilter means that the intersection of any family of less than κ many its elements belongs to the ultrafilter.

It was Kurt Gödel who had already emphasized the necessity of looking for new axioms in set theory, possibly in the form of some maximality axioms. These axioms should be *natural*, which implies that they should be intuitively obvious. What genuinely mathematical intuition is (say, with respect to set theory only) is a different story, far outside the scope of this chapter. Anyway, the nature of mathematical intuition is dynamic. It is regulated, among others, by the appearance of paradoxes and methods of resolving them. It should not be confused with *familiarity*, however. We write more on mathematical intuition in the next chapter.

There are several criteria to be met when formulating new axioms (of the existence of large cardinal numbers), among others: *necessity* (or *non-arbitrariness*) and *fruitfulness in their consequences*. Adding the axiom of infinity to the other axioms of ZF enables us to prove theorems about infinite sets. Likewise, adding an axiom stating the existence of strongly inaccessible cardinals makes it possible to extend operations of set formation beyond what is provable in ZFC. Axioms of the existence of large cardinals (called large cardinal axioms in brief) are of decisive importance for descriptive set theory and in this sense they appear fruitful. The same concerns their applications in, say, infinitary combinatorics.

Joan Bagaria recalls fundamental principles by which (according to Hao Wang's quotations of Gödel's ideas in Wang 1974, 1996) new axioms of set theory should be introduced:

> According to Gödel there are five such principles: *Intuitive Range*, the *Closure Principle*, the *Reflection Principle*, *Extensionalization*, and *Uniformity*. The first, *Intuitive Range*, is the principle of intuitive set formation, which is embodied into the ZFC axioms. The *Closure Principle* can be subsumed into the principle of *Reflection*, which may be summarized as follows: The universe V of all sets cannot be uniquely characterized, i.e., distinguished from all its initial segments, by any property expressible in any reasonable logic involving the membership relation. A weak form of this principle is the ZFC-provable reflection theorem of Montague and Levy (see Kanamori 1994):
>
> *Any sentence in the first-order language of Set Theory that holds in V holds also in some V_α.*
>
> Gödel's *Reflection* principle consists precisely of the extension of this theorem to higher-order logics, infinitary logics, etc.
>
> The principle of *Extensionalization* asserts that V satisfies an extensional form of the Axiom of Replacement and it is introduced in order to justify the existence of inaccessible cardinals. [...]
>
> The principle of *Uniformity* asserts that the universe V is uniform, in the sense that its structure is similar everywhere. In Gödel's words (Wang 1996, 8.7.5): *The same or analogous states of affairs reappear again and again (perhaps in more complicated*

> *versions).* He also says that this principle may also be called *the principle of proportionality of the universe*, according to which, analogues of the properties of small cardinals lead to large cardinals. Gödel claims that this principle makes plausible the introduction of measurable or strongly compact cardinals, insofar as those large-cardinal notions are obtained by generalizing to uncountable cardinals some properties of ω. (Bagaria 2005, 47–48)

There is extensive literature concerning the necessity of accepting new axioms in set theory. Some recent developments are described in detail in monographs (see for example Kanamori 1994), as well as in the numerous papers. We attempt below to summarize a few important trends in this area of study, following the expositions contained in Bagaria 2005 and Koellner 2010.

According to Andrzej Mostowski (Mostowski 1967) there are two principles governing the introduction of new axioms of infinity:

1. *The principle of passing from potential to actual infinity.* We build new sets using the axioms of infinity and replacement of ZF. The universe of all sets is potentially infinite and closed with respect to some operations. We postulate the existence of a *set* which itself is closed with respect to these operations. In this way we obtain for instance *strongly inaccessible* cardinals.

2. *The principle of existence of peculiar sets.* Suppose that while constructing sets according to the known operations on them we always meet sets with a certain property P. If there are no evident reasons which should force us to assume that *all* sets have P, then we propose a new axiom saying that there exist sets *without* the property P. In this way we obtain for example *measurable* cardinals.

In the last few decades several kinds of large cardinals have been investigated. Postulating the existence of a large cardinal (whose existence cannot be proved from the axioms of ZF) is, of course, a kind of maximality condition. Large cardinal axioms are also closely related to the deductive strength of the theories obtained by adjoining such axioms. Let us look at some very elementary examples.

Let Z_0 denote ZFC without the axioms of infinity and replacement. The standard model for this theory is V_ω. The existence of this set follows from the axiom of infinity. Let Z_1 denote Z_0 with the axiom of infinity. Then we can prove in Z_1:

- Z_0 is consistent.
- There exists a standard model for Z_0.

The standard model for Z_1 is $V_{\omega+\omega}$. The existence of this set follows from the axiom of replacement. Let Z_2 denote Z_1 with the axiom of replacement. Then we can prove in Z_2:

- Z_1 is consistent.
- There exists a standard model for Z_1.

The standard model for Z_2 is V_κ, where κ is an inaccessible cardinal. Thus, the next axiom of infinity in this hierarchy will be the sentence "There exists an inaccessible cardinal." The next theory, that is ZFC together with this sentence, proves the existence of a level of the cumulative hierarchy which is a model for ZFC. And so on: in this way we obtain stronger and stronger set theories.

Let $Con(PA_n)$ be the sentence expressing consistency of the n-th order arithmetic PA_n. Then $Con(PA_n)$ cannot be decided in PA_n, but it can be decided in PA_{n+1}. These sentences are connected with the levels of the cumulative hierarchy of sets. Recall that PA_1, that is the first-order system of Peano arithmetic PA is mutually interpretable with ZF minus the axiom of infinity, because V_ω, the first infinite level of the cumulative hierarchy, consists of hereditarily finite sets which can be coded by natural numbers.

However, the sentences $Con(PA_n)$ are not the only undecidable statements:

> The trouble is that when one climbs the hierarchy of sets in this fashion the greater expressive resources that become available lead to more intractable instances of undecidable sentences and this is true already of the second and third infinite levels. For example, at the second infinite level one can formulate the statement PM (that all projective sets are Lebesgue measurable) and at the third infinite level one can formulate CH (Cantor's continuum hypothesis). [...] [I omit the fragment where Koellner briefly summarizes Gödel's and Cohen's results showing together the independence of CH from ZFC – JP]
>
> These instances of independence are more intractable in that no simple iteration of the hierarchy of types leads to their resolution. They led to a more profound search for new axioms. (Koellner 2010, 3–4)

Due to Gödel's and Cohen's results concerning the independence of CH from ZFC one can see that ZFC is mutually *interpretable* with ZFC+CH, as well as with ZFC+¬CH. The situation with PM is, however, different. The method of inner models shows that ¬PM holds in the constructible universe L. Hence ZFC and ZFC+¬PM are mutually interpretable. But Shelah has shown that ZFC+PM implies the consistency of ZFC and therefore, due to Gödel's second incompleteness theorem, ZFC+PM *is not* interpretable in ZFC. It follows that in order to establish the independence of PM from ZFC we need to assume the consistency of some *stronger* theory – namely that of ZFC plus the sentence "There exists a strongly inaccessible cardinal".

This is only a very elementary example. There is a profusion of axioms of existence of large cardinals which have relevant impact on the independence proofs. Let us only add that there exists a pattern of formulating large cardinal axioms in terms of elementary embeddings. Generally speaking, one considers non-trivial (meaning different from identity) elementary embeddings $j : V \to M$ of the cumulative hierarchy V into a transitive class M. The least ordinal moved by such an embedding is called the *critical point* of j and denoted by $\mathrm{crit}(j)$. For example, a cardinal is *measurable* if and only if it is the critical point of some such embedding. Further conditions imposed on j and M enable us to create several sorts of large cardinal axioms. As Kunen has shown, there is no elementary embedding $j : V \to V$, so there exists an upper bound for this procedure.

The structure of degrees of interpretability of theories is very complicated. However, *natural* theories having practical mathematical applications happen to be orderly comparable, which of course is only an *empirical* fact. Theories can be compared through large cardinal axioms corresponding to them:

> Given ZFC+φ and ZFC+ψ one finds large cardinal axioms Φ and Ψ such that (using the methods of inner and outer models) ZFC+φ and ZFC+Φ are mutually interpretable and ZFC+ψ and ZFC+Ψ are mutually interpretable. One then compares ZFC+φ and ZFC+ψ (in terms of interpretability) by mediating through the natural interpretability relationship between ZFC+Φ and ZFC+Ψ. So large cardinal axioms (in conjunction with the dual method of inner and outer models) lie at the heart of the remarkable empirical fact that natural theories from completely distinct domains can be compared in terms of interpretability. (Koellner 2010, 10–11)

Sometimes the procedure outlined above is the only known way to compare theories, which provides a *pragmatic* justification for the investigations of large cardinal axioms.

7.5 Sentences independent from the axioms

Georg Cantor dedicated a lot of work to prove the continuum hypothesis, without success. It took several decades to recognize that the hypothesis could neither be disproved (Gödel's result) nor proved from the axioms of ZF (Cohen's result). The method of forcing introduced by Cohen and developed later by many mathematicians made it possible to present a lot of other sentences independent from the axioms of ZF. Some important examples are as follows:

1. *The continuum hypothesis.* $2^{\aleph_0} = \aleph_1$.
2. *The generalized continuum hypothesis.* $2^{\aleph_\alpha} = \aleph_{\alpha+1}$ for any ordinal number α.
3. *The axiom of constructibility.* Every set is constructible.
4. *Suslin problem.* Let R be a linearly ordered set without endpoints and let the considered ordering be dense and complete. Let every family of mutually disjoint non-empty open intervals in R be countable. Any set ordered by such a relation R and *not* order isomorphic to the real numbers is called a Suslin line. The Suslin hypothesis states that Suslin lines do not exist.

There are numerous statements not only in set theory itself but also in many other areas of mathematics (number theory, algebra, order theory, measure theory, topology, or functional analysis, and so on) which are independent from ZFC. It is debatable whether one could invent (discover?) some "miraculous" additional axioms which could remove those independencies. It may be the case that independence phenomena are, in a sense, an inherent property of mathematics.

Part III

Cognitive aspects

Chapter 8

Mathematical intuition

In the first part of this chapter we give a brief mention to some selected philosophical standpoints concerning mathematical intuition. The second part contains selected opinions of professional mathematicians concerning mathematical intuition. The last part is devoted to the role of mathematical intuition in learning and teaching mathematics.

8.1 Philosophical remarks

So far, we have discussed extremal axioms from a historical perspective and we have focused our attention on their logical and mathematical aspects. Extremal axioms are related to the intended models of theories and the latter notion is in turn related to mathematical intuitions. The terms *intended model* and *standard model* are sometimes used interchangeably in literature. However, we would like to propose the following distinction (already mentioned in chapter 1):

1. By *intended model* we mean the structure which has been investigated for its own sake, typically for a long time, so that we have collected a large amount of data and have proven many fundamental theorems about it. One could say that intended models are such structures which became *domesticated*, easily accessible cognitively, and which are responsible for the basic mathematical intuitions. One can see that this is only an intuitive characterization of the term *intended model* which, in turn, is also an intuitive term. Natural number series (with arithmetical operations) may serve as an example of intended model understood in this way. Further examples include the geometrical universe described by Euclid, and

could also include the universe of Cantorian set theory before it was axiomatized by Zermelo.

2. One can only talk about a *standard model* after having obtained a developed formal theory, ultimately an axiomatic one. In such a case one can establish the class of all possible models of the theory in question. Then the standard model of such a theory is its model most closely related to the intended one. It may happen that the standard model of a theory obtains a precise characterization, for example in terms of isomorphism or elementary equivalence; for instance, Tennenbaum's theorem says that the standard model of first-order Peano arithmetic is the only (up to isomorphism) recursive model of that theory. Examples of standard models in the proposed sense are, for instance, the standard model of first-order Peano arithmetic and the completely ordered real field $\mathbb{R}$.

A theory may obviously also have *non-standard* models which do not resemble the intended model at all or resemble it only to a certain degree. For instance, the first-order Peano arithmetic has a lot of non-standard models; all its countable non-standard models have the same order type $\omega + (\omega^* + \omega) \cdot \eta$ but they differ with respect to the arithmetical structure.

In general, standard models can be determined neither syntactically nor semantically. It is our epistemic choice to call a given model the standard one. That choice is of course influenced by the research practice of professional mathematicians, by the accumulation of results characterizing the intended model.

The intended models emerge as the result of mathematicians' research which has to be considered in both contexts – that of *discovery* and that of *justification*:

1. *Context of discovery.* The process of mathematical discovery remains mysterious to a certain degree. Mathematical discoveries and inventions rely on the (mathematical) *intuitions* of professional mathematicians. These include beliefs (sometimes not verbalized explicitly) concerning mathematical facts.

2. *Context of justification.* This primarily embraces methods of establishing mathematical truths. The most important procedure in this respect is that of *deduction*, meaning the presentation of proofs of theorems. A proof is often said to be a confirmation of the previous intuition.

The concept of *intuition* is widely discussed in philosophical literature. *The Encyclopedia of Philosophy* (Edwards 1967) has the following entry concerning intuition:

> ***Intuition***. The broadest definition of the term "intuition" is "immediate apprehension". "Apprehension" is used to cover such disparate states as sensation, knowledge, and mystical rapport. "Immediate" has as many senses as there are kinds of mediation: It may be used to signify the absence of inference, the absence of causes, the absence of the ability to define a term, the absence of justification, the absence of symbols, or the absence of thought. Given this range of uses, nothing can be said about intuition in general. Instead, it is necessary to pick out those principal meanings of the term which have played the most important roles in philosophical controversy and to discuss each of these individually. (Edwards 1967, volume 4, 204, article by Richard Rorty, 204–212)

This entry goes on to discuss the "principal meanings" mentioned above:

1. Intuition as unjustified true belief not preceded by inference.
2. Intuition as immediate knowledge of the truth of a proposition, where "immediate" means "not preceded by inference."
3. Intuition as immediate knowledge of a concept.
4. Intuition as nonpropositional knowledge of an entity – knowledge that may be a necessary condition for, but is not identical with, intuitive knowledge of the truth of propositions about the entity.

The article on intuition in *The Stanford Encyclopedia of Philosophy* written by Joel Pust enumerates several possible ways of understanding the fact that a subject has certain intuitions (Pust 2012):

1. S has the intuition that p if and only if S believes that p.
2. S has the intuition that p if and only if S forms the occurrent belief that p without consciously deriving it from other beliefs.
3. S has the intuition that p if and only if S forms the occurrent belief solely on the basis of the competence with the concepts involved in p.
4. S has the intuition that p if and only if S is disposed to believe p.

5. S has the intuition that p if and only if S is disposed to believe p merely on the basis of understanding p.

6. S has the intuition that p if and only if it seems to S that p.

7. S has the intuition that p if and only if it intellectually seems to S that p.

8. S has the *rational intuition* that p if and only if it intellectually seems to S that necessarily p.

9. S has the *rational intuition* that p if and only if either [A] (1) it intellectually seems to S that p and (2) if S were to consider whether p is necessarily true, it would intellectually seem to S that necessarily p, or [B] it intellectually seems to S that necessarily p.

The commonly accepted view is that knowledge is characterized as justified true belief. Let us list the possible combinations of the properties of being true and being justified as applied to beliefs:

Justified	True	Example
yes	yes	knowledge
yes	no	illusion
no	yes	intuition
no	no	prejudice

This tentative classification of beliefs can possibly be further elaborated by taking into account, for instance, types of justification in question.

Here we are interested only in *mathematical intuitions*. It is necessary to make certain distinctions as far as the understanding of this term is concerned. We propose stratifying the mathematical intuitions of cognitive subject into the following levels or stages:

1. *Proto-intuitions*. Such beliefs are related to our cognitive abilities. They may include the likes of subitation, distinctions between inside and outside, determination of locations and directions, recognition of simple shapes, and recognition of rhythm, and so on.

2. *Intuitions based on symbolic violence in education*. These concern ordinal and cardinal aspects of numbers, simple arithmetical properties, abilities to operate on symbols, and achieving fluency in applications of prescribed algorithms, and so forth.

3. *Intuitions of professional mathematicians.* These are the most difficult to describe, because they depend to a great extent on individual talent. We have some reports from introspection referring to the circumstances of particular mathematical discoveries but it is difficult (if possible at all) to measure the reported phenomena. We can also try to reconstruct the intuitions in question from the source texts, although that would require much caution: we should not impose our own way of thinking on that of mathematicians from earlier epochs.

Discussion of the nature of mathematical intuition is of course connected to other important problems investigated in the philosophy of mathematics. These include, among others, the following questions:

1. *What are mathematical objects?*
2. *What access do we have to them?*
3. *What is mathematical truth?*
4. *What is a correct method in mathematics?*
5. *Why is mathematics successful in science?*
6. *Is mathematics invented or discovered?*

It should be stressed that the arguments used in the defence of a chosen standpoint are what is most important and not the ultimate answers to philosophical questions concerning mathematics. We have recently witnessed a large diversification of positions in the philosophy of mathematics. They grew from the classical standpoints (logicism, formalism, intuitionism) or emerged as new approaches to vital problems in the philosophy of mathematics. We are not going to present this diversity of opinions in detail; instead we shall only name the existing standpoints (together with the names of their most prominent defenders) with a short commentary concerning the main assumptions of the position in question.

Important pre-20th century claims about mathematical intuition can be found, for instance, in the works of Rene Descartes and Immanuel Kant.

Rule III in Descartes' *Rules for the Direction of the Mind* concerns obtaining knowledge: *Concerning objects proposed for study, we ought* *to investigate what we can clearly and evidently intuit or deduce with certainty, and not what other people have thought or what we ourselves conjecture. For knowledge can be attained in no other way.* Then there follows a comment explaining what the author means by *intuition* (citing the English translation of Descartes' treatise credited to Elizabeth Anscombe and Peter Thomas Geach in 1954):

> By intuition I mean, not the wavering assurance of the senses, or the deceitful judgment of a misconstructed imagination, but a conception, formed by unclouded mental attention, so easy and distinct as to leave no room for doubt in regard to the thing we are understanding. It comes to the same thing if we say: It is an indubitable conception formed by an unclouded mental mind; one that originates solely from the light of reason, and is more certain even than deduction, because it is simpler (though, as we have previously noted, deduction, too, cannot go wrong if it is a human being that performs it). Thus, anybody can see by mental intuition that he himself exists, that he thinks, that a triangle is bounded by just three lines, and a globe by a single surface, and so on; there are far more of such truths than most people observe, because they disdain to turn their mind to such easy topics. (Descartes 1954)

For Kant, the judgments of mathematics are *synthetic a priori*. Philip Kitcher observes that Kant's point of view has to be considered in two aspects:

> Kant usually offers this as one thesis, but it is fruitful to regard it as consisting of two separate claims, a metaphysical subthesis and an epistemological subthesis.
>
> (KM) The truths of pure mathematics are necessary, although they do not owe their truth to the nature of our concepts.
>
> (KE) The truths of pure mathematics can be known independently of particular bits of experience, although one cannot come to know them through conceptual analysis alone. (Kitcher 1975, 23)

Kant's views have tremendous influence on subsequent investigations in the philosophy of mathematics. Let us summarize the most important features of Kant's proposals regarding mathematics:

1. Kant's analytical judgments correspond to Leibniz's truths of reason.

2. Kant's synthetical judgments correspond to Leibniz's truths of facts.

3. According to Kant, all statements of pure mathematics are synthetical *a priori* statements. He writes in §6 of *Prolegomena to Any Future Metaphysics That Will Be Able to Present Itself as a Science*:

   > All mathematical cognition has this peculiarity: it must first exhibit its concept in a visual form [*Anschauung*] and indeed a priori, therefore in a visual form which is not empirical, but pure. Without this mathematics cannot take a single step; hence its judgments are always visual, viz., "intuitive"; whereas philosophy must be satisfied with discursive judgments from mere concepts, and though it may illustrate its doctrines through a visual figure, can never derive then from it.

4. Space and time are forms of pure intuition. Space forms the foundations of geometry and the pure intuition of time is responsible for the foundations of arithmetic and the concept of number.

5. Kant distinguished between the construction of an object and the postulate of its existence.

6. Kant's critical epistemology concerned conceptions and phenomena; it did not penetrate into reality and "things in themselves".

7. According to Kant, actual infinity is an idea of reason, which is to say a concept that is in itself consistent, though its realizations are neither perceived nor explicitly constructed.

We are not going to discuss the classical standpoints in the philosophy of mathematics in detail but we shall restrict ourselves to outlining only the most important:

1. *Platonism* (Plato, Pythagoras, Cantor, Gödel, Erdős). Mathematical platonism claims that mathematical objects are abstract, have no spatiotemporal properties and are inert with respect to causal dependencies. Thus platonists believe in the existence of a transcendental mathematical realm which is eternal and unchanging. We have access to that world through mathematical intuition – a position explicitly formulated by Kurt Gödel. There are several kinds

of platonism, for example full-blooded platonism, or set-theoretical realism. It is usually claimed that most working mathematicians are adherents of platonism, but one has to remember that as a rule professional mathematicians are not experts in philosophical considerations. Perhaps the platonic standpoint is a convenient way of talking about mathematical objects.

2. *Logicism* (Frege, Russell, Carnap, Wright). Defenders of this standpoint claim that mathematics can be reduced to logic. All mathematical statements are necessary logical truths. The initial version of logicism presented in Frege's system got into trouble (Russell's antinomy) while the later phase of logicism, defended by Russell, became embroiled in some formal complications (axiom of reducibility).

3. *Formalism* (Hilbert, Bernays, Gentzen, Tarski, Carnap, Curry, von Neumann). Strict formalists view mathematics as pure syntax: manipulation with strings of symbols according to the admissible rules. The origins of formalism are ascribed to David Hilbert, who formulated his famous foundational program. This primarily called for the formalization of mathematics followed by the development of metamathematics, a formal methodology of mathematics which could enable us to prove that mathematical theories are consistent and complete. This program may only be realized in part, due to the well known limitative theorems.

4. *Intuitionism* (Kronecker, Brouwer, Weyl, Heyting). Mathematics is the result of human cognitive activity alone, and this mental activity does not resemble discoveries but is based on constructive inventions. The warrant of mathematical truth is proof alone. Proofs should be constructive, and therefore some laws of classical logic (for example excluded middle) are not accepted.

Modern standpoints in the philosophy of mathematics include:

1. *Constructivism* (Shanin, Markov, Martin-Löf, Lorenzen, Bishop). In order to prove the existence of an object, one should present a method of its construction. Constructivists therefore do not accept proofs by contradiction. An extreme form of constructivism is finitism, according to which a mathematical object exists if and only if it can be constructed from natural numbers in a finitary way.

2. *Structuralism* (Resnik, Shapiro, Benacerraf). Mathematics is about structures: mathematical theories describe abstract structures, and mathematical objects are considered to be places in those structures. There are several versions of structuralism: *ante rem*, *in re*, *post rem.*

3. *Empiricism* (Quine, Putnam, Lakatos, Chaitin). Mathematical truths are not *a priori* statements. We discover mathematical facts in a similar way to that followed in empirical sciences. What is important for empiricism (which is a kind of realism) is the indispensability argument: because mathematics is indispensable for empirical sciences, then – if we want to believe that the phenomena described by these sciences are real – we ought to believe in the reality of the mathematical objects to which we refer.

4. *Fictionalism* (Field). It is claimed that mathematics is a collection of (useful) fictions. Field tried for instance to show that one can develop parts of physics without the explicit use of mathematics.

5. *Social constructivism* (Vygotsky, Ernest, Davis, Hersh, Tymoczko). Mathematics is a social construct and as such it is subject to revisions and corrections. Mathematical objects are semiotic objects.

6. *Embodied cognition* (Lakoff and Núñez, Tall). This stresses that mathematical activities emerge as result of human cognitive apparatus. Defenders of this standpoint claim that mathematics exists only in human brains, and there is no such thing as transcendental mathematics. They also support the view that the major mechanism of concept formation in mathematics is based on forming conceptual metaphors.

7. *Mathematical agnosticism* (Azzouni, Balaguer, Boscolo). Belief in the existence of Platonic realm (causally inert objects outside spacetime) does not, in general, influence the research conducted by professional mathematicians.

We propose thinking of the above standpoints not as contradicting each other, but rather complementary to each other. Of course, the reader may choose (or invent) her own position, meaning the only one she considers ultimately right. Numerous presentations of the philosophy of mathematics are accessible – one of the most recent is Bedürftig and

Murawski 2018 (see also Batóg 1996). There are also many selections of source texts, for instance: Benacerraf and Putnam 1983, Shapiro 2005.

8.2 Research practice

What can be said about mathematical intuitions based on the research practice of professional mathematicians? The style of mathematical publications recently finding acceptance forces the authors to present only the results of their investigations, that is theorems supported by proofs. Occasionally, mathematicians may report on the intuitions which guided their reasoning. We find such "confessions" in memoirs or mathematicians' reflections about their past work. We may also try to reconstruct the intuitions behind the route to a mathematical result in a source text, taking into account for instance the ideas, concepts, results, and proofs from earlier works known to the author in question.

8.2.1 Context of discovery: selected views

Mathematical inventions and discoveries do not emerge as a result of any algorithmic process. They depend on the individual abilities of mathematicians, on the state of mathematical knowledge of a given epoch, on inspirations from natural sciences, on an aesthetic sense, and possibly on some other factors as well. Specific cases of mathematical discoveries are described in textbooks on the history of mathematics, and they are either described as examples of continuation in the development of a given discipline or as revolutionary moves which are responsible for a radical change of perspective or the creation of a completely new discipline.

The classical work Lakatos 1976 provides a detailed description of a process of coming up with a mathematical result – in that case Euler's formula about the number of vertices, edges and faces of a polyhedron. The small booklet Hadamard 1954 (first published by Princeton University Press in 1945) is still repeatedly quoted as a source of ideas about mathematical inventions. Most famous are his reflections on the stages (both conscious and unconscious) in mathematical invention, including the preparation, incubation, illumination, verifying and "precising" stages. He also wrote about the role of language in mathematical thought and analyzed several individual cases of modes of thinking in famous mathematicians. The booklet contains appendices: *An inquiry into the working methods of mathematicians* (translated from the text from *L'Enseignement Mathématique*, vol IV, 1902 and vol. VI, 1904), a letter from Albert Einstein about his style of doing research, and a note on the invention of the infinitesimal calculus.

Henri Poincaré published *Intuition and Logic in mathematics* as part of *La valeur de la science* in 1905. It was translated into English by George B. Halsted and published in 1907 as part of Poincaré's *The Value of Science*. The text of that note is accessible online:
http://www-history.mcs.st-and.ac.uk/Extras/Poincare_Intuition.html

We have chosen the following passages from the said text illustrating the author's views on the diversity of intuitions:

> We believe that in our reasonings we no longer appeal to intuition; the philosophers will tell us this is an illusion. Pure logic could never lead us to anything but tautologies; it could create nothing new; not from it alone can any science issue. In one sense these philosophers are right; to make arithmetic, as to make geometry, or to make any science, something else than pure logic is necessary. To designate this something else we have no word other than intuition. But how many different ideas are hidden under this same word? Compare these four axioms:
>
> 1. Two quantities equal to a third are equal to one another;
> 2. if a theorem is true of the number 1 and if we prove that it is true of $n + 1$ if true for n, then it will be true of all whole numbers;
> 3. if on a straight the point C is between A and B and the point D between A and C, then the point D will be between A and B;
> 4. through a given point there is not more than one parallel to a given straight.
>
> All four are attributed to intuition, and yet the first is the enunciation of one of the rules of formal logic; the second is a real synthetic a priori judgment, it is the foundation of rigorous mathematical induction; the third is an appeal to the imagination; the fourth is a disguised definition.
>
> Intuition is not necessarily founded on the evidence of the senses; the senses would soon become powerless; for example, we can not represent to ourselves a chiliagon, and yet we reason by intuition on polygons in general, which include the chiliagon as a particular case.
>
> [. . .]
>
> We have then many kinds of intuition: first, the appeal to the senses and the imagination; next, generalization by induction, copied, so

> to speak, from the procedures of the experimental sciences; finally, we have the intuition of pure number, whence arose the second of the axioms just enunciated, which is able to create real mathematical reasoning. I have shown above through example that the first two cannot give us certainty; but who would seriously doubt the third, who would doubt arithmetic?

William Thurston wrote an interesting essay concerning, among others, possible ways of understanding a given mathematical concept – in this case the concept of derivative (Thurston 1994). After discussing several such ways, he added:

> This is a list of different ways of *thinking about* or *conceiving of* the derivative, rather than a list of different *logical definitions*. Unless great efforts are made to maintain the tone and flavor of the original human insights, the differences start to evaporate as soon as the mental concepts are translated into precise, formal and explicit definitions.
>
> [...]
>
> These differences are not just a curiosity. Human thinking and understanding do not work on a single track, like a computer with a single central processing unit. Our brains and minds seem to be organized into a variety of separate, powerful facilities. These facilities work together loosely, "talking" to each other at high levels rather than at low levels of organization. (Thurston 1994, 163–164)

The most important in this quotation is the very last sentence. This contradicts the claim that mathematical reasoning is mostly unconscious, inaccessible to direct introspection (as some cognitive scientists claim).

Terence Tao, one of the most prominent contemporary mathematicians, writes on his blog about three stages in mathematical education (Tao 2009). His main ideas may be summarized as follows:

1. *The "pre-rigorous" stage.* Here mathematics is taught in an informal manner, referring to intuitions, concrete examples and "hand waving". Simple calculations are in focus here.

2. *The "rigorous" stage.* Here the emphasis is put more on theory. Students should be aware of what it means that a given procedure is applied properly. Also the ability of a free manipulation of symbols is required here.

3. *The "post-rigorous" stage.* Here one should be able to confront the previous informal intuitions with the meaning of concepts established in formal theories. The emphasis here is put on applications, intuition, and the "big picture".

He adds further comments claiming that: *The point of rigour is* not *to destroy all intuition; instead, it should be used to destroy* bad *intuition while clarifying and elevating* good *intuition.*

Davis and Hersh summarize their characterization of mathematical intuition as follows (Davis and Hersh 1981):

1. Intuitive is a converse of precise.
2. Intuitive means visual.
3. In the absence of proof, intuitive means persuasive or probable.
4. Intuitive means incomplete.
5. Intuitive means based on a physical model.
6. Intuitive means holistic and synthetic (in contraposition to particular and analytic).

Davis and Hersh also comment on the main trends in the philosophy of mathematics, and claim that:

1. All standard points of view in the philosophy of mathematics are essentially based on some understanding of the concept of intuition.
2. None of them has ever tried to explain the meaning of the accepted notion of intuition.
3. The notion of intuition is complex and difficult but not unexplainable or non-analyzable. Realistic analysis of intuition is a reasonable goal of the philosophy of mathematics.

According to the above authors, the main trouble of Platonism in mathematics is the nature of the alleged contact between ideal mathematical objects and the human mind. In turn, a difficulty for the constructive approach is seen in the fact that the alleged common intuition of the system of natural numbers is not in fact common for the human population; it is common to specialists only. Finally, the formalist view has troubles with an adequate explanation of changes in mathematical thinking, with the dynamics of intuition.

Davis and Hersh formulate their own position with respect to mathematical intuition referring to the processes of learning and teaching mathematics. Thus, mathematical intuition is not a direct perception of something existing externally and eternally. Rather, it is an effect emerging in the mind after numerous experiences with concrete objects (including signs and mental representations).

Charles Parsons has introduced an interesting distinction between *intuitions that* and *intuitions of* (Parsons 2008). The first of them concerns mathematical facts, it is propositional in content. The second, in turn, concerns mathematical objects. He writes:

> I propose to use the term 'intuition' so that a mode of evidence does not count as intuition unless it is analogous to perception in a definite way. In the case of some proposed kind of intuition *that*, one way in which the analogy can be made out is that it involves intuition of certain objects. Unlike Gödel, I will not argue that all rational evidence of principles that are not the conclusions of deductive or empirical arguments is a case of intuition, or even that intuition extends very far into the conceptual domain. The following inquiry will be in the tradition of Kant, for whom intuition and Reason are of a different nature, rather than in the tradition of Spinoza and Leibniz, for whom intuitive knowledge is possible at high levels of abstraction and rational integration. (Parsons 2008, 147)

Richard Tieszen analyzes mathematical knowledge in terms of a phenomenological approach:

> Intuition is understood in terms of the notion of *fulfillment* of intensions. Intentions may be thought of as fulfilled, partially fulfilled, frustrated, and so on. We *intuit* an object when our intention to the object is fulfilled, for then the object intended is actually displayed or presented, precisely *as* it is intended, and this is not the case where the object intended is not intuited. The difference between intuiting an object and merely having an intention directed toward the object can be roughly described as the difference between actually seeing something and merely thinking about it or having a concept of it. There is an obvious resemblance here to Kant's distinction between having a concept of an object and having an intuition of the object, even though as we shall see later Husserl believed that he was extending and developing a Kantian view of knowledge for areas of human experience – logic and mathematics

> – that Kant did not fully appreciate. In the case of mathematical objects like natural numbers and finite sets the problem therefore is to give an account of just what is to count as fulfillment of mental intensions to such objects. (Tieszen 1989, 24)

Efraim Fischbein takes into account the cognitive-behavioral function of intuition and writes:

> In our opinion intuition is the analog of perception at the symbolic level. It has the same behavioral task as perception, namely to prepare and to guide our mental or practical activity. Therefore an intuitive conception must possess a number of features analogous to that of perception: globality, structurality, imperativeness, direct evidence, a high level of intrinsic credibility. In this way intuitions are able to inspire and guide our intellectual endeavors firmly and promptly even in a situation of uncertain or incomplete information. (Fischbein 1972, 201)

One of the most eminent defenders of the fundamental role of mathematical intuition in the context of mathematical discovery was Kurt Gödel. According to Gödel, mathematical concepts form an objective reality. We cannot create or change them, but we can describe and perceive them. He was not very specific regarding what should be meant by mathematical intuition, yet some of his remarks on it are nevertheless among those most frequently quoted, for example (citing *Collected Works*, i.e. Gödel 1986–2003):

> Thereby [by a Platonistic view – JP] I mean the view that mathematics describes a non-sensual reality, which exists independently both of the acts and the dispositions of the human mind and is only perceived, and probably perceived very incompletely, by the human mind. (*Some basic theorems on the foundations of mathematics and their implications.* Gödel CW III, 323)

> However, mathematical intuition in addition produces conviction that, if these sentences express observable facts and were obtained by applying mathematics to verified physical laws (or if they express ascertainable mathematical facts), then these facts will be brought out by observation (or computation). (Gödel CW III, 340, version III *Is mathematics syntax of language?*)

> If the possibility of a disproval of mathematical axioms is frequently disregarded, this is due solely to the convincing power of mathematical intuition. (Gödel CW III, 361, version V *Is mathematics syntax of language?*)

> But, despite their remoteness from sense-experience, we do have something like a perception of the objects of set theory, as is seen

> from the fact that the axioms force themselves upon us as being true. I don't see any reason why we should have less confidence in this kind of perception, i.e. in mathematical intuition, than in sense-perception. (*What is Cantor's continuum problem?* In: Benacerraf and Putnam 1964, 271)

8.2.2 Context of justification

A mathematical result is considered as justified if it has a proof accepted by the mathematical community. Ideally, any mathematical proof should be a chain of inferences based on logical entailment, without any gaps and leading from the accepted assumptions to the defended thesis. In practice, mathematical proofs contain gaps and shortcomings; it is understood that they can be filled by the reader.

The style of mathematical publications has recently been very sterile. Their authors present theorems and lemmas with complete or only sketched proofs, while the paths taken to achieve these proofs are usually hidden. However, one may try to reconstruct the main factors responsible for proof creation, for example:

1. *Imagination.* This factor is related to individual talent. Mathematicians have different styles of imagination: geometrical, combinatorial, or algebraic, and so on.

2. *Intuition.* This cognitive ability seems to be the most fundamental in the context of mathematical discovery (see the previous section).

3. *Generalization and abstraction.* It is a very natural tendency in mathematical research to generalize a given result to a wider scope of phenomena.

4. *Analogy.* Reasoning based on analogy supports many new mathematical results, obtained by projecting a known case in one domain into another domain.

5. *Induction.* Inductive hints are often useful in the formulation of conjectures to be proven. If we find a recurring regularity in several cases, we try to formulate a general hypothesis embracing the observed cases and putting them into a general schema.

6. *Abduction.* Abductive reasoning seems to occur in those cases when mathematicians try to find sufficient conditions for truth of a statement which they find appealing.

Several types of proof exist; they may be differentiated with respect to their logical structure, the techniques used, or the sources of inspiration, and so forth:

1. *Direct proof.* The thesis is inferred from the assumptions in a direct way, by applying consecutively the rules of inference.

2. *Proof by contradiction.* The thesis is negated and if this leads to a contradiction, then the thesis itself is considered to be proven.

3. *Proof by cases.* The proof is conducted by inspection of several particular cases.

4. *Proof by induction.* There are several kinds of inductive proof, for example by standard mathematical induction or by transfinite induction.

5. *Elementary proof.* An elementary proof (in a given mathematical discipline) means a proof using only basic techniques typical for that discipline. For instance, the first proof of the prime number theorem was not elementary, as it referred to complex analysis (half a century later an elementary proof of that theorem was found).

6. *Constructive proof.* In constructive proofs, the existence of an object should be witnessed by the construction of it, with the use of admissible methods.

7. *Explanatory proof.* Recently, there has been much discussion of regarding what it means for a proof to *explain* something. If a proof is called explanatory, then it should possess certain virtues, as for instance depth and generality; see Mancosu 2015.

8. *Infinitary proofs.* Standard proofs investigated in mathematical logic are finitary syntactic objects. But it is also possible to consider proofs in which infinitary rules are applied, meaning rules with an infinite set of premisses, as for instance the ω-rule.

9. *Proofs based on empirical considerations.* The idea of a mathematical proof may be suggested by investigation of empirical phenomena. For instance, facts from optics may suggest geometrical theorems, facts related to heat flow may suggest theorems from complex analysis, and so on.

10. *Computer based proofs.* In recent decades we have witnessed the development of computer-assisted proofs, in other words proofs which are partly based on a computer program which can in a relatively short time inspect thousands of cases. It is debatable whether we are still dealing with a mathematical proof in the former meaning of the term (that is as a procedure which can be completely surveyed by a human subject), and as such the concept of mathematical proof might be different in the near future.

The history of mathematics has been witness to several cases of incomplete or even incorrect proofs. Here are a few examples (mathematical and logical) of incomplete proofs and false results, later corrected:

Euclid's Vth postulate. The parallel postulate in Euclid's geometry is more complicated in formulation than the other axioms of the system. Many investigators suspected that it could be a consequence of the other axioms: Proclus, Alhazen, Omar Chajjiam, Nasir ad-Din Tusi, among others. However, their efforts in proving that were unsuccessful. The attitude then changed and the possibility of geometry with negation of the parallel postulate was allowed for (Saccheri, Lambert). Finally, systems of non-Euclidean geometry were shown to be alternative systems of geometry (Gauss, Lobachevsky, Bolyai, Riemann, Poincaré) and the independence of the parallel postulate was proved (Beltrami). It is worth noting that:

1. It took a long time to overturn the belief that Euclid's geometry was the valid geometry, describing adequately physical space.
2. Some intuitive prejudices prevented earlier discovery of the system of non-Euclidean geometry. For instance, Saccheri did not accept the possibility that a straight line could be limited (as is the case in elliptic geometry) or the possibility of there existing a triangle with maximum area (as it is the case in hyperbolic geometry).

Galileo: brachistochrone. Galileo was convinced that brachistochrone (the curve of fastest descent) was a segment of a circle and not any succession of line segments. This claim was wrong – the curve in question is the cycloid. Investigation into problems of this sort gave birth to the calculus of variations.

Origins of calculus, infinitesimals. Newton based his construction on kinematic and geometric intuitions while the approach to calculus by Leibniz was based on geometric and algebraic intuitions. Mathematicians of the 17th and 18th century developed calculus showing its numerous applications, although the precise meaning of its fundamental notions remained unclear. The first mathematically precise representation of infinitely small quantities was given in non-standard analysis from the 20th century.

Infinite series: Grandi series, Euler proofs. In the 17th and 18th century the precise criteria for convergence of infinite series had not yet been formulated and mathematicians sometimes obtained curious results. Let us mention that Euler, solving the Basel Problem used an argument from analogy (writing about infinite products involving trigonometric functions). Only later did Weierstrass prove a factorization theorem justifying Euler's argument.

Solutions by radicals: naive intuitions abolished. In the presence of solutions by radicals for polynomials of the first, second, third and fourth degree one was tempted to believe in the existence of a general method of solution by radicals for polynomials of any degree. Due to the results obtained by Galois and others this false belief was discarded.

Cauchy: continuity and infinitesimals. There is still heated discussion regarding whether Cauchy was right or wrong claiming that a limit of (pointwise) convergent sequence of continuous functions is itself continuous. On the one hand, there is the claim that he was wrong: a stronger assumption of uniform convergence is necessary (within classical analysis). On the other hand, it is important to remember that Cauchy made essential use of the concept of an infinitesimal, and as such he may have understood something different by continuity than the classical concept.

Lebesgue: projections of Borel sets. Henri Lebesgue claimed that projections of Borel sets were Borel, which was a mistake. The mistake was corrected by Luzin. This gave rise to a whole domain of descriptive set theory dealing with projective sets.

Malfatti circles. Gian Francesco Malfatti formulated the problem of carving three circular columns out of a triangular block of marble, using as much of the marble as possible. The problem can of course be reduced to a planimetry problem (concerning the bases of the pillars). Malfatti conjectured that three mutually-tangent circles inscribed within the triangle would provide the optimal solution. Such circles are called *Malfatti circles*. His conjecture was wrong which was shown no sooner than in 1930.

Lion and man problem. This problem was posed by Radó in 1925. It concerns a game of pursuit. A man and a lion have identical maximum speed. The lion tries to catch the man. They are situated on an island in the shape of a circle. Who has a winning strategy? For a few decades it was accepted that the lion has a winning strategy in this game. However, using a result of Besicovitch it can be shown that it is the man who has the winning strategy, although the lion becomes infinitely close to the man (as always, we make the funny assumption that both the lion and the man are points). The solution is based on the fact that the harmonic series is divergent.

Gottlob Frege: unrestricted separation axiom. This assumption led to antinomy in set theory, and the antinomy was eliminated by a proper formulation of the separation axiom.

Lewis Carroll: heuristic rules involving resolution. Carroll suggested a method of resolving soriteses based on a kind of resolution rule. However, the rule did not work in a general case. He noticed that, and proposed a completely new approach: the method that he named *The Method of Trees*. This was in fact a prototype of the analytic tableaux method developed half a century later.

Mathematical mistakes reported in literature are numerous, and there are several factors involved, for example:

Incompetence. Some people still stubbornly try to prove for instance that problems which were shown to be insolvable can be nevertheless solved – see for example Dudley 1992, where there are numerous stories about "mathematical cranks" and their efforts.

Oversight. Special cases are occasionally forgotten when proposing a classification – see for example the case of the Perko pair in graph theory.

Suggestive drawings. When a drawing is not general enough, it may suggest wrong results.

Verbal problems. This kind of mistake is caused by the illegal use of meaning from vernacular usage in considerations about notions which are named by the term in question.

Wrong conceptualizations. Some of the troubles with the proof of the independence of Euclid's Vth postulate seem to have been caused by insufficiently clear conceptualization of what the "straight line" could mean.

Lack of sound logical foundations. When notions are understood only in an informal intuitive manner, they may lead to wrong results.

Complexity of notions. Perhaps several flaws in attempted proofs of Fermat's Last Theorem were caused by the (hidden) complexity of the problem itself.

No complete classification or typology of mathematical mistakes is possible. However, certain major types may be listed:

Material mistakes (wrong assumptions). This kind of mistake is rather easy to recognize – a careful inspection of assumptions allow us to eliminate such errors.

Formal mistakes (non sequitur). Due to the fact that mathematical proofs often use abbreviations in inferences, it can happen that such gaps contain hidden formal mistakes.

Incomplete proofs. A proof may be incomplete because not all cases were taken into account. Applying a rule as not yet approved in the mathematical community may lead to a result which will not be accepted unless the rule in question is recognized and justified.

Errors in calculations. Humans are fallible, and as such might commit errors even in simple calculations. For instance, there are reports about the incorrect calculation of a decimal expansion of π in early attempts.

Improper inductive hints. When one collects, say, a hundred thousand cases confirming a hypothesis, then one may feel pressured to accept the hypothesis. Numerous examples of such situations can be found, for example in number theory.

False analogies. Hamilton's efforts to define multiplication for triples of real numbers (in analogy with real and complex numbers) were, as is well known, unsuccessful. The operation in question is not possible, though it is possible to define multiplication for quadruples of real numbers which, however, is not commutative.

Wrong generalizations. It might seem natural to consider an analogon of the Jordan curve theorem in three dimensions. However, as the counterexample of Alexander's horned sphere shows, such generalization is wrong.

Wrong intuitions. Wishful thinking may force non-experienced mathematicians to accept beliefs leading them astray.

Correct results formerly rejected. Occasionally the mathematical community initially rejects a result which after some time and thorough investigation finally becomes accepted as true. Such doubts were formulated for example in the case of early proofs of the four color theorem.

False conjectures. It is debatable whether conjectures which ultimately turned out to be false should be counted as mathematical mistakes. A conjecture without proof is only a hypothesis, and it is not claimed to be stating a mathematical truth before it is proved. However, some conjectures may be so influential that they can drive away mathematical research.

Collapsed programs. Hilbert's program concerning the foundations of mathematics cannot be fully realized. However, it is possible to realize it partially. In this case, the collapse had fruitful consequences – see for example the development of a program of *reverse mathematics*.

Discussing the role of mistakes in the development of mathematical knowledge is justified (cf. *erro ergo disco*). There are several collections of mathematical mistakes, for example De Morgan 1915, Lecat 1935, or Posamentier, Lehmann 2013. The text of Euclid's *Pseudaria* is lost (it is mentioned by Proclus). It is presumed to have contained several examples of false proofs (sophisms) presented as warnings for the students of geometry – see Acerbi 2008.

8.2.3 Mathematical intuition in action

In this section we take into account mathematical intuitions from the third of the above discussed stages, meaning the intuitions of professional mathematicians.

Mathematical intuitions are encoded in axioms. However one should bear in mind that the path towards the formulation of axiom systems was rather long for the most important mathematical theories. The first axiom system in mathematics – that presented in Euclid's *Elements* – was practically the only one used in mathematics until the second half of the 19th century.

In the first chapter we discussed the problem of which mathematical objects are considered standard. Mathematicians often use the phrase "object X is well behaving". This means that the object in question is, in some sense, natural, standard, prototypical in the investigated domain, that it has predictable properties, and so on. Of course, well behavior in this sense is relativized to the theory in question or to its applications. Which objects are called standard also depends on tradition. An important role in establishing standards is played by several classification theorems and representation theorems as well as theorems showing that investigated objects can be transformed into suitable normal, canonical, or standard forms, and so forth.

One can draw a distinction between standard objects (as characterized above) and *generic* objects and generic properties. A property holding for "typical" objects from a given class is called *generic* and objects possessing it are called *generic* objects from this class. In this sense, "generic" usually means the same thing as "almost all", where the latter term obtains a precise mathematical meaning, depending on the context. A dual concept to "almost all" is "negligible" and its meaning is derived from the meaning associated with "almost all".

In several domains of mathematics we talk about *degenerate* objects and *limiting* cases. An object is degenerate if some of its properties take the most possible values. In general, degenerate objects in some class belong to a class of simpler objects (for example a point as a degenerate circle). We distinguish between degenerate objects and objects called *exceptions*. Exceptions are always connected with some results established within a certain mathematical theory. The most typical situation occurs when objects are classified (according to some chosen criteria) and a few objects do not fit into any of the classes.

Mathematical objects named *pathological* (sometimes also: *paradoxical*) appear as unexpected and, moreover, unwilling. There seem to be at least three typical situations in which one speaks about pathological objects in mathematics. The first two (a clash with established intuitions and new definitions of concepts formerly understood in an intuitive way) were already mentioned in chapter 1. The third case concerns pathologies created on purpose. Many objects initially called pathological were introduced intentionally, as counterexamples or as indicators that one should modify the previously accepted intuitive views. There are numerous examples of such pathologies in general topology. A famous example of a pathological object created on purpose is the (ternary) *Cantor set* which is the subset of the unit interval defined as follows:

$$C = \bigcap_{n=1}^{\infty} \bigcup_{k=0}^{3^{n-1}-1} \left(\left[\tfrac{3k+0}{3^n}, \tfrac{3k+1}{3^n}\right] \cup \left[\tfrac{3k+2}{3^n}, \tfrac{3k+3}{3^n}\right]\right).$$

The Cantor set can also be defined arithmetically, by considering triadic expansions of real numbers. The above construction is often represented graphically: we remove the central interval of the length one-third from the unit segment, then repeat this operation (in the corresponding scale) to the two remaining intervals, and so on. The Cantor ternary set is then the intersection of all sets obtained at these stages. The following may be listed among the interesting properties of this set:

1. C does not contain any interval of positive length.

2. C is uncountable.

3. All members of the Cantor set are either rational or transcendental (because they are not normal and all algebraic irrational numbers are normal).

4. C is closed, compact, nowhere dense and perfect.

Taken together, these properties may give the impression that C is a very bizarre set (and originally it was so). However, at the present moment the Cantor set is a well-recognized mathematical object and it has numerous applications in many areas of mathematics. One can prove for instance that:

1. Every separable metric space is the continuous image of a subset of the Cantor set.

2. Every compact metric space is the continuous image of the Cantor set.

3. C is homeomorphic to 2^{ω}.

Some other well-known examples of pathologies include the space filling curves. These are functions whose graph fills an area with positive measure – for example a square. Famous examples of such curves were given by Peano and Hilbert. The functions in question are continuous (as limits of uniformly converging sequences of continuous functions) but they are not differentiable at any point. In chapter 1 we also mentioned further objects, currently still recognized as pathological, in a sense: Alexander's horned sphere, exotic spheres and exotic $\mathbb{R}^4$.

Unlike the intuitions related to vernacular experience (and to common sense), mathematical intuitions are flexible. This is true primarily of third-stage intuitions, that is the intuitions of professional mathematicians. This should be obvious because the changes in this case are related to development in mathematics, to the emergence of new ideas and to the discoveries of new results. Experts' intuitions are based on creativity of the mind.

Factors responsible for changes in mathematical intuitions are numerous; among the most important are the following:

Antinomies and paradoxes. Inconsistent theories are of no use, and hence once we find an antinomy in a theory we should take every effort to avoid it. For instance, the intuitive assumption that any property defines a set causes an antinomy and therefore we are forced to change that intuition by assuming a new form of the axiom of separation in set theory.

Thus antinomies require changes in the underlying theory. Paradoxes, in turn, are responsible for changes in our intuitions but without change in the theory. Paradoxes show that our initial intuitions were insufficiently subtle to understand a given problem. For instance, the Bertrand paradox shows that we have to base our notion of probability on the notion of measure – probability is not an absolute notion, it depends on the previously accepted measure.

Scientific programs. The history of mathematics has frequently seen the formulation of an entire elaborated program for further research. Examples are:

1. *Arithmetization of analysis.* The origins of the differential and integral calculus were lacking a sound logical basis (according not only to the current criteria of correctness, but also to those of that era). The intuitions concerning infinitely small quantities or the convergence of infinite series, for example, were still *in statu nascendi*, so to speak. The replacement of kinematic and geometric references in calculus by precisely formulated arithmetical counterparts was considered in the 19th century as a remedy to that foundational crisis.

2. *Algebraic topology.* Not long after the development of the foundations of general topology, it became clear that many topological problems could be resolved much more easily with the use of suitable algebraic counterparts of topological notions.

3. *Geometrization conjecture.* Thurston's geometrization program is an example of setting up new directions in mathematics. Out of the seven Millenium Problems proposed by the Clay Institute, only one (the Poincaré conjecture) has been solved so far (quite recently). The problem itself is closely related to Thurston's program.

4. *Axioms of the existence of large cardinal numbers.* We have already discussed the reasons for accepting those axioms in the previous chapter. They are useful not only with respect to metamathematical considerations (such as the consistency strength of theories), but they also shape anew our intuitions about sets.

5. *Hilbert program.* David Hilbert formulated an ambitious program for the foundations of mathematics. Its goal was to prove the consistency and completeness of the most fundamental mathematical

theories. After results obtained by Gödel, Turing, Tarski and others it became clear that the program could not be fully realized. Thus our intuitions concerning certain metamathematical properties changed due to those results.

6. *Hilbert problems.* In 1900 David Hilbert formulated a list of 23 problems fundamental for mathematics. His vision appeared successful – a tremendous amount of mathematical research in the 20th century focused on these problems.

New definitions. The fact that some collections are equinumerous with their proper parts was long considered paradoxical. Dedekind's idea of calling this a definitional property of infinite sets led to the emergence of a new intuition concerning the notion of infinity.

New results. Mathematical intuitions are flexible and the growth of mathematical knowledge is one of the most important factors of this flexibility. The discovery of non-Euclidean geometries caused an essential change in our thinking about space.

Aesthetic values. Most mathematicians declare that mathematical beauty strongly influences their research. It is possible to characterize the beauty of mathematical ideas and constructions with the help of such notions as generality, simplicity, symmetry, and others. What is interesting is that mathematicians share aesthetic valuations, meaning that these values are not only subjective but there is a common agreement about them.

Empirical experiments. Inspirations for mathematical discoveries frequently derive from the natural sciences, most notably from physics. The origins of arithmetic lie in the processes of counting, the origins of geometry lie in measuring lengths, areas and volumes, and the origins of differential and integral calculus are influenced by reflections on motion, change, and velocity, and so on.

Mathematical fashion. This is a sociological factor, not belonging to mathematics itself. It can be seen for instance in overemphasis of the role that certain ideas play for mathematics as a whole. This was the case, for instance, with the invention of many-valued logics or the introduction of quaternions.

The cognitive accessibility of mathematical objects can be investigated in at least two aspects:

1. *External aspect.* This aspect depends on human cognitive abilities, for instance: abstraction, generalizations, building mental representations, or the ability to conduct reasoning.

2. *Internal aspect.* This aspect depends on the role played by the object in question in mathematics. Those objects which are fundamental for the creation and development of mathematical theories are becoming more and more accessible.

In the second of the above cases one may ask about *measures* of accessibility which characterize degrees of accessibility of mathematical objects. Examples of such measures include:

Logical complexity. Mathematical notions can be classified with respect to the degree of complexity of their definitions measured for example by the number of quantifiers in the definitions (arithmetical hierarchy, analytical hierarchy). For instance, the notion of limit requires three quantifiers, the notion of truth in the standard model of arithmetic is beyond the whole arithmetical hierarchy. Functions are classified into Baire classes – for instance the Dirichlet function (the function which takes value 1 for any rational argument and value 0 for any irrational argument) belongs to the second Baire class: $f(x) = \lim_{k\to\infty} \lim_{j\to\infty} (\cos(k!\pi x))^{2j}$.

Effectiveness of construction. Objects which are obtained by an effective construction may be considered more accessible than, for example, objects whose construction requires the use of the axiom of choice.

Categoricity or completeness. One may accept the opinion that objects characterized in a categorical way (that is, up to isomorphism) are more accessible than objects lacking such characterization. Similarly, objects indistinguishable with respect to elementary equivalence can be seen as more accessible than objects lacking such characterization.

Multiplicity of representations. Objects which have several representations are better "domesticated" in mathematics. For example, rational numbers can be represented as the (minimal) field of characteristic zero, a tree-like structure (the Stern-Brocot tree or Calkin-Wilf tree), a dense subset of real numbers (dense either in the topological sense or dense as an ordered subset), a geometrical structure (all straight lines through the point $(0, 0)$ and lattice points), and so on.

Fruitfulness in applications. If an object finds many applications in several domains of mathematics, then it also becomes more "domesticated". For example, the Cantor set which at the moment of its discovery (creation?) was considered a pathological object, later became one of the fundamental structures not only in topology but also in other mathematical domains.

Access to mathematical objects is connected with important oppositions: finite–infinite, discrete–continuous, regular–random, and computable–uncomputable. The first members of these oppositions correspond to more accessible objects. The regular–random opposition is a more difficult case, because neither regularity nor randomness has a commonly accepted formal definition.

8.3 Teaching practice

Let us turn now to the mathematical intuitions from the second stage, and so to intuitions shaped in the process of education.

The main goals of mathematical education include "forming proper mathematical intuitions", as frequently emphasized in many official curricula. This can be viewed from two perspectives: that of the teacher and that of the student.

8.3.1 Teaching mathematics: selected views

For obvious reasons most of the literature on mathematical education concerns teaching mathematics at a very elementary level. Our own teaching experience is limited to teaching mathematics at university level. In addition, we face the difficulty of teaching mathematics to students of humanities (mainly cognitive science), and this audience is, to a certain degree, hostile to mathematical considerations. It is sometimes very hard to convince this audience that mathematical thinking is essential in any problem-solving activity.

It is a sad and to a certain degree mysterious fact that most of the population finds it hard to understand mathematics. Poincaré wrote:

> How is it that there are so many minds that are incapable of understanding mathematics? Is there not something paradoxical in this? Here is a science which appeals only to the fundamental principles of logic, to the principle of contradiction, for instance, to what forms, so to speak, the skeleton of our understanding, to what we could not be deprived of without ceasing to think, and yet there are people who find it obscure, and actually are the majority. That they should be incapable of discovery we can understand, but that they should fail to understand the demonstrations expounded to them, that they should remain blind when they are shown a light

> that seems to us to shine with a pure brilliance, it is altogether miraculous.
>
> And yet one need have no great experience of examinations to know that these blind people are by no means exceptional beings. We have here a problem that is not easy of solution, but yet must engage the attention of all who wish to devote themselves to education. (*Science and Method*, citing Sierpińska 1994, 112)

Our own teaching practice is of course influenced by the results, advice and hints contained in literature on the subject, in particular in the Polish tradition of teaching mathematics (see for example Krygowska 1967). This is not the place to report extensively on that research, and we shall mention just a few selected standpoints:

Piaget. His works in cognitive psychology (for example Piaget 1970, 1977) influenced many teachers of mathematics – numerous references to Piaget's results are still to be found in literature focusing on mathematical education. Teachers rely on the stages in cognitive development suggested by Piaget when preparing their programs: the sensorimotor stage, preoperational stage, concrete operational stage, and formal operational stage (see for instance Ojose 2008).

Polya. He became famous for his analysis of the process of mathematical problem solving (Polya 1957, 2009, 2014). The rules he proposed may be summarized as follows:

1. First, you have to understand the problem.
2. After understanding, then make a plan.
3. Carry out the plan.
4. Look back on your work. How could it be better?

These general pieces of advice are further elaborated, giving more detailed hints, for example the following in the case of plan making:

Guess and check	Look for a pattern
Make an orderly list	Draw a picture
Eliminate possibilities	Solve a simpler problem
Use symmetry	Use a model
Consider special cases	Work backward
Use direct reasoning	Use a formula
Solve an equation	Be creative

Schoenfeld. Alan Schoenfeld is most famous for his works concerning mathematical problem-solving strategies (Schoenfeld 1985). He stresses the role of heuristic procedures in problem solving and the importance of monitoring our problem-solving activity, which may be called *metacognition.* An important piece of practical advice which should be followed in mathematical education is to convey it in such a way that at any moment students are convinced about the sense of what they are doing.

Sierpińska. Anna Sierpińska applied Kazimierz Ajdukiewicz's notion of understanding to mathematical concepts, constructions, proofs, and problems, and so on. Ajdukiewicz's original formulation runs as follows: *a person understands an expression if on hearing it he directs his thoughts to an object other than the word in question* (Ajdukiewicz 1974, 7). Design features of Sierpińska's approach may be summarized as follows:

1. Understanding is a relation holding between the *object of understanding* and *basis of understanding.*

2. Mathematical objects are *seen as* something and not plainly *seen.*

3. The *objects of understanding in mathematics* include: concepts, problems, formalism, and text.

4. The *basis of understanding* includes: representations (for instance mental images or conceptual representations, procedural representations), mental models, thoughts that (so and so), and apperception.

5. *Mental operations* involved in understanding include: identification, discrimination, generalization, and synthesis.

6. *Attention* and *intention* are necessary conditions for the act of understanding.

7. The author points to the role of *epistemological obstacles* in the process of understanding.

Tall. David Tall promotes his vision of the *three worlds of mathematics* (Tall 2013). They appear as the result of combining two classifications. On the one hand the author indicates the following division:

1. *Practical Mathematics*: with experiences in shape and space and arithmetic.

2. *Theoretical Mathematics*: with definitions based on known objects and operations.

3. *Formal Mathematics*: with formal objects based on formal definitions.

On the other hand, Tall sees the partition of mathematical activities into conceptual embodiment, operational symbolism and axiomatic formalism.

Important notions introduced by Tall are those of *pro-cept* and *met-before*. Pro-cepts emerge as a result of transforming a procedure (for example summation) into a concept (correspondingly, the concept of sum). Met-befores refer to that which is known to the student before she acquires a new concept. Beliefs included in particular met-befores may influence (positively or negatively) the ability to understand a new concept.

8.3.2 The context of transmission

We propose adding a new context to the traditionally discussed contexts of discovery and justification, one which we tentatively name the *context of transmission* (of mathematical knowledge, skills and abilities).

The context of discovery is responsible for the creation of new mathematical facts. The context of justification provides methods for establishing their validity. However, mathematics is also present in culture in different roles than just the creative work of professional mathematicians. There are at least five domains which should be taken into account here:

1. *Teaching mathematics.*

2. *Learning mathematics.*

3. *Popularizing mathematics.*

4. *Applications of mathematics in art and design, and so forth.*

5. *Philosophical reflection on mathematics.*

All the above belongs to the context of transmission, which includes situations when we talk *about* mathematics. The first three of the mentioned domains are of special interest here. In all of them the focus is on the *transmission* of mathematical knowledge, skills and abilities. It is obvious that in order to achieve this goal we can apply tools which transcend mathematics itself. We find ourselves on the metalevel, so to speak. Anna Sierpińska writes in this respect:

> The quest for an explanation in mathematics cannot be a quest for proof, but it may be an attempt to find a rationale of a choice of axioms, definitions, methods of construction of a theory. A rationale does not reduce to logical premisses. An explanation in mathematics can reach for historical, philosophical, pragmatic arguments. In explaining something in mathematics, we speak *about* mathematics: our discourse becomes more metamathematical than mathematical.(Sierpińska 1994, 76)

In our opinion, the most important role in the context of transmission is played by *intuitive explanations*. We propose thinking about the problem as follows:

1. The meaning of mathematical concepts is determined by the underlying theory (ultimately, an axiomatic one).

2. Intuitive explanations connect that meaning with mental representations.

3. An understanding of a mathematical concept (such as a problem, proof, or construction) is obtained as a result of joining the (abstract) meaning together with a bunch of its intuitive explanations.

We distinguish several types of intuitive explanation. The most important are the following:

Linguistic explanations. Mathematical ideas are expressed in a formal language but their exposition in teaching is possible only with the help of verbal explanations. Of course, linguistic explanations are not limited to "translations" of formal expressions into an ethnic language. The reading of a formal expression sometimes contains a certain redundancy of content, as for example when the phrase $\forall\varepsilon > 0$ is read as "for all positive *arbitrarily small* ε" (in definitions of limits or continuity, and so on). The proper choice of mathematical terminology often appears helpful – for example such terms as "dense set", "nowhere dense set", and "overspill lemma". A metaphorical mode of speaking is also useful at times, for example when we say that a real variable moves, walks, tends, approaches, runs, flies, or swims (but not that it jumps) in the continuous domain. All these metaphors are innocent as long as we remember that behind the linguistic usage there lies a precise meaning: the variable accepts (or takes) particular values in the domain.

Perception. These explanations primarily use drawings, schemas, pictures and the like. Pictures do not replace formal proofs but may significantly help the students better understand the matters in question. Pictures should not be too suggestive, however. It is also important to make the students aware of the fact that we use several conventions (for example in representations of solid objects on the plane) and that the picture itself depends on the surface on which it is drawn. Visual representations are helpful for instance in complex analysis, elementary topology, and the theory of knots (see Needham 1997, Prasolov 2011, Adams 2004).

Physical models. Inspirations for the creation of large parts of mathematics come from investigations in physics. It is thus natural to explain mathematical ideas going back to their physical roots.

A sharp distinction between *examples* from physics and *explanations* based on physics is difficult to describe. Perhaps such alleged distinction is an illusion, and when referring to physical situations in mathematical education we always work with examples.

People may have very bizarre images or representations concerning the behavior of physical bodies. For example, as Talia Ben-Zeev and Jon Star write:

> It has been found that when people were asked to draw the path of a moving object shot through a curved tube, they believed that the object would move along a curved (instead of a straight) path even in the absence of external forces. (Ben-Zeev and Star 2001, 29)

Such views are classified as belonging to *folk physics*, also known as *naive physics*. The views in question are very often simplifications or misunderstandings of real phenomena, and they are investigated in cognitive psychology. Quite where they originate from is an interesting matter.

References to physics in mathematical works were already present in antiquity (Euclid's reference to motion or Archimedes' references to balances and centers of mass). Nice collections of such arguments can be found in several modern works, including Levi 2009 (mechanical models illustrating several mathematical theorems), Ghrist 2014 (linkages), Havil 2007 (several phenomena in which the appearance of harmonic series is essential), Polya 2014 (variational problems), or Laraudogoitia 1996 and Romero 2014 (supertasks). For example:

1. Center of mass may be useful in argumentations associated with theorems in geometry.
2. Effects connected with gas pressure can be helpful in illustrations related to curvature problems (like Gauss-Bonnet Theorem).
3. Results from complex analysis (like, say, the Cauchy Integral Formula or the Riemann Mapping Theorem) can be represented by fluid flow or heat flow.
4. Linkages are capable of transforming rotary motion into perfect straight-line motion, and vice versa, thus illustrating mathematical transformations in terms of movement.
5. Buffon's needle experiment, the Bertrand paradox, Galperin's billiard problem and the like are examples of physical contexts illustrating mathematical ideas in measure theory, probability and approximation theory.
6. Intuitive explanations of some variational problems may include for instance *minimal surfaces* – the mathematical description can be accompanied by physical models (say, soap bubbles).

Common sense. Explanations may refer to everyday experience and common sense. Thus we often imagine topological objects as "rubber-like", we interpret operations on negative numbers as operations on debts (or floors going downwards, temperature, game scoring, jumping frogs, and so on). Intuitive explanations in algebra or topology can refer to our perception of size. For example, *filters* are often said to consist of *big* elements, while *ideals* consist of *small* elements.

Internal explanations. Notions (ideas, constructions, and so forth) from one domain of mathematics can be explained by reference to another domain. For instance, textbook introduction of the notion of derivative (of a function of one real variable) is usually accompanied by illustrative reference to geometrical facts (tangent of curve).

Geometrical representations of natural numbers can be used in cases, when several summation formulas are illustrated by suitably chosen diagrams depicting triangular, quadratic, pentagonal, or hexagonal numbers (see for example Conway and Guy 1998). Formulas such as:

1. $1 + 2 + 3 + \ldots + n = \frac{n \cdot (n+1)}{2}$
2. $1 + 3 + 5 + \ldots + (2n - 1) = n^2$

are usually proven by mathematical induction. Geometrical illustrations accompanying such proofs are likely to contribute to a better understanding of the processes of summation being investigated.

The introduction of some very sophisticated notions in set theory may be accompanied by reference to (simpler) algebraic constructions – this is the case when the method of *forcing* and *generic sets* is explained by pointing to the analogy with transcendental extensions of fields.

Davis and Hersh discuss an interesting case concerning cross-domain explanation related to the Riemann hypothesis (Davis and Hersh 1981). This hypothesis can be shown to be equivalent to a certain statement in which one uses the *intuitive* assumption that Mertens' function could be considered a random variable. In consequence, one obtains the claim that the Riemann hypothesis is true *with probability one.*

The critique of restriction axioms in set theory described in the previous chapter also made use of an *intuitive* piece of advice: we give up the axioms of restriction and accept the axioms of the existence of large cardinals because we want the universe of sets to be as rich as possible, in an analogy with Hilbert's axiom of completeness in geometry.

8.3.3 Mathematical puzzles

Finally, let us discuss a special method in mathematical education used for the development of correct mathematical intuitions, namely that based on the consideration of *mathematical puzzles.* We have in mind here a special sort of such puzzles, namely those whose very formulation suggests a simple solution based on everyday experiences and intuitions and which at the same time is illusory. The first answer to the puzzle, given without reflection, is simply wrong from a mathematical point of view and only after a deeper analysis is the correct answer obtained. To put it briefly, such puzzles may be called paradoxical (with respect to everyday experiences and beliefs). They also show explicitly the difference between quick and slow thinking in solving problems (in the sense considered in cognitive science).

There is a clear difference between mathematical puzzles and routine mathematical exercises. Solving exercises is recommended when students are supposed to master their mathematical skills and achieve a full understanding of newly-introduced concepts and theorems. A typical exercise demands calculation (sometimes a lot) or the presentation of standard proofs. After calculating a few dozen integrals the students should be confident in their understanding of the process of integration. In order to fully grasp the method of proofs by mathematical induction, the student should go through several such proofs. The students may get the feeling that they understand natural deduction only after they have conducted many particular proofs in this system by themselves. Thus exercises should be conceived of as tools for mastering standard techniques. On the other hand, a typical mathematical puzzle should involve the following design features:

1. It should contain an unusual plot and an interesting story, although the latter should be presented in everyday language.

2. In order to solve a puzzle, one is supposed to be creative and not only follow, say, a prescribed algorithm.

3. The solution should be surprising, unexpected, and revealing a new insight.

Contrary to the usual mathematical exercises, mathematical puzzles are often connected with that which is unexpected, that which contradicts our everyday experience. As such, puzzles are instructive as far as a critical attitude towards informal intuitions is concerned. They teach us that we should be cautious in relying on these intuitions, which are sometimes very illusory.

In literature, mathematical puzzles are sometimes called conundrums, brain-teasers, riddles, or curiosities, and so on. We are not interested in all such intellectual entertainments (recreational puzzles); we take into account only "serious" puzzles, those which are responsible for showing the distinction between jumping to a conclusion based on (illusory) intuitions and an analytical approach to the problem in question, ultimately leading to the correct answer.

The problems discussed during our class on mathematical problem solving were classified into groups:

The Infinite	Numbers and Magnitudes	Change and Motion
Shape and Space	Algorithms and Computation	Patterns and Structures
Orderings	Measure and Probability	Games and Strategies

In addition, we also discussed the following in our course:

1. *Paradoxes*. Something that seems to be false, but nevertheless (after proper analysis) reveals itself to be true.
2. *Sophisms*. Deliberately constructed fallacies.
3. *Illusions*. Primarily optical illusions, but also illusions occurring as the result of fallacious argument.
4. *Paralogisms*. These are mistakes committed without intention – in contradistinction to intentionally created sophisms.

The puzzles were chosen in order to achieve a better understanding of mathematical concepts involved in their solutions. That is say, we pointed to typical contexts in which a concept in question occurs, we showed how to operate with this concept, and sometimes we discussed the origin of the concept.

Contemporary authors of mathematical puzzles include Martin Gardner, Hugo Steinhaus, Raymond Smullyan, John Conway, Ian Stewart, Julian Havil, Peter Winkler, and Presh Talwalkar. For some time now thousands of math puzzles have been accessible in the internet. Among others, the following post recommendable educational movies on YouTube: *Mind your decisions*, *Mathologer*, *3Blue1Brown*, *Numberphile*, *Khan Academy*, *TED-ED*, *PBS Infinite Series*, *TEDx Talks*, *Math Antics*, *World Science Festival*.

We presented a selection of puzzles used during the course in the paper Pogonowski 2018a. Let us recall three examples, all of them surprising for the audience in the sense that the expected solution disagrees with the intuitive view.

Ant on the rubber rope

This cute puzzle has several versions, a typical one being the following:

> An ant starts to crawl along a taut rubber rope 1 km long at a speed of 1 cm per second (relative to the rope it is crawling on), starting from its left fixed end. At the same time, the whole rope starts to stretch with the speed 1 km per second (both in front of and behind the ant, so that after 1 second it is 2 km long, after 2 seconds it is 3 km long, etc). Will the ant ever reach the right end of the rope?

It should be stressed that this is a purely mathematical puzzle; we ignore the ant's mortality, and we assume that infinitely elastic ropes exist, and so forth.

People usually doubt that the ant could achieve the goal in a finite period of time. However, the answer is affirmative; the ant certainly will reach the right end of the rope, though it takes a really long time interval.

The dynamic aspects of the problem may cause some difficulties in its solution. In general, one should solve a (rather simple) differential equation describing the motion in question. However, the problem can also be approached in a discrete manner, as follows.

The main question is: which part of the rope is crawled by the ant in each consecutive second? It is easy to see that:

During second	the ant crawls	part of the whole rope
first	1cm out of 1km	$\frac{1}{100000}$
second	1cm out of 2km	$\frac{1}{200000}$
third	1cm out of 3km	$\frac{1}{300000}$
n-th	1cm out of $n$km	$\frac{1}{n \cdot 100000}$

Hence the problem reduces to the question of existence of a number n such that:

$$\frac{1}{100000} + \frac{1}{200000} + \frac{1}{300000} + \ldots + \frac{1}{n \cdot 100000} \geqslant 1.$$

This is of course equivalent to the existence of n such that:

$$1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} \geqslant 100000.$$

Recall the *harmonic numbers*: $H_n = \sum_{k=1}^{n} \frac{1}{k}$. We know that the *harmonic series* $\sum_{n=1}^{\infty} \frac{1}{n}$ is divergent. Here is a simple argument of its divergency:

$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} + \frac{1}{9} + \frac{1}{10} + \ldots >$
$1 + \frac{1}{2} + (\frac{1}{4} + \frac{1}{4}) + (\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}) + (\frac{1}{16} + \ldots + \frac{1}{16}) + \ldots =$
$1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \ldots = \infty.$

Therefore, there exists a number n such that $1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} \geqslant 100000$. This number is really huge, it equals approximately $e^{100000-\gamma}$, where γ is the Euler-Mascheroni constant:

$$\gamma = \lim_{n\to\infty} (\sum_{k=1}^{n} \frac{1}{k} - \ln n) = 0,5772156649501\ldots$$

This constant remains a little bit mysterious – for instance, we do not know at the present whether it is rational or irrational.

Essential in this puzzle is the fact that the considered velocities are constant. If, for instance, the rope is doubled in length at each second, then the poor ant has no chance of reaching the right end of the rope (crawling, as before, with constant speed). The puzzle also has interesting connections with recent views concerning the Universe. Remember: the space of the Universe is expanding, but the speed of light is constant. What are the consequences of these facts for the sky viewed at night in a far, far future?

The fact that the harmonic series is divergent may be used in formulations of several further puzzles, for example Gabriel's horn (infinite surface embracing a finite volume), the jeep problem (travel through a desert of any breadth with a limited amount of fuel), and so on.

The next term of a sequence

There are many popular mathematical puzzles concerning the discovery of regularities in a given set of numerical data. However, one should treat them with caution, as the following example clearly shows (Klymchuk and Staples 2013).

The first terms of a sequence are: 2, 4, 8, 16. Find (or guess) the next term. Such a formulation is misleading, because – given the data – one can define the fifth term of the sequence so that some kind regularity will be visible. Of course, the first – intuitive – answer could be 32, because the four first terms are evidently consecutive powers of 2. But let us consider the following formula for the n-th term of the sequence:

$$a_n = 2^n + (n-1)(n-2)(n-3)(n-4)x.$$

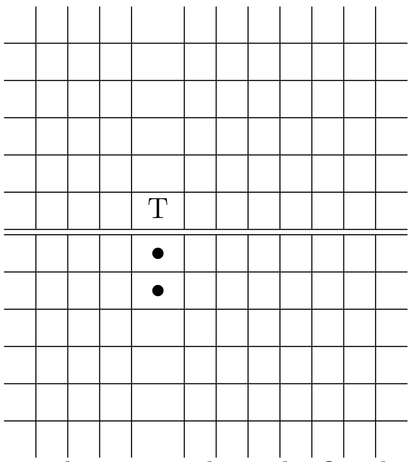
In this case the first four terms are obviously 2, 4, 8, 16 but if we put $x = \frac{a-32}{24}$, then a can be arbitrary.

Conway's army

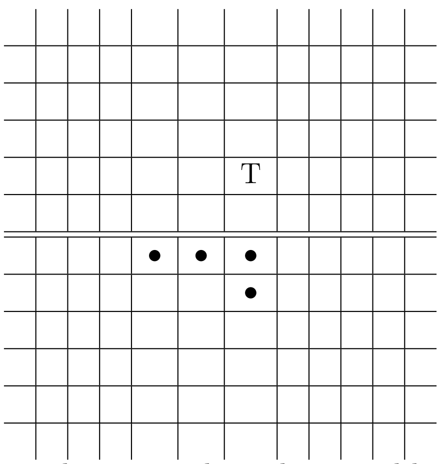
John Horton Conway is famous for his inventiveness in creating mathematical puzzles. Who hasn't hear of his celebrated *Game of Life*? Conway also proposed a game now known as *Conway's army*. He actually proved a surprising fact about this game, which may, at first glance, seem very unexpected.

The game is played on an infinite board; just imagine the whole Euclidean plane divided into equal squares and with a horizontal border somewhere. You may gather your army of checkers below the border. The goal is to reach a specified line above the border. The checkers move only vertically or horizontally. Thus diagonal moves are excluded. As in the genuine checkers, your soldier jumps (horizontally or vertically) over another soldier on an adjacent square (which means that he kills him), provided that it lands on an unoccupied square next to the square occupied previously by the killed soldier.

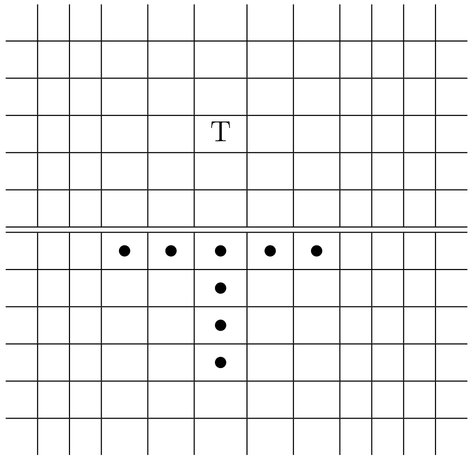
It is easy to show that one can reach the first, second, third and fourth line above the border. Here are examples:



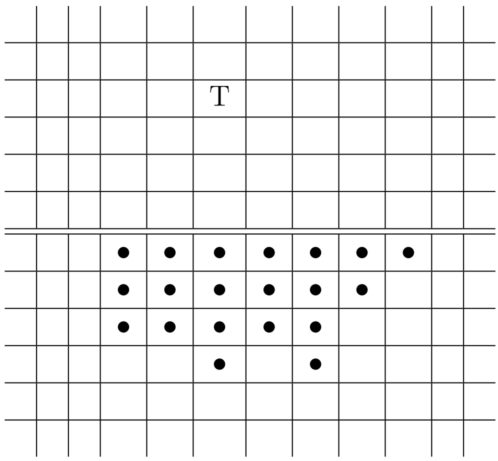
Minimal army reaching the first level.



Minimal army reaching the second level.



Minimal army reaching the third level.



Example of a minimal army reaching the fourth level.

However, no finite amount of soldiers gathered below the border can ever reach (by at least one surviving soldier) the fifth line above the border! The main idea of the proof is as follows (Havil 2007).

We consider the Manhattan metric on the plane. The target T to be reached, a chosen square on the fifth line above the border, is given numerical value 1.

			x^1	T	x^1			
		x^3	x^2	x^1	x^2	x^3		
x^6	x^5	x^4	x^3	x^2	x^3	x^4	x^5	x^6
	x^6	x^5	x^4	x^3	x^4	x^5	x^6	
	x^7	x^6	x^5	x^4	x^5	x^6	x^7	
	x^8	x^7	x^6	x^5	x^6	x^7	x^8	
		x^8	x^7	x^6	x^7	x^8		
			x^8	x^7	x^8			
				x^8				

Distances of checkers from the target on the fifth level.

Your army is described by a formal polynomial in an unknown x. The Manhattan distance of a given soldier x to the target T is coded in the exponent of x. For instance, if your soldier is two steps from T (horizontally or vertically) than it becomes x^2, if it is on the line just below the border and vertically just below T, then it becomes x^5, etc. The soldiers with their exponents are added and in this way your army looks like a polynomial. For instance, the reader may check that the polynomials for the armies in the first three cases look as follows:

1. $P_1: \ x^5 + x^6$ (it can reach the first line above the border)
2. $P_2: \ x^5 + 2x^6 + x^7$ (it can reach the second line above the border)
3. $P_3: \ x^5 + 3x^6 + 3x^7 + x^8$ (it can reach the third line above the border).

The rules of moving soldiers can be summarized as follows:

1. $x^{n+2} + x^{n+1}$ is replaced by x^n
2. $x^n + x^{n-1}$ is replaced by x^n
3. $x^n + x^{n+1}$ is replaced by x^{n+2}.

We will choose the (positive) value for x in such a way that the move from point 1 above does not change the total value of the polynomial describing the army and the moves from points 2 and 3 diminish this value.

Because $x > 0$, we have $x^n + x^{n-1} > x^n$. If we want $x^n + x^{n+1} > x^{n+2}$, then $1 + x > x^2$ and this gives the inequality $0 < x < \frac{1}{2}(\sqrt{5} + 1)$. Moreover, we want that $x^{n+2} + x^{n+1} = x^n$. This means that $x + x^2 = 1$. Hence the choice $x = \frac{1}{2}(\sqrt{5} - 1)$ satisfies all our requirements and we have $x + x^2 = 1$. Now, each army is described by a formal polynomial described above. Its total value is of course less than the sum P of an *infinite* army, occupying *all* the squares below the border.

					T						
x^{10}	x^9	x^8	x^7	x^6	x^5	x^6	x^7	x^8	x^9	x^{10}	x^{11}
x^{11}	x^{10}	x^9	x^8	x^7	x^6	x^7	x^8	x^9	x^{10}	x^{11}	x^{12}
x^{12}	x^{11}	x^{10}	x^9	x^8	x^7	x^8	x^9	x^{10}	x^{11}	x^{12}	x^{13}
x^{13}	x^{12}	x^{11}	x^{10}	x^9	x^8	x^9	x^{10}	x^{11}	x^{12}	x^{13}	x^{14}
x^{14}	x^{13}	x^{12}	x^{11}	x^{10}	x^9	x^{10}	x^{11}	x^{12}	x^{13}	x^{14}	x^{15}
x^{15}	x^{14}	x^{13}	x^{12}	x^{11}	x^{10}	x^{11}	x^{12}	x^{13}	x^{14}	x^{15}	x^{16}

The infinite army below the border.

Let us calculate the sum P:

1. $P = x^5 + 3x^6 + 5x^7 + 7x^8 + \ldots = x^5(1 + 3x + 5x^2 + 7x^3 + \ldots)$.
2. Let $S = 1 + 3x + 5x^2 + 7x^3 + \ldots$. Then:
3. $xS = x + 3x^2 + 5x^3 + 7x^4 + \ldots$
4. $S - xS = S(1 - x) = 1 + 2x + 2x^2 + 2x^3 + \ldots$
5. $S(1 - x) = 1 + 2(x + x^2 + x^3 + \ldots)$
6. $S(1 - x) = 1 + \frac{2x}{1-x} = \frac{1+x}{1-x}$
7. $S = \frac{1+x}{(1-x)^2}$
8. Hence $P = \frac{x^5(1+x)}{(1-x)^2}$ (because $P = x^5S$).

Remember that we have $x + x^2 = x(1 + x) = 1$, and thus $1 + x = \frac{1}{x}$ and $1 - x = x^2$. We see that:

$$P = \frac{x^5(1+x)}{(1-x)^2} = \frac{x^5(\frac{1}{x})}{(x^2)^2} = \frac{x^4}{x^4} = 1.$$

This means that the value of any *finite* army below the border must be less than 1. No acceptable move ever increases the value of the marching army and hence it is impossible that any soldier (from a finite army) will ever reach the fifth line above the border.

There exist several generalizations of this game with their own limitations as far as the accessible level above the border is concerned.

* * *

It seems, judging from the discussions with students and from the content of their essays that puzzle solving – including brain-storming and group discussion on the way to solutions proved a very effective way for achieving improvement in students' mathematical skills.

A final word

As the reader has surely noticed, this book does not offer new solutions to the problems concerning the characterization of intended models of theories. Neither does it propose a new approach to extremal axioms. We have endeavored to gather together certain results already present in subject literature. We mention the first formulations of the extremal axioms and discuss the attempt by Carnap and Bachmann towards a general logical approach to such axioms. The said authors considered the axiom of completeness in Hilbert's system of geometry, the induction axiom in arithmetic and Fraenkel's axiom of restriction in set theory. Perhaps a certain novelty of our presentation is its more comprehensive list of extremal axioms (compared to that investigated by Carnap and Bachmann). To the ones just mentioned, we have added Gödel's axiom of constructibility, Suszko's axiom of canonicity, von Neumann's axiom of the limitation of size, axioms of the existence of large cardinal numbers (those proposed some time ago by Zermelo as well as the ones proposed much later) and Ehrlich's axioms constituting a generalization of Dedekind's ideas. We have pointed to the role of isomorphism theorems in algebra, which – in a sense – are counterparts of extremal axioms. Further expansion of this list cannot be ruled out, for instance in the case of extensions of algebraic structures.

The characteristics of the fundamental mathematical structures proposed in the late 19th century (and made more precise later) are still considered the common standards. This concerns axiomatic descriptions and constructions of number systems (natural numbers, integers, rational, real and complex numbers) as well as improved systems of Euclidean geometry (Hilbert's and Tarski's axioms, among others). Set theory, originated by Cantor and axiomatized by Zermelo (with improvements by Skolem, Fraenkel, and von Neumann), belongs among the standard tools of contemporary professional mathematicians. In their research practice they make free use of natural numbers as the uniquely determined least inductive set. Similarly, the real numbers are uniquely (up to isomorphism) characterized as the completely ordered field (which is at the same time the maximal Archimedean field). Complex numbers form the unique algebraically closed field of characteristic zero whose transcendence degree over the field of rational numbers equals continuum. The ordering of rational numbers, in turn, was already characterized in a unique way by Cantor. Finite dimensional Euclidean spaces are considered as vector spaces over the field of real numbers and have ubiquitous applications in many areas of mathematics (remember that the dimension of such a vector space determines it uniquely). "Normal" mathematicians (meaning those who do not conduct their research in set theory itself or, more generally, in the foundations of mathematics) make free use of the whole cumulative hierarchy of sets, though in many cases the hierarchy of constructible sets is sufficient in that respect.

All this does not mean that the recently accepted set theoretical paradigm along with the standard characterization of the real numbers (as an arithmetical representation of the continuum) will not change in the future. It cannot be excluded that usage of both standard real numbers and hyperreal numbers (as representations of the continuum), depending on the purposes of research and accepted points of view, will be attractive for the mathematics of the future. The decisive role is always played by mathematical research practice. The same concerns systems of set theory; in some important mathematical domains the approach based on category theory seems more fruitful than the set theoretical approach.

We do not refer in the book to the intense discussion in the philosophy of science, inspired by the work Putnam 1980, dealing with the notion of intended model in applications of mathematics to science. We share the critical opinion of Jan Woleński about Putnam's argumentation (see Woleński 2005, 480–481). The intended model of a theory (as well as its standard model) cannot be characterized syntactically or semantically within the theory itself; one has to enter the level of metatheory for that purpose.

In chapter 4 we reported on certain results concerning categoricity (categoricity in power) and completeness obtained in model theory. Recent investigations of those issues, in the case of first-order logic as well as certain stronger logics, are highly sophisticated and it was not possible to take them into account in this book. Readers interested in the topic are kindly invited to consult modern textbooks on model theory and comprehensive surveys; we recommend the very interesting paper Baldwin 2014. As we have seen in chapter 4, one can associate different senses with the intuitive notion of minimality or maximality of structure. The investigation of the space of types realized in models as well as the family of sets definable in models is useful for that purpose. For instance, the atomic models are "poor" and the saturated models are "rich".

Is it possible to obtain a general logical characterization of extremal axioms including those not taken into account by Carnap and Bachmann? In our opinion it is doubtful whether one could achieve that goal. But it is also questionable whether one needs such a characterization at all. The limitative theorems mentioned in chapter 3 revealed the possibilities and limitations of the deductive method by showing the incompleteness of sufficiently rich mathematical theories. Modern set theory is still a relatively young discipline, and it may undergo modification in the near future. The relationships between the notions of categoricity and completeness, which were of great interest to Carnap, have been shown in a new perspective in modern model theory. The characterization in question is of course a logical challenge, but it does not seem to be of essential relevance for the mathematical research itself.

The last chapter focused on mathematical intuition and thus only partially related to the extremal axioms themselves. As said before, mathematical intuitions (those of professional mathematicians but of course emerging from the intellectual activities in all subjects) are among the most important factors in the process of forming intended models. Philosophical reflections concerning mathematical intuitions are obviously valuable on their own, but (in our opinion at least) remarks expressed by mathematical educators who observe and modify the emergence of the mathematical intuitions of pupils and students are often more interesting. Intuitive explanations are essential in the context of transmission (of mathematical knowledge, skills and abilities). The notion of the *context of transmission* was introduced in Pogonowski 2016 and elaborated a little in Pogonowski 2018b. It remains the subject of our current research.

Bibliography

Acerbi, F. (2008). Euclid's Pseudaria. *Archive for History of Exact Sciences*, 62 (5): 511–551.

Adams, C.C. (2004). *The knot book. An elementary introduction to the mathematical theory of knots.* American Mathematical Society, Providence, Rhode Island.

Ajdukiewicz, K. (1974). *Pragmatic logic.* D. Reidel Publishing Company, Dordrecht.

Alling, N.L. (1962). On the existence of real-closed fields that are η_α-sets of power $\aleph_\alpha$. *Transactions of the American Mathematical Society*, 103: 341–352.

Alling, N.L. (1987). *Foundations of analysis over surreal number fields.* Mathematics Studies 141, North-Holland Publishing Company.

Artin, E., Schreier, O. (1927). Eine Kennzeichnung der reell abgeschlossenen Körper. *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg*, 5: 225–231.

Awodey, S., Butz, S. (2000). Topological completeness for higher-order logic. *The Journal of Symbolic Logic*, 65 (3): 1168–1182.

Awodey, S., Carus, A.W. (2001). Carnap, completeness, and categoricity: the *Gabelbarkeitssatz* of 1928. *Erkenntniss*, 54: 145–172.

Awodey, S., Reck, E.H. (2002a). Completeness and categoricity. Part I: nineteenth-century axiomatics to twentieth-century metalogic. *History and Philosophy of Logic*, 23: 1–30.

Awodey, S., Reck, E.H. (2002b). Completeness and categoricity. Part II: twentieth-century metalogic to twenty-first-century semantics. *History and Philosophy of Logic*, 23: 77–94.

Baer, R. (1928). Über ein Vollständigkeitsaxiom in der Mengenlehre. *Mathematische Zeitschrift*, 27: 536–539.

Bagaria, J. (2005). Natural axioms of set theory and the continuum problem. In Hájek, P., Valdés-Villanueva, L., Westerståhl, D., editors, *Proceedings of the 12th International Congress of Logic, Methodology and Philosophy of Science*, 43–64. Kings College Publications, London.

Baldus, R. (1928). Zur Axiomatik der Geometrie I. Über Hilberts Vollständigkeitsaxiom. *Mathematische Annalen* 100: 321–333.

Baldwin, J.T. (2014). Completeness and categoricity (in power): foundations without foundationalism. *The Bulletin of Symbolic Logic* 20 (1): 39–79.

Barwise, J., editor, (1977). *Handbook of mathematical logic.* North-Holland Publishing Company, Amsterdam, New York, Oxford.

Barwise, J., Feferman, S., editors, (1985). *Model-theoretic logics.* Springer-Verlag, New York.

Barwise, J. (1985). Model-theoretic logics: background and aims. In Barwise, J., Feferman, S., editors, (1985), 3–23.

Batóg, T. (1996). *Dwa paradygmaty matematyki. Studium z dziejów i filozofii matematyki* [Two paradigms of mathematics: An essay on the history and philosophy of mathematics]. Wydawnictwo Naukowe UAM, Poznań.

Bedürftig, T., Murawski, R. (2018). *Philosophy of mathematics.* Walter de Gruyter, Boston, Berlin.

Bell, J.L. (2005). Continuity and infinitesimals. *Stanford Encyclopedia of Philosophy.* Available at:

https://plato.stanford.edu/entries/continuity/

Bell, J.L., Machover, M. (1977). *A course in mathematical logic.* North-Holland Publishing Company, Amsterdam.

Bell, J.L., Slomson, A.B. (1969). *Models and ultraproducts: an introduction.* North-Holland Publishing Company, Amsterdam, London.

Benacerraf, P., Putnam, H., editors, (1983). *Philosophy of mathematics. Selected readings.* Cambridge University Press, Cambridge.

Ben-Zeev, T., Star, J. (2001). Intuitive mathematics: theoretical and educational implications. In Torff, B., Sternberg, R.J., editors, (2001), 29–56.

Bernays, P. (1937–1948). A system of axiomatic set theory. *The Journal of Symbolic Logic*, 2: 65–77 (Part I, 1937), 6: 1–17 (Part II, 1941), 7: 65–89 (Part III, 1942), 7: 133–145 (Part IV, 1942), 8: 89–106 (Part V, 1943), 13: 65–79 (Part VI, 1948).

Birkhoff, G.D. (1932). A set of postulates for plane geometry (based on scale and protractors). *Annals of Mathematics*, 33: 329–345.

Błaszczyk, P. 2007. *Analiza filozoficzna rozprawy Richarda Dedekinda "Stetigkeit und irrationale Zahlen"* [Philosophical analysis of Richard Dedekind treatise "Stetigkeit und irrationale Zahlen"]. Wydawnictwo Naukowe Akademii Pedagogicznej, Kraków.

Borsuk, K., Szmielew, W. (1975). *Podstawy geometrii* [Foundations of geometry]. Państwowe Wydawnictwo Naukowe, Warszawa.

Bourbaki, N. (1980). *Elementy historii matematyki* [Elements of the history of mathematics]. Państwowe Wydawnictwo Naukowe, Warszawa.

Boyer, C. (1949). *The history of the calculus and its conceptual development.* Dover Publications, Inc., New York.

Buckley, B.L. (2012). *The continuity debate: Dedekind, Cantor, Du Bois-Reymond, and Pierce on continuity and infinitesimals.* Docent Press, Boston.

Buldt, B. (2002). Kompaktheit und Endlichkeit in der formalen Logik. In Buldt, B. u.a., editors, *Kurt Gödel. Wahrheit und Beweisbarkeit.* Band 1: *Dokumente und historische Analysen*, Band 2: *Kompendium zum Werk*, 31–49, öbv&hpt VerlagsgmbH & Co., Wien.

Button, T., Smith, P.R. (2012). The philosophical significance of Tennenbaum's theorem. *Philosophia Mathematica*, 20 (1): 114–121.

Button, T., Walsh, S. (2019). *Philosophy and model theory.* Oxford University Press, Oxford.

Cantor, G. (1872). Über der Ausdehnung eines Satzes aus der Theorie der trigonometrischen Reihen. *Mathematische Annalen*, 5: 123–132.

Carnap, R. (1927). Eigentliche und uneigentliche Begriffe. *Symposion*, 4: 355–374.

Carnap, R. (1929). *Abriss der Logistik. Mit besonderer Berücksichtigung der Relationstheorie und ihrer Anwendungen.* Springer-Verlag, Wien.

Carnap, R. (1930). Bericht über Untersuchungen zur allgemeinen Axiomatik. *Erkenntnis*, 1: 303–307.

Carnap, R. (1931). Versuch einer Metalogik. [22.1.31, im Bett geschrieben. In der schlaflosen Nacht vorher ausgedacht (Halsentzündung).] Carnap's *Nachlaß*, Archives for Scientific Philosophy, University of Pittsburgh.

Carnap, R. (1934). *Logische Syntax der Sprache.* Verlag von Julius Springer, Wien.

Carnap, R. (1937). *The logical syntax of language.* Kegan Paul, London.

Carnap, R. (1954). *Einführung in die symbolische Logik mit besonderer Berücksichtigung ihrer Anwendungen.* Springer-Verlag, Wien.

Carnap, R. (1958). *Introduction to symbolic logic and its applications.* Dover Publications, New York.

Carnap, R. (2000). *Untersuchungen zur Allgemeinen Axiomatik.* Bonk, T., Mosterin, J., editors, Wissenschaftliche Buchgesellschaft, Darmstadt.

Carnap, R., Bachmann, F. (1936). Über Extremalaxiome. *Erkenntnis*, 6: 166–188.

Carnap, R., Bachmann, F. (1981). On extremal axioms. [English translation of Carnap, Bachmann, 1936, by H.G. Bohnert] *History and Philosophy of Logic*, 2: 67–85.

Chang, C.C., Keisler, J.H. (1973). *Model theory.* North-Holland Publishing Company, Amsterdam, London; American Elsevier Publishing Company, Inc., New York.

Cohen, P. (1966). *Set theory and the continuum hypothesis.* New York.

Conway, J.H. (1976). *On numbers and games.* A K Peters, Natick, Massachusetts.

Conway, J.H., Guy, R.K. (1998). *The book of numbers.* Springer-Verlag, New York.

Corcoran, J. (1980). Categoricity. *History and Philosophy of Logic*, 1: 187–207.

Corcoran, J. (1981). From categoricity to completeness. *History and Philosophy of Logic*, 2: 113–119.

Cori, R., Lascar, D. (2001). *Mathematical logic. A course with exercises.* Oxford University Press, Oxford.

Couturat, A.L. (1904). IIme Congrès International de Philosophie, Génève. *Revue de métaphisique et de morale*, 12: 1037–1077.

Davis, P.J., Hersh, R. (1981). *Mathematical experience.* Birkhäuser, Boston.

Dawson, J.W. Jr. (1993). The compactness of first-order logic: from Gödel to Lindström. *History and Philosophy of Logic*, 14: 15–38.

Dedekind, R. (1872). *Stetigkeit und irrationale Zahlen.* Friedr. Vieweg und Sohn, Braunschweig.

Dedekind, R. (1888). *Was sind und was sollen die Zahlen?* Friedr. Vieweg und Sohn, Braunschweig.

De Morgan, A. (1915). *A budget of paradoxes.* Volume I. The Open Court Publishing Co., Chicago, London.

Descartes, R. (1954). *Philosophical writings.* A selection translated and edited by Elizabeth Anscombe and Peter Thomas Geach with an introduction by Alexandre Koyré. Thomas Nelson and Sons, Ltd., Edinburgh.

Dreben, B., Heijenoort, J. van (1986). Introductory note to 1929, 1930 and 1930a. In S. Feferman, S. *et al.*, editors, *Kurt Gödel: Collected works, Volume I.* Oxford University Press, New York, 44–59.

Du Bois-Reymond, P. (1866). *Über die Grundlagen der Erkenntniss in den exakten Wissenschaften.* Wissenschaftliche Buchgesellschaft, Darmstadt.

Du Bois-Reymond, P. (1877). Ueber die Paradoxen des Infinitärskalküls. *Mathematische Annalen*, 11: 149–167.

Du Bois-Reymond, P. (1882). *Allgemeine Functionentheorie.* H. Laupp, Tübingen.

Dudley, U. (1992). *Mathematical cranks.* The Mathematical Association of America, Washington, DC.

Ebbinghaus, H.D., Flum, J., Thomas, W. (1996). *Mathematical logic.* Springer.

Edwards, P., editor, (1967). *The encyclopedia of philosophy.* Macmillan Publishing Co., Inc. & The Free Press, New York, Collier Macmillan Publishers, London.

Ehrlich, P. (1988). An Alternative construction of Conway's ordered field *No*. *Algebra Universalis*, 25: 7–16.

Ehrlich, P. (1992). Universally extending arithmetic continua. In Sinaceur, H., Salanskis, J.M., editors, *Le labyrinthe du continu: Colloque du Cerissy*, 168–178, Springer-Verlag France, Paris.

Ehrlich, P. (1995). Hahn's *Über die nichtarchimedischen Grössensysteme* and the development of the modern theory of magnitudes and numbers to measure them. In Hintikka, J., editor, *From Dedekind to Gödel: essays on the development of the foundations of mathematics.* Kluwer Academic Publishers. Available at:

http://www.ohio.edu/people/ehrlich/HahnNew.pdf

Ehrlich, P. (2006). The rise of non-Archimedean mathematics and the roots of a misconception I: the emergence of non-Archimedean systems of magnitudes. *Archive for the History of Exact Sciences*, 60: 1–121.

Ehrlich, P. (2012). The absolute arithmetic continuum and the unification of all numbers great and small. *The Bulletin of Symbolic Logic*, 18 (1): 1–45.

Ewald, W. (1996). *From Kant to Hilbert. A source book in the foundation of mathematics.* Clarendon Press, Oxford.

Feferman, S. (2009). Conceptions of the continuum. *Intellectica*, 51: 169–189.

Feferman, S., Friedman, H.M., Maddy, P., Steel, J.R. (2000). Does mathematics need new axioms? *The Bulletin of Symbolic Logic*, 6 (4): 401–446.

Ferreirós, J. (1999). *Labyrinth of thought. A history of set theory and its role in modern mathematics.* Birkhäuser Verlag, Basel, Boston, Berlin.

Finsler, P. (1926). Über die Grundlegung der Mengenlehre I. *Mathematische Zeitschrift*, 25: 683–713.

Fischbein, E. (1987). *Intuition in science and mathematics: an educational approach.* Kluwer Academic Publishers, New York, Boston, Dordrecht, London, Moscow.

Fraenkel, A.A. (1922). Zu den Grundlagen der CANTOR-ZERMELOschen Mengenlehre. *Mathematische Annalen*, 86: 230–237.

Fraenkel, A.A. (1928). *Einleitung in die Mengenlehre.* Verlag von Julius Springer, Berlin.

Fraenkel, A.A., Bar-Hillel, Y. (1958). *Foundations of set theory.* North-Holland Publishing Company, Amsterdam, London.

Fraenkel, A., Bernays, P. (1958). *Axiomatic set theory.* North-Holland Publishing Company, Amsterdam.

Fraenkel, A.A., Bar-Hillel, Y., Levy, A. (1973). *Foundations of set theory.* North-Holland Publishing Company, Amsterdam, London.

Frege, G. (1879). *Begriffsschrift: eine der arithmetischen nachgebildete Formelsprache des reinen Denkens.* Halle a.S.

Frege, G. (1884). *Die Grundlagen der Arithmetik: eine logisch-mathematische Untersuchung über den Begriff der Zahl.* Breslau.

Frege, G. (1893–1903). *Grundgesetze der Arithmetik, begriffsschriftlich abgeleitet.* Verlag Hermann Pohle, Jena (Band I: 1893, Band II: 1903).

Friedman, H. (1992). The incompleteness phenomena. *Proceedings of the AMS Centennial Symposium, August 8–12, 1988.* American Mathematical Society, 49–84.

Gaifman, H. (2004). Nonstandard models in a broader perspective. In Enayat, A., Roman, R., editors, *Nonstandard models in arithmetic and set theory*, 1–22, AMS Special Session Nonstandard Models of Arithmetic and Set Theory, January 15–16, 2003, Baltimore, Maryland, *Contemporary Mathematics*, 361, American Mathematical Society, Providence, Rhode Island.

Gelbaum, B.R., Olmsted, J.M.H. (1990). *Theorems and counterexamples in mathematics.* Springer-Verlag, New York.

Gelbaum, B.R., Olmsted, J.M.H. (2003). *Counterexamples in analysis.* Dover Publications, Inc., Mineola, New York.

George, B. (2006). Second-order characterizable cardinals and ordinals. *Studia Logica*, 84 (3): 425–449.

Ghrist, R. (2014). *Elementary applied topology.* Createspace, ISBN 978-1502880857.

Gödel, K. (1930). Die Vollständigkeit der Axiome des logischen Funktionenkalküls. *Monatshefte für Mathematik und Physik*, 37: 349–360.

Gödel, K. (1931). Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme. *Monatshefte für Mathematik und Physik*, 38: 173–198.

Gödel, K. (1940). The consistency of the axiom of choice and of the generalized continuum hypothesis with the axioms of set theory. *Annals of Mathematics Studies* 3, Princeton.

Gödel, K. (1964). What is Cantor's continuum problem? In P. Benaceraff, P., Putnam, H., editors, *Philosophy of mathematics. Selected Readings*, 470–485, Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Gödel, K. (1986–2003). Feferman, S. *et al.*, editors, *Kurt Gödel: Collected works, Volume I 1986, Volume II 1990, Volume III 1995, Volume IV 2003, Volume V, 2003.* Oxford University Press, New York.

Grattan-Guiness, I. (2000). *The search for mathematical roots 1870–1940. Logics, set theories and the foundations of mathematics from Cantor through Russell to Gödel.* Princeton University Press, Princeton and Oxford.

Grassmann, H. (1861). *Lehrbuch der Arithmetik*. Verlag von Th. Chr. Fr. Enslin, Berlin.

Grzegorczyk, A. (1962). On the concept of categoricity. *Studia Logica*, 13: 39–66.

Hadamard, J. (1954). *An essay on the psychology of invention in the mathematical field*. Dover Publications, Inc., New York.

Hallet, M. (1984). *Cantorian set theory and limitation of size*. Clarendon Press, Oxford.

Hamilton, W.R. (1853). *Lectures on quaternions*. Hodges and Smith, Dublin.

Hanusek, J. (2015). Uwagi o arytmetyce Grassmanna [Remarks on Grassmann's arithmetic]. *Diametros*, 45: 107–121.

Hardy, G.H. (1954 (1910)). *Orders of infinity. The 'Infinitärcalcül' of Paul Du Bois-Reymond*. Cambridge University Press, Cambridge.

Hausdorff, F. (1906). Untersuchungen über Ordnungstypen I, II, III. *Ber. über die Verhandlungen der Königl. Sächs. Ges. der Wiss. zu Leipzig. Math.-phys. Klasse*, 58: 106–169.

Hausdorff, F. (1907). Untersuchungen über Ordnungstypen IV, V. *Ber. über die Verhandlungen der Königl. Sächs. Ges. der Wiss. zu Leipzig. Math.-phys. Klasse*, 59: 84–159.

Hausdorff, F. (1914). *Grundzüge der Mengenlehre*. Veit, Leipzig.

Havil, J. (2007). *Nonplussed! Mathematical proof of implausible ideas*. Princeton University Press, Princeton and Oxford.

Heath, T.L. (1968). *The thirteen books of Euclid's Elements. Translated from the text of Heiberg with introduction and commentary*. Cambridge University Press, Cambridge.

Heath, T.L. (Ed.) (2002). *The works of Archimedes*. Dover Publications, Inc., Mineola, New York.

Hedman, S. (2004). *A first course in logic. An introduction to model theory, proof theory, computability, and complexity*. Oxford University Press, Oxford.

Heijenoort, J. van, editor, (1967). *From Frege to Gödel: A source book in mathematical logic, 1879–1931.* Cambridge, Mass.

Heine, E. (1872). Die Elemente der Functionenlehre. *Journal für die reine und angewandte Mathematik*, 74: 172–188.

Henkin, L. (1961). Some remarks on infinitely long formulas. *Infinitistic Methods*, Pergamon Press, Qxford, 167–183.

Henkin, L. (1996). The discovery of my completeness proofs. *The Bulletin of Symbolic Logic*, 2 (2): 127–158.

Hilbert, D. (1899). *Grundlagen der Geometrie.* Festschrift zur Feier der Enthüllung des Gauss-Weber-Denkmals in Göttingen. Teubner, Leipzig.

Hilbert, D. (1900). Über den Zahlbegriff. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 8: 180–194.

Hilbert, D. (1950). *The foundations of geometry.* Authorized translation by E.J. Townsend. Reprint edition. The Open Court Publishing Company, La Salle, Illinois.

Hilbert, D., Ackermann, W. (1928). *Grundzüge der theoretischen Logik.* Verlag von Julius Springer, Berlin.

Hilbert, D., Bernays, P. (1934/1939). *Grundlagen der Mathematik.* Springer, Berlin.

Hinman, P.G. (2005). *Fundamentals of mathematical logic.* A K Peters, Wellesley, Massachusetts.

Hintikka, J. (1986). Extremality assumptions in the foundations of mathematics. *Philosophy of Science Association*, 2: 247–252.

Hintikka, J. (1991). Carnap, the universality of language and extremality axioms. *Erkenntnis*, 35 (1–3): 325–336.

Hintikka, J. (1992). Carnap's work in the foundations of logic and mathematics in a historical perspective. *Synthese*, 93: 167–189.

Hoborski, A. (1923). *Nowa teoria liczb niewymiernych* [A new theory of irrational numbers]. Kraków.

Hodges, W. (1993). *Model theory*. Cambridge University Press, Cambridge.

Hodges, W. (2000). Model theory (Draft 20 Jul 00). Available at:
http://wilfridhodges.co.uk/history07.pdf

Hodges, W. (2007). Tarski on Padoa's method. Available at:
www.maths.qmul.ac.uk/∼wilfrid/padoa.pdf

Hodges, W. (2019). A short history of model theory. Historical appendix D in: Button, Walsh (2019).

Howard, P.E., Yorke, M.F. (1989). Definitions of finite. *Fundamenta Mathematicae*, 133 (3): 169–177.

Huntington, E.V. (1902). A complete set of postulates for the theory of absolute continuous magnitude. *Transactions of the American Mathematical Society*, 3: 264–279.

Huntington, E.V. (1905). A set of postulates for real algebra, comprising postulates for a one-dimensional continuum and for the theory of groups. *Transactions of the American Mathematical Society*, 6: 17–41.

Huntington, E.V. (1913). A set of postulates for abstract geometry, expressed in terms of the simple relation of inclusion. *Mathematische Annalen*, 73: 522–559.

Jech, T. (2003). *Set theory. The third millenium edition, revised and expanded.* Springer-Verlag, Berlin, Heidelberg, New York, Hong Kong, London, Milan, Paris, Tokyo.

Jensen, R. (1972). The fine structure of the constructible hierarchy. *Annals of Mathematical Logic*, 4 (3): 229–308.

Kanamori, A. (1994). *The higher infinite. Large cardinals in set theory from their beginnings.* Springer-Verlag, Berlin.

Kanovei, V., Shelah, S. (2003). A definable nonstandard model of the reals. Available at:
https://arxiv.org/pdf/math/0311165.pdf

Kaye, R. (1991). *Models of Peano arithmetic.* Clarendon Press, Oxford.

Kharazishvili, A.B. (2006). *Strange functions in real analysis.* Chapman & Hall/CRC, Taylor & Francis Group, Boca Raton, London, New York, Singapore.

Kitcher, P. (1975). Kant and the foundations of mathematics. *The Philosophical Review*, 84 (1): 23–50.

Klein, F. (1872). Vergleichende Betrachtungen über neuere geometrische Forschungen. *Mathematische Annalen*, 43: 63–100.

Kline, M. (1972). *Mathematical thought from ancient to modern times.* Oxford University Press, New York, Oxford.

Klymchuk, S., Staples, S. (2013). *Paradoxes and sophisms in calculus.* Mathematical Association of America.

Knuth, D.E. (1974). *Surreal numbers: how two ex-students turned on to pure mathematics and found total happiness.* Addison-Wesley Publishing Company, Upper Saddle River, Boston, Indianapolis, San Francisco, New York, Toronto, Montréal, London, Munich, Paris, Madrid, Mexico City, Capetown, Sydney, Tokyo, Singapore.

Koellner, P. (2010). Independence and large cardinals. *Stanford Encyclopedia of Philosophy.* Available at:

http://plato.stanford.edu/entries/independence-large-cardinals/

Krygowska, Z. (1967). Rozumowanie empiryczne, intuicyjne i formalne w nauczaniu matematyki [Empirical, intuitive, and formal reasoning in teaching mathematics]. *Roczniki Polskiego Towarzystwa Matematycznego*, series II: *Wiadomości Matematyczne* 10: 49–91.

Krynicki, M., Mostowski, M., Szczerba, L.W., editors, (1995). *Quantifiers: logics, models and computation.* Kluwer Academic Publishers, Dordrecht, Boston, London.

Kuratowski, K. (1920). Sur la notion d'ensemble fini. *Fundamenta Mathematicae*, 1: 129–131.

Lakatos, I., editor, (1967). *Problems in the philosophy of mathematics, Proceedings of the international colloquium in the philosophy of science, London 1965.* North-Holland Publishing Company, Amsterdam.

Lakatos, I. (1976). *Proofs and refutations. The logic of mathematical discovery.* Cambridge.

Laraudogoitia, J.P. (1996). A beautiful supertask. *Mind*, 105: 81–83.

Lecat, M. (1935). *Erreurs de mathematiciens des origines à nos jours.* Castaigne, Brüssel.

Levi, M. (2009). *The mathematical mechanic. Using physical reasoning to solve problems.* Princeton University Press, Princeton and Oxford.

Levy, A. (1958). The independence of various definitions of finiteness. *Fundamenta Mathematicae*, 46: 1–13.

Lindenbaum, A., Mostowski, A. (1938). Über die Unabhängigkeit des Auswahlaxioms und einiger seiner Folgerungen. *C.R. des Séances de la Société des Sciences et des Lettres de Varsovie* Cl. III, 31: 27–32.

Lindenbaum, A., Tarski, A. (1936). Über die Beschränkheit der Ausdruckmittel deduktiver Theorien. *Ergebnisse eines mathematischen Kolloquiums* 1934–1935, 7: 15–22.

Lindström, P. (1966a). On characterizability in $L_{\omega_1\omega_0}$. *Theoria*, 32: 165–171.

Lindström, P. (1966b). On relations between structures. *Theoria*, 32: 172–185.

Lindström, P. (1966c). First order predicate logic with generalized quantifiers. *Theoria*, 32: 186–195.

Lindström, P. (1969). On extensions of elementary logic. *Theoria*, 35: 1–11.

Linnebo, Ø. (2017). *Philosophy of mathematics.* Princeton University Press, Princeton and Oxford.

Löwenheim, L. (1915). Über Möglichkeiten im Relativkalkül. *Mathematische Annalen*, 68: 169–207.

Mancosu, P. (2015). Explanation in mathematics. *The Stanford encyclopedia of philosophy.* Available at:

http://plato.stanford.edu/archives/sum2015/entries/mathematics-explanation/

Mancosu, P. (2010). *The adventure of reason. Interplay between philosophy and mathematical logic, 1900–1940*, Oxford University Press, Oxford.

Marcja, A., Toffalori, C. (2003). *A guide to classical and modern model theory.* Kluwer Academic Publishers, Dordrecht, Boston, London.

Marker, D. (2002). *Model theory: an introduction.* Springer-Verlag, New York, Berlin, Heidelberg.

Merzbach, U.C., Boyer, C.B. (2010). *A history of mathematics.* John Wiley & Sons, Inc., Hoboken, New Jersey.

Moore, G.H. (2013). *Zermelo's axiom of choice: its origins, development, and influence.* Dover Publications, Mineola, New York.

Mostowski, A. (1938). Über den Begriff einer endlichen Menge. *C.R. des Séances de la Société des Sciences et des Lettres de Varsovie* Cl. III, 31: 13–20.

Mostowski, A. (1955). Współczesny stan badań nad podstawami matematyki [Recent state of research in the foundations of mathematics]. *Prace matematyczne*, 1: 13–55.

Mostowski, A. (1957). On generalization of quantifiers. *Fundamenta Mathematicae*, 44: 12–36.

Mostowski, A. (1967). Recent results in set theory. In Lakatos, G., editor, (1967), 82–108.

Myhill, J. (1952). The hypothesis that all classes are nameable. *Proc. Nat. Acad. Sci. USA*, 38: 979.

Needham, T. (1997). *Visual complex analysis.* Clarendon Press, Oxford.

Neumann, J. von (1925). Eine Axiomatisierung der Mengenlehre. *Journal für die reine und angewandte Mathematik*, 154: 219–240.

Ojose, B. (2008). Applying Piaget's theory of cognitive development to mathematics instruction. *The Mathematics Educator*, 18 (1): 26–30.

Parsons, C. (2008). *Mathematical thought and its objects*. Cambridge University Press, Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo.

Pasch, M. (1926). *Vorlesungen über neuere Geometrie*. Springer, Berlin.

Peano, G. (1889). *Arithmetices principia, nova methodo exposita*. Bocca, Torino.

Piaget, J. (1970). *Science of education and the psychology of the child*. Viking, New York.

Piaget, J. (1977). *Epistemology and psychology of functions*. D. Reidel Publishing Company, Dordrecht.

Pogonowski, J. (2011). Remarks on intended models of mathematical theories. *Lingua Posnaniensis*, 53 (2): 83–100.

Pogonowski, J. (2016). Kontekst przekazu w matematyce [The context of transmission in mathematics]. *Annales Universitatis Paedagogicae Cracoviensis. Studia ad Didacticam Mathematicae Pertinentia*, 8: 119–137.

Pogonowski, J. (2017). On conceptual metaphors in mathematics. *Annales Universitatis Paedagogicae Cracoviensis. Studia ad Didacticam Mathematicae Pertinentia*, 9: 85–98.

Pogonowski, J. (2018a). Paradox resolution as a didactic tool. In Błaszczyk, P., Pieronkiewicz, B., editors, *Mathematical Transgressions 2015*, 324–339, Universitas, Kraków.

Pogonowski, J. (2018b). Intuitive explanations of mathematical ideas. *Annales Universitatis Paedagogicae Cracoviensis. Studia ad Didacticam Mathematicae Pertinentia*, 10: 123–137.

Poizat, B. (1999). *Course in model theory: an introduction in contemporary mathematical logic*. Springer-Verlag.

Polya, G. (1957). *How to solve it*. Doubleday, Garden City, NY.

Polya, G. (2009). *Mathematical discovery. On understanding, learning, and teaching problem solving*. Ishi Press International, New York, Tokyo.

Polya, G. (2014). *Mathematics and plausible reasoning.* Vol.I: *Induction and analogy in mathematics*, Vol. II: *Patterns of plausible inference.* Martino Publishing, Mansfield Centre, CT.

Posamentier, A.S., Lehmann, I. (2013). *Magnificent mistakes in mathematics.* Prometheus Books, Amherst (New York).

Prasolov, V.V. (2011). *Intuitive topology.* American Mathematical Society.

Pust, J. (2012). Intuition. *The Stanford encyclopedia of philosophy.* Available at:

https://plato.stanford.edu/entries/intuition/

Putnam, H. (1980). Models and reality. *The Journal of Symbolic Logic*, 45: 464–482.

Quine, W.V.O. (1941). Element and number. *Journal of Symbolic Logic*, 6 (4): 135–149.

Quine, W.V.O. (1946). Concatenation as a basis for arithmetic. *Journal of Symbolic Logic*, 11 (4): 105–114.

Quinon, P., Zdanowski, K. (2006). The intended model of arithmetic. An argument from Tennenbaum's theorem. Available at:

https://pdfs.semanticscholar.org/c92a/4bc83cf67d5f0f4d054301d5b7ebb2ff8ed9.pdf

Robinson, A. (1996). *Non-standard analysis.* Princeton Landmarks in Mathematics (2nd ed.), Princeton University Press.

Romero, G.E. (2014). The collapse of supertasks. *Foundation of science*, 19 (2): 209–216.

Scanlan, M. (1991). Who were the American Postulate Theorists? *The Journal of Symbolic Logic*, 56 (3): 981–1002.

Scanlan, M. (2003). American Postulate Theorists and Alfred Tarski. *History and Philosophy of Logic*, 24: 307–325.

Scheepers, M. (1993). Gaps in ω^ω. In *Set theory of the reals (Ramat Gan, 1991).* Volume 6 of *Israel Math. Conf. Proc.*, Bar-Ilan University, Ramat Gan, 439–561.

Scheepers, M. (1994). Gaps in ${}^{\omega}\omega$. Update: January 29, 1994. Department of Mathematics, Boise State University, Boise, Idaho 83725. Available at:

https://math.boisestate.edu/∼marion/research/order/gaps.pdf

Schiemer, G. (2010a). Fraenkel's axiom of restriction: axiom choice, intended models, and categoricity. In Löwe B., T. Müller, T., editors, *Philosophy of mathematics: sociological aspects and mathematical practice*, 307–340, College Publications, London.

Schiemer, G. (2010b). *Carnap's early semantics.* PhD Dissertation, Universität Wien.

Schiemer, G. (2012). Carnap on extremal axioms, "completeness of models", and categoricity. *The Review of Symbolic Logic*, 5 (4): 613–641.

Schiemer, G. (2013). Carnap's early semantics. *Erkenntniss*, 78 (3): 487–522.

Schiemer, G., Reck, E.H. (2013). Logic in the 1930s: type theory and model theory. *The Bulletin of Symbolic Logic*, 19 (4): 433–472.

Schiemer, G., Zach, R., Reck, E.H. (2015). Carnap's early metatheory: scope and limits. Available at: arXiv:1508.05867v1.pdf

Schoenfeld, A. H. (1985). *Mathematical problem solving.* Academic Press, Inc., Orlando.

Schwabhäuser, W., Szmielew, W., Tarski, A. (1983). *Metamathematische Methoden in der Geometrie.* Springer-Verlag.

Shapiro, S., editor, (1996). *The limits of logic: higher-order logic and the Löwenheim-Skolem theorem.* Dartmouth Publishing Company, Aldershot.

Shapiro, S. (Ed.) (2005). *Philosophy of mathematics and logic.* Oxford University Press, Oxford.

Shepherdson, J.C. (1951–1953). Inner models for set theory. *Journal of Symbolic Logic*, 16: 161–190 (1951), 17: 225–237 (1952), 18: 145–167 (1953).

Sierpińska, A. (1994). *Understanding in mathematics.* The Falmer Press, London.

Skolem, T. (1920). Logisch-kombinatorische Untersuchungen über die Erfüllbarkeit oder Beweisbarkeit mathematischer Sätze nebst einem Theoreme über dichte Mengen. *Videnskappselskapets skrifter, I. Matematisk-naturvedenskabelig klasse*, 4. Translated in Heijenoort, J. van, editor, (1967), 252–263.

Skolem, T. (1922). Einige Bemerkungen zur axiomatischen Begründung der Mengenlehre. *Matematikerkongressen i Helsingfors den 4–7 Juli 1922, Den femte skandinaviska matematikerkongressen, Redogörelse*, (Akademiska Bokhandeln, Helsinki, 1923), 217–232.

Skolem, T. (1933). Über die Unmöglichkeit einer vollständigen Charakterisierung der Zahlenreihe mittels eines endlichen Axiomensystems. *Norsk matematisk forenings skrifter*, 2 (10): 73–82.

Skolem, T. (1934). Über die Nicht-charakterisierbarkeit der Zahlenreihe mittels endlich oder abzählbar unendlich vieler Aussagen mit ausschließlich Zahlenvariablen. *Fundamenta Mathematicae*, 23 (1): 150–161.

Smith, J.T. (2010). Definitions and nondefinability in geometry. *American Mathematical Monthly*, 117: 475–489.

Stammler, G. (1928). *Begriff Urteil Schluss. Untersuchungen über Grundlagen und Aufbau der Logik.* Niemeyer, Halle/Saale.

Stäckel, P. (1907). Zu H. Webers elementarer Mengenlehre. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 16: 425–428.

Steen, L.A., Seebach, J.A., Jr. (1995). *Counterexamples in topology.* Dover Publications, Inc., New York.

Stegmüller, W., Varga von Kibéd, M. (1984). *Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie, Band III: Strukturtypen der Logik. Teil C.* Springer-Verlag, Berlin, Heidelberg, New York, Tokyo.

Stillwell, J. (2018). *Reverse mathematics. Proofs from the inside out.* Princeton University Press, Princeton and Oxford.

Surma, S., editor, (1973a). *Studies in the history of mathematical logic.* Zakład Narodowy imienia Ossolińskich, Wydawnictwo Polskiej Akademii Nauk, Wrocław, Warszawa, Kraków, Gdańsk.

Surma, S. (1973b). A historical survey of the significant methods of proving Post's theorem about the completeness of the classical propositional calculus. In Surma, S., editor, (1973a), 19–32.

Surma, S. (1973c). A survey of various concepts of completeness of the deductive theories. In Surma, S., editor, (1973a), 279–288.

Suszko, R. (1949). *O analitycznych aksjomatach i logicznych regułach wnioskowania. Z teorii definicji* [On analytical axioms and logical rules of inference. From the theory of definitions]. Poznańskie Towarzystwo Przyjaciół Nauk, Prace Komisji Filozoficznej, 7 (5), Poznań.

Suszko, R. (1951). Canonic axiomatic systems. *Studia Philosophica*, 4: 301–330.

Tall, D. (2013). *How humans learn to think mathematically. Exploring the three worlds of mathematics*. Cambridge University Press, Cambridge.

Tao, T. (2009). There's more to mathematics than rigour and proofs. Available at:

https://terrytao.wordpress.com/career-advice/theres-more-to-mathematics-than-rigour-and-proofs/

Tarski, A. (1924). Sur les ensembles finis. *Fundamenta Mathematicae*, 6: 45–95.

Tarski, A. (1929). Les fondements de la gémetrié des corps. In *Księga Pamiątkowa Pierwszego Polskiego Zjazdu Matematycznego*, 29–33, Polskie Towarzystwo Matematyczne, Kraków.

Tarski, A. (1933). Einige Betrachtungen über die Begriffe der ω-Widerspruchsfreiheit und ω-Vollständigkeit. *Monatshefte für Mathematik und Physik*, 40: 97–112.

Tarski, A. (1934). Z badań metodologicznych nad definiowalnością terminów [Methodological investigations into definability of terms]. *Przegląd Filozoficzny*, 37: 438–460.

Tarski, A. (1940). On the completeness and categoricity of deductive theories. Appendix in Mancosu, P. (2010), 485–492.

Tarski, A. (1951). *A decision method for elementary algebra and geometry.* University of California Press, Berkeley and Los Angeles.

Tarski, A. (1954). Contributions to the theory of models I. *Indagationes Mathematicae*, 16: 572–581.

Tarski, A. (1956). A general theorem concerning primitive notions of Euclidean geometry. *Indagationes Mathematicae*, 18: 468-474. Reprinted in Tarski 1986b, vol. 3, 611–619.

Tarski, A. (1959). What is elementary geometry? In Henkin, L., Suppes, P., Tarski, A., editors, *The axiomatic method. With special reference to geometry and physics. Proceedings of an International Symposium held at the Univ. of Calif., Berkeley, Dec. 26, 1957-Jan. 4, 1958*, 16–29, Studies in Logic and the Foundations of Mathematics, North-Holland Publishing Company, Amsterdam.

Tarski, A. (1986a). What are logical notions? *History and Philosophy of Logic*, 7: 143–154.

Tarski, A. (1986b). *Collected papers.* Edited by S.R. Givant and R.N. McKenzie, 4 volumes, Birkhäuser, Basel.

Tarski, A. (1994). *Introduction to logic and to the methodology of deductive sciences.* Oxford University Press.

Tennant, N. (2000). Deductive versus expressive power: a pre-Gödelian predicament. *Journal of Philosophy*, 97: 257–277.

Thomson, W. (1842). *Outline of the laws of thought.* Pickering, London; Graham, Oxford.

Thurston, W. (1994). On proof and progress in mathematics. *Bulletin of the American Mathematical Society*, 30 (2): 161–177.

Tieszen, R.L. (1989). *Mathematical intuition: phenomenology and mathematical knowledge.* Kluwer Academic Publishers, Dordrecht.

Torff, B., Sternberg, R.J., editors, (2001). *Understanding and teaching the intuitive mind: student and teacher learning.* Lawrence Erlbaum Associates Publishers, Mahwah, New Jersey, London.

Vaught, R. (1974). Model theory before 1945. In Henkin, L., Addison, J., Chang, C.C., Craig, W., Scott, D., Vaught, R., editors, *Proceedings of the Tarski Symposium*, 153–172, Proceedings of Symposia in Pure Mathematics, 25, American Mathematical Society, Rhode Island, New York.

Veblen, O. (1904). A system of axioms for geometry. *Transactions of the American Mathematical Society*, 5: 343–384.

Veblen, O. (1906). The foundations of geometry. *Popular Science Monthly*, 68: 21–28.

Wang, H. (1955). On denumerable bases of formal systems. In Skolem, T., Hasenjaeger, G., Kreisel, G., Robinson, A., Wang, H., Henkin, L., Łoś, J., editors, *Mathematical interpretation of formal systems*, 57–84, North-Holland Publishing Company, Amsterdam.

Wang, H. (1957). The axiomatization of arithmetic. *The Journal of Symbolic Logic*, 22: 145–158.

Wang, H. (1974). *From mathematics to philosophy*. Routledge and Kegan Paul.

Wang, H. (1996). *A logical journey: from Gödel to philosophy*. MIT Press, Cambridge, Mass.

Weaver, G., George, B. (2002). Quasi-finitely characterizable and finitely characterizable Dedekind algebras. *Bulletin of the Section of Logic*, 31: 145–157.

Weaver, G., George, B. (2003). The Fraenkel-Carnap question for Dedekind algebras. *Mathematical Logic Quarterly*, 49: 92–96.

Weaver, G., George, B. (2005). Fraenkel-Carnap properties. *Mathematical Logic Quarterly*, 51: 285–290.

Weber, H. (1906). Elementare Mengenlehre. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 15: 173–184.

Westerståhl, D. (1989). Quantifiers in formal and natural languages. In Gabbay, D.M., Guenthner, F., editors, *Handbook of philosophical logic.* Volume IV *Topics in the philosophy of language*, 1–131, D. Reidel Publishing Company.

Wilkosz, W. (1932). *Arytmetyka liczb całkowitych. System aksjomatyczny* [Integer arithmetic. Axiomatic system]. Bibljoteczka Kółka Mat.-Fiz. U.U.J. Nr 1, Kraków.

Wise, G.L., Hall, E.B. (1993). *Counterexamples in probability and real analysis*. Oxford University Press, New York.

Woleński, J. (2005). *Epistemologia. Poznanie, prawda, wiedza, realizm* [Epistemology. Cognition, truth, knowledge, realism]. Wydawnictwo Naukowe PWN, Warszawa.

Wright, C. (1983). *Frege's conception of numbers as objects*. Aberdeen University Press, Aberdeen.

Zach, R. (1999). Completeness before Post: Bernays, Hilbert, and the development of propositional logic. *The Bulletin of Symbolic Logic*, 5 (3): 331–366.

Zermelo, E. (1908). Untersuchungen über die Grundlagen der Mengenlehre I. *Mathematische Annalen*, 65: 261–281.

Zermelo, E. (1909a). Sur les ensembles finis et le principe de l'induction complète. *Acta mathematica*, 32: 185–193.

Zermelo, E. (1909b). Über die Grundlagen der Arithmetik. In: Castelnuovo, G., editor, *Atti del IV Congresso Internazionale dei Mathematici (Roma, 6–11 Aprile 1908)*, vol. 2: *Communicazioni delle sezioni I e II*, 8–11, Tipografia della R. Academia dei Lincei, Rome.

Zermelo, E. (1930). Über Grenzzahlen und Mengenbereiche: Neue Untersuchungen über die Grundlagen der Mengenlehre. *Fundamenta Mathematicae*, 16: 29–47.

Zermelo, E. (2010). *Ernst Zermelo. Collected works. Gesammelte Werke.* Edited by Ebbinghaus, H.D., Fraser, C.G., Kanamori, A., Springer, Heidelberg, Dordrecht, London, New York.

Zygmunt, J. 1973. On the sources of the notion of the reduced product. *Reports on Mathematical Logic*, 1: 53–67.

Author index

Subject index