Springer Monographs in Mathematics

Pavel Pudlák

Logical Foundations of Mathematics and Computational Complexity

A Gentle Introduction

Springer

S*pringer* **M***onographs* in **M***athematics*

For further volumes:
www.springer.com/series/3733

Pavel Pudlák

Logical Foundations of Mathematics and Computational Complexity

A Gentle Introduction

Springer

Pavel Pudlák
ASCR
Prague, Czech Republic

ISSN 1439-7382 Springer Monographs in Mathematics
ISBN 978-3-319-00118-0 ISBN 978-3-319-00119-7 (eBook)
DOI 10.1007/978-3-319-00119-7
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013936799

Mathematics Subject Classification: 03D15, 03E30, 03E35, 03F03, 03F20, 03F30, 03F40, 68Q15

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Dedicated to my parents
Anna Pudláková and Ján Pudlák

Preface

As the title states, this book is about logic, foundations and complexity. My aim is to present these topics in a readable form, accessible to a wide spectrum of readers. The message that I want to convey is that complexity, either in the form of computational complexity or in the form of proof complexity, is as important for foundations as the more traditional concepts of computability and provability are. Rather than presenting my own philosophical doctrine in the foundations, my goal is to isolate the most important problems and invite the reader to think about them.

The foundations of mathematics has always attracted mathematicians and philosophers. There were periods of time when many mathematicians were involved in the discussion of foundations. The most important such period was at the beginning of the 20th century. At that time the set-theoretical foundations were laid down, but set theory itself ran into problems—paradoxes were found showing that the intuitive use of set theory sometimes leads to contradictions. This problem was solved by accepting a particular axiomatic system for set theory, and things settled down. Later the interest in the foundations was stirred by several events. In the 1930s, it was Gödel's Incompleteness Theorem that showed that Hilbert's program to prove the consistency of the foundations was not possible. The second major event was Cohen's proof of the independence of the Continuum Hypothesis in the 1960s. This was an open problem concerning a basic question about the cardinality of the real numbers, posed by Cantor already in the 1870s. Also in the late 1960s a new field emerged that seemed to be somehow connected with foundations. This was the computational complexity theory.

Achievements in foundations can be viewed as solutions of important problems, but in fact they present us with much deeper open problems. Do the axioms of set theory describe the real universe of sets? Can we trust the axiomatic system for set theory to be free of contradiction? When the consistency of a theory is only provable in a stronger theory, according to the Incompleteness Theorem, what are we going to do with the consistency problem? How are we going to decide the Continuum Hypothesis, when it is independent of the axioms of set theory? In computational complexity there are a number of open problems. They may just be very difficult solvable problems, but their nature, which is similar to logical problems, and their resilience with which they resist any attempts to solve them, rather suggest that there are more fundamental reasons why they are still open.

These examples show that, in spite of all the progress that has been achieved, there are problems in the foundations that are still widely open. Many mathematicians and philosophers are aware of this fact and are thinking about the problems. But not only them; also physicists have realized that they must know something about the foundations of mathematics if they want to find the unified foundations of physics. One can observe a renewed interest in the foundations in the past decade notwithstanding the fact that there has been no breakthrough result obtained recently.

However, a mathematician with a deeper interest in this subject does not have much choice of suitable sources: on the one hand, there are many popular books that present the subject in a very superficial manner, and often incorrectly; on the other hand, there are monographs about various parts of logic, set theory and computational complexity theory that can only be read with considerable effort. Furthermore, these monographs always cover much more than is needed for understanding the basic questions about the foundations, and someone not acquainted with the field does not know what to read and what to skip.

This book is intended to fill this gap by presenting a survey of results related to the foundations of mathematics and complexity theory in a readable form and with a sufficient amount of detail. It focuses on explaining the essence of concepts and the ideas of proofs, rather than presenting precise formal statements and full proofs. Each section starts with concepts and results that can easily be explained, and gradually proceeds to more difficult ones. The idea is that the readers should not be lost before they get to the heart of the matter. But since mathematicians are always curious how the things are actually done, some formal definitions and sketches of proofs are provided in the notes to the sections.

The prospective readers of this book are mathematicians with an interest in the foundations, philosophers with a good background in mathematics and, perhaps, also philosophically minded physicists. Most of the book should be accessible to graduate students of mathematics. Logicians may find much of the material familiar, but they can profit from the chapters about computational and proof complexities, unless they also work in these fields.

I should also say what the reader should not expect from the book. Although the style of the presentation is often light (such as in the quotations from science fiction stories), the book is not popular science—its primary aim is not to entertain, but to educate the reader. So the readers will need to stop from time to time and ponder what they have read, or even to skip a part and return to it later. But the book is also not a typical dry monograph consisting of definitions, theorems and proofs. Concerning the history of mathematics, the facts that I occasionally mention are only meant to make the text more readable and are not intended to give a complete picture of the development of the field.

The book consists of seven chapters. The first two chapters are an introduction to the foundations of mathematics and mathematical logic. The material is explained very informally and more detailed presentation is deferred to later chapters. For example, set theory is introduced by means of several informal principles that are presented more precisely as the axioms of Zermelo-Fraenkel Set Theory in Chap. 3. Similarly, the Incompleteness Theorem is only stated and the proof and the consequences are discussed in Chap. 4.

Chapter 3 is devoted to set theory, which is the most important part of the foundations of mathematics. The two main themes in this chapter are: (1) higher infinities as a source of powerful axioms, and (2) alternative axioms, such as the Axiom of Determinacy.

Proofs of impossibility, the topic of Chap. 4, are proofs that certain tasks are impossible, contrary to the original intuition. Nowadays we tend to equate impossibility with unprovability and non-computability, which is a rather narrow view. Therefore, it is worth recalling that the first important impossibility results were obtained in different contexts: geometry and algebra. The most important result presented in this chapter is the Incompleteness Theorem of Kurt Gödel. I believe that the essence of the proof of this theorem can be explained with very little formalism and this is what try to I do in this chapter. Due to the diversity of results and connections with concrete mathematics, this is probably the most interesting chapter.

Proofs of impossibility are, clearly, important in foundations. One field in which the most basic problems are about proving impossibility is computational complexity theory, the topic of Chap. 5. But there are more connections between computational complexity and the foundations. I think that one cannot study the foundations of mathematics without understanding computational complexity.

In fact, there is a field of research that studies connections between computational complexity and logic. It is called '*Proof Complexity*' and it is presented in Chap. 6. Although we do have indications that complexity should play a relevant role in the foundations, we do not have any results proving this connection. In the last section of this chapter I present some ideas of mine about the possible nature of these connections. I state several conjectures which, if true, would give an explicit link between these two areas.

Every book about the foundations of mathematics should mention the basic philosophical approaches to the foundations of mathematics. I also do it in Chap. 7, but as I am not a philosopher, the main part of the chapter rather concentrates on mathematical results and problems that are at the border of mathematics and philosophy. Since I feel that the field lacks innovative approaches, I present one at the end of the chapter. It is based on the idea that natural numbers that can be represented in the physical universe are different from those studied in mathematics.

I tried to be as neutral as possible, but one cannot avoid using a certain philosophical standpoint when explaining the foundations. At the beginning of the book I assume the point of view of a realist, because it is easier to explain logic to a beginner from this viewpoint. My actual philosophy is the one of a moderate formalist, which is certainly apparent from my comments throughout the book. The only special feature of my philosophy is the stress on the importance of the complexity issues.

Even a thick volume like this cannot cover everything that is relevant to the foundations of mathematics. The main omission that I am aware of concerns intuitionistic type theories. These theories play a central role in the current research into the intuitionistic foundations of mathematics. The reasons for this omission is my lack of expertise in this field and the fact that the book is already fairly long as it is.

Prague, Czech Republic
January 2013

Pavel Pudlák

Acknowledgements

I would like to thank all who helped me by reading parts of the manuscript, pointing out errors, suggesting improvements or answering questions related to the text: Paul Beame, Arnold Beckmann, Lev Beklemishev, Samuel Buss, Lorenzo Carlucci, Stephano Cavagnetto, Dmitri Gavinsky, Stefan Hetzl, Edward Hirsch, Radek Honzík, Pavel Hrubeš, Peter Koellner, Leszek Kolodziejczyk, Jan Krajíček, Sebastian Müller, Jan Nekovář, Adam Nohejl, Jeffrey Paris, Ján Pich, Michael Rathjen, Zenon Sadowski, Neil Thapen, Iddo Tzameret, Eva Vachková.

I am also grateful to the anonymous reviewers of the manuscript, whose critical remarks were very useful and helped me to correct several errors.

My thanks are further due to Julie Cismosky, Sean Miller and Neil Thapen for correcting the English, and to Petr Pudlák for helping me with computer related issues.

The photographs were kindly provided by: Fachbereich Mathematik, Universität Hamburg (Cantor, Dedekind); Kurt Gödel Society, Vienna (Gödel); Archives of the Mathematisches Forschungsinstitut Oberwolfach (D. Hilbert); King's College, Cambridge University (Turing); Princeton University Library (A. Church); Bertrand Russell Archives, McMaster University Library (Russell); Universitätsarchiv Zürich (Zermelo).

I appreciate the help of Lynn Brandon, Lauren Stoney and Catherine Waite from Springer-Verlag London during the preparation of the manuscript for publication.

Through all the work I was supported by the Institute of Mathematics of the Academy of Sciences of the Czech Republic and received additional support from several grants of the Grant Agency of the Academy of Sciences.

Last, but not least, I want to thank my wife Věra for her understanding and encouragement over the years of writing this book.

Contents

1 Mathematician's World . . . 1

1.1 Mathematical Structures . . . 2

1.2 Everything Is a Set . . . 25

1.3 Antinomies of Set Theory . . . 36

1.4 The Axiomatic Method . . . 43

1.5 The Necessity of Using Abstract Concepts . . . 54

Main Points of the Chapter . . . 64

2 Language, Logic and Computations . . . 65

2.1 The Language of Mathematics . . . 66

2.2 Truth and Models . . . 80

2.3 Proofs . . . 92

2.4 Programs and Computations . . . 123

2.5 The Lambda Calculus . . . 146

Main Points of the Chapter . . . 155

3 Set Theory . . . 157

3.1 The Axioms of Set Theory . . . 159

3.2 The Arithmetic of Infinity . . . 176

3.3 What Is the Largest Number? . . . 196

3.4 Controversial Axioms . . . 215

3.5 Alternative Set-Theoretical Foundations . . . 231

Main Points of the Chapter . . . 253

4 Proofs of Impossibility . . . 255

4.1 Impossibility Proofs in Geometry and Algebra . . . 256

4.2 The Incompleteness Theorems . . . 272

4.3 Algorithmically Unsolvable Problems . . . 300

4.4 Concrete Independence . . . 319

4.5 The Independent Sentences of Set Theory . . . 340

Main Points of the Chapter . . . 364

5 The Complexity of Computations 365
5.1 What Is Complexity? 366
5.2 Randomness, Interaction and Cryptography 410
5.3 Parallel Computations 437
5.4 Quantum Computations 448
5.5 Descriptional Complexity 479
Main Points of the Chapter 493

6 Proof Complexity 495
6.1 Proof Theory 496
6.2 Theories and Complexity Classes 523
6.3 Propositional Proofs 540
6.4 Feasible Incompleteness 562
Main Points of the Chapter 580

7 Consistency, Truth and Existence 583
7.1 Consistency and Existence 584
7.2 The Attributes of Reality 609
7.3 Finitism and Physical Reality 646
Main Points of the Chapter 664

Bibliographical Remarks 667

References 671

Name Index 683

Subject Index 687

Symbols and Abbreviations 695

Chapter 1
Mathematician's World

The real universe arched sickeningly away beneath them. Various pretend ones flitted silently by, like mountain goats. Primal light exploded, splattering space-time as with gobbets of junket. Time blossomed, matter shrank away. The highest prime number coalesced quietly in a corner and hid itself away for ever.

Douglas Adams*, The Hitchhiker's Guide to the Galaxy*

FOR an ordinary person, it is a strange, imaginary world. At the entrance we meet very familiar creatures, such as the natural numbers $0, 1, 2, \dots$, but further on there will appear many strange aliens, like the imaginary unit i, the first uncountable cardinal number $\aleph_1$ and things even stranger than these. In some sense it is like the artificial worlds of science fiction, or like a detective story made up of mysteries with logical solutions, but still in many respects it is very different. The main difference is, perhaps, not in the artificial nature of the things that we encounter in mathematics, which apparently have very little to do with our everyday life, but in the strict rules that they obey. In a good detective story the detective eventually solves a mysterious crime by applying logical deduction. The author usually pretends that you could also have deduced who the murderer was already at the beginning of the story, knowing only the basic data presented on the first few pages. But in fact this is not true; on the contrary, the author chooses the most unlikely person. In mathematics you really can solve problems using only deduction and, in fact, no initial data are needed, except for the statement of the problem; the only things you need are patience and determination.

If you read a good novel or regularly watch a TV series, you enter into the world of the heroes of the story and often forget, at least for a while, that it is not real. In science fiction stories you can even experience a completely different world than ours here on Earth. Science fiction gives writers the opportunity to construct new worlds, even worlds that are in contradiction with firmly established laws of physics. There is nothing wrong with this if it has its own logic. Similarly, mathematicians invent worlds which are sometimes completely alien to ordinary people. In their minds they create mental pictures of the concepts about which they are thinking, as if they could really see numbers, sets, functions, infinitely dimensional spaces and a lot more, and move in this environment arranging these objects until they construct

P. Pudlák, *Logical Foundations of Mathematics and Computational Complexity*,
Springer Monographs in Mathematics, DOI 10.1007/978-3-319-00119-7_1,

the one they were looking for. Active mathematicians actually spend a big portion of their lives in this world. The more time they spend there, the more real this world seems to them. Like many teenagers who spend a lot of time in the virtual realities of computer games, mathematicians live part of their lives in what I would call *real virtuality*. Whereas virtual reality is pretend reality, what mathematicians do is the opposite: their worlds seem virtual, but are in some sense very real.

So is the mathematical world real or not? Most mathematicians would defend the true existence of at least some mathematical objects; in fact, most people would agree that the numbers $0, 1, 2, \ldots$ in some abstract sense do exist. As I will explain later, this is not just an important philosophical question, it is a question which is very important for the foundations of mathematics independent of our philosophical view, or our lack of interest in philosophy. But before we discuss such problems we have to know what kind of "things" mathematicians deal with.

1.1 Mathematical Structures

In biology we study animals, plants, bacteria, etc., in astronomy stars, planets, etc. So we can define biology as the science studying living organisms, astronomy as the science of the universe, and so on. But how can we describe mathematics? The answer to this question used to depend on what the main topic in contemporary mathematics was. For ancient Greeks, mathematics was essentially geometry and thus mathematics was the science of space. In the 18th century, when mathematics was tightly connected with physics, an answer to the question '*What is the subject of study of mathematics?*' would most likely be that it is *quantities and the relations between them.* A '*quantity*' was a real number that possibly depended on other numbers. For example, when describing a motion of a physical object, quantities could be position, speed, and momentum, all depending on time. The views on what the subject of mathematics is changed gradually. In roughly the 19th century mathematicians realized that there could be other objects of study on top of the traditional ones. The discovery of non-Euclidean geometries was an important step towards realizing that one does not have to study only objects which occur naturally in real life. An especially dramatic shift happened in algebra, where mathematicians realized that the usual number-theoretic structures are merely special instances from classes of structures sharing properties with the standard ones. Later on, new mathematical fields appeared where the objects studied had little to do with numbers or geometry. A systematic treatment of all mathematical objects became possible only after calculus had been given rigorous foundations and when there was a sufficiently general tool at hand: the concept of set.

I will describe the current standard approach to the question of what mathematical objects are. It is based on the concept of a *mathematical structure*, which gradually developed in the first half of the 20th century and was finally adopted as a key concept by the Bourbakists. *Nicolas Bourbaki* was a pseudonym under which, in 1939, a group of young French mathematicians started publishing an encyclopedic series of monographs covering the main fields in mathematics. Naturally, an attempt to give a unified treatment to the whole of mathematics needed a general concept such as the concept of mathematical structure.

This is certainly not the only possible view of contemporary mathematics. If I were not interested in foundations and wanted rather to explain the source of ideas which led to the most profound results, I would choose a different vantage point. Quite often it is difficult to formalize general ideas by a single mathematical concept. In fact, the main progress in modern mathematics has in most cases been achieved by realizing that the same idea was present in several fields and thus results and proof techniques could be transferred from one field to another. A prominent example is algebraic geometry, a field which applies geometric ideas to various non-geometric objects, including some discrete structures. Mathematics has always been a never ending struggle to express general ideas in a comprehensible, general and rigorous way and thus it cannot be explained completely by a single concept such as the mathematical structure. Nevertheless, Bourbaki's structuralist approach is the best that we have.

The ancient mathematicians considered only a few structures: the natural numbers, the plane and three dimensional space. Gradually new structures appeared in mathematics, although it was not an easy process to accept them. For instance, the complex numbers turned out to be very useful, but for a long time they were treated as a strange auxiliary means to solve problems about real numbers. We still use the terminology of *real* and *imaginary* numbers, but now we treat these words as purely technical terms and do not attribute more existence to real numbers than to complex ones. In mathematical analysis people realized that functions can be added, multiplied, etc. just as numbers can be, though they are not numbers. An important turning point was when mathematicians realized that they did not have to study only one of the few standard structures, instead they could choose any structure from a large variety. It was as if the objects of study were not given to them, but they could design them according to their own will and need, just following certain rules. (Whether one views it as the possibility to choose, or the possibility to create, depends on one's philosophical standpoint.)

Let us turn to the definition of a structure. Roughly speaking, a mathematical structure is a toy or a gadget that you can play with. You push or turn knobs and something happens. It is also like a painting where a single brush stroke makes no sense, but together the strokes give some meaning. You can also think of a structure as a game. In a game you have certain *objects*, and *rules* that determine what you can do with the objects.

A nice example is Rubik's cube, the well-known toy: the objects are the 26 small cubes and the rules are fixed by the ingenious mechanism of Rubik's cube that allows you to move only certain groups of small cubes together, namely those that form a face of the cube. Though it was important to design the mechanical construction of the cube, so that it worked well and could be mass produced, the essence of it is not the mechanism. The only thing that is important is that you have 26 pieces and particular rules how to move them. You can do "mathematical research" on Rubik's cube by studying what configurations are possible, which are symmetric, how many steps you need to transform a particular configuration into another one and so on. This is, in fact, what mathematicians actually do with structures.

There are many different structures; some are, in some sense, unique, while some are just members of large classes of similar structures. Let us consider the most familiar structure which is the natural numbers $0, 1, 2, 3, \ldots$. The structuralist point of view is that a single number, say 4, does not have any meaning. It has a meaning only as a part of the structure, namely, that *there are four numbers less than it*. Notice that we need the relation '*less than*', without it we could not distinguish 4 in this way. Furthermore we can add and multiply numbers (this is the 'playing with a toy' alluded to above). Thus we arrive at the following description of the natural numbers as a mathematical structure: they consists of

1. the *set* of nonnegative integers $\{0, 1, 2, 3, \ldots\}$, called the *universe*, or the *base set*, or the *underlying set* of the structure;
2. the *operations* of addition $+$ and multiplication $\cdot$;
3. the *relation* of being less than or equal $\leq$.

Notice the stressed words *set*, *operations*, and *relation*. This is, in fact, the form of all basic structures: they consist of a set on which there are some operations and relations defined. We do not restrict the number of operations and relations, except that their number must be finite. In particular, a structure can have only relations or only operations. For example, we may consider the natural numbers only with the ordering relation, or, on the contrary, we may add more operations. The natural numbers with no operations and $\leq$ as the only relation form a much simpler structure, but they are important when we are interested in a particular class of structures, namely, ordered sets.

In our example above the operations are *binary*, which means that they produce an element from 2 elements. Obviously, one can consider operations with this parameter 2, called the *arity*, replaced by any natural number.[1] In particular, operations of arity 0 are called *constants* and operations of arity 1 are called *functions*. Operations with arity greater than 2 are rare. The arity of a relation can be any number greater or equal than 1. A unary relation, that is a relation of arity 1, is usually called a *predicate*, or a *property*. An example of a ternary relation is the relation '*x is between y and z*' used in the formalization of plane geometry.

It probably required a considerable psychological effort for mathematicians to realize that the underlying set, the universe of a structure, does not determine the relations and operations. For example, originally people thought of the natural numbers as something intrinsically associated with the natural ordering and the two basic operations. The realization that we are completely free to choose operations and relations (and that the resulting structures can be interesting and useful) led to a dramatic development of mathematics in the 19th century, especially in algebra. A similar revolution occurred in physics one century later. In the 20th century theoretical physicists discovered that mathematics offers not only the classical structures of

[1] 'Arity' is not an English word, but it is common in mathematical jargon. The word is derived from the suffix *-ary*.

mathematical analysis, but many more, and they can be very useful in physics. This started with Einstein's use of the tensor calculus on manifolds in general relativity theory and Heisenberg's use of matrices in quantum mechanics.

Now, what happened with quantities? Modern mathematics has replaced this informal term by the concept of *function*. When describing some real phenomenon by two numbers x and y, where the number y is uniquely determined by the number x, we say that *y is a function of x*. This is formally written as

$$y = f(x).$$

We call x a *variable* and y the *value*, and f is a symbol by which we denote the function. The basic functions have names, such as '*square of*', '*sine*', '*exponential*', ..., and they are often expressed using special notation,

$$x^2, \quad \sin x, \quad e^x, \quad \ldots.$$

More generally, y may depend on several variables. Thus, in particular, operations are also functions. We use the word '*operation*' in situations when the function of several variables possesses some "nice" properties. This is the case of the operations of addition and multiplication on the natural numbers: they are commutative and associative (which means that the sums and products do not depend on the order in which they are computed).

If f is, say, a function defined on the real numbers, then it can be studied as the structure consisting of

1. the universe $\mathbb{R}$, which is the set of real numbers, and
2. the function f, as an operation.

It may seem that I am too fastidious about details when mentioning the universe. Isn't the structure already determined by the function? When we are stating theorems about the structure, it must be clear what the elements we are talking about are. We use the universe to determine the range of elements. It is a sort of a universe in which things concerning the structure take place.

I assume that the reader already knows most of these elementary concepts, but it is good to recall the terminology before discussing more difficult ones.

Ordered Sets

Let us now consider an example of a *class* of structures. The structures in the class are called *ordered sets*. This is probably the most ancient kind of structure. As soon as people started to organize their things they made lists by ordering the items that they considered. In fact this structure is imposed on essentially all data people use. We use language which is a sequence of words; written records are also sequences. So things are communicated in some order, whether we want to stress it or not. It is also interesting to note that the word '*ordering*' comes from '*order*' which also

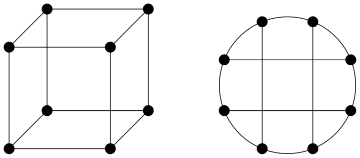

Fig. 1.1 Two drawings of the graph of the cube

means that things are properly organized, the opposite of disorder. And this is in fact the main purpose of mathematical structures, namely, to organize things, to introduce some order into our observations and data, so that we are able to manipulate them efficiently, physically and mentally.

The most obvious example of an ordered set is the set of natural numbers with the ordering relation $\leq$ that I mentioned above. Other familiar examples are the structure of the integers with ordering and the structure of the real numbers with ordering. These three structures are essentially different, not only because they have different universes, but because they have different structures, now using the word in the usual meaning. It does not matter how we represent the natural numbers, the integers and the real numbers, there will always be something different. The natural numbers are distinguished from the integers and the reals by the fact that they contain a smallest element. In the integers there are pairs of numbers such that there is no element between them, for example, there is no element strictly between 0 and 1. This is not true for the structure of reals: for every two elements, there exists an element between them (their mean is such an element).

Graphs

The word '*graph*' is used in two meanings. The traditional one is the diagram of a function, such as the dependence of the price of some commodity on time. It has a different meaning in the modern branch of mathematics that studies discrete structures, *the theory of graphs.* In this theory a *graph* consists of *points* and *arcs* that connect some points. This looks like a geometric concept, and it did originate in geometry, but it has more to do with topology than geometry. Consider for instance a cube. A cube determines in a natural way a graph, where we take the vertices of the cube as points and the edges of the cube as arcs, see Fig. 1.1. In fact, the standard terminology uses '*vertices*' and '*edges*' for all graphs. The reason why graph theory is so different from the classical fields of mathematics is that we completely abstract from the nature of vertices and edges and we only consider facts that depend on information about which vertices are connected and which are not. So if our cube is made of rubber and we twist it, the graph will be the same.

As another example of a graph, let us consider the graph of the flight connections of an airline. You can think of it as cities on a world map connected by arcs. On most such maps the arcs have little to do with the actual routes that an aircraft takes when flying between the two cities. An actual route must follow particular corridors, which is irrelevant for a passenger who only wonders whether there is a direct flight from city X to city Y.

Groups

If mathematicians voted for the most important class of structures, they would probably elect *groups*. The name is just a historical accident, so do not try to guess the meaning from normal use of the word. This concept is slightly more difficult, but worthwhile to learn. A group is a structure with one binary operation which in some sense behaves nicely. What this means precisely can be defined by postulating some simple laws that the operation must meet, which I will state shortly (page 10). Here I will only explain the concept in plain words.

The best way to imagine a group is to think of the elements of the group as *reversible actions* and the group operation as the composition of actions. As usual in mathematics, taking no action also counts as an action, called the *unit element*. Note that there is an important conceptual shift here: the actions themselves are elements, not the objects on which they act. Rubik's cube and similar toys are excellent examples. For Rubik's cube group, an action is, for example, turning the front face clockwise 90°, or turning the top by 180°. These are just some elementary actions. An action, however, may be more complex. For instance, we can compose the first one with the second one and this is also an action. We will get a different action, if we start with the second one and then apply the first one. The trick to solving this puzzle is to have several complex actions which do some particular things, such as turning two neighboring corners in opposite directions while keeping the rest the same. To transform a particular position into the original position is also an action. The goal is to compose this action from the elementary ones.

As you can imagine, the group of Rubik's cube is not a very simple one, it has $2^{12} \cdot 11! \cdot 3^8 \cdot 8!$ elements. There are groups which have infinitely many elements, but whose structure is simpler. Namely, one of the basic groups is the group of integers where the group operation is addition. To visualize it as a group of actions think of it as adding money to and withdrawing money from an account, say, starting with balance 0. Adding money is represented by positive integers, withdrawing by negative ones. This structure is the additive part of the structure of the integers that we considered earlier.

Groups are also essential in the study of *symmetries*. Consider a simple symmetric object, say an equilateral triangle A, B, C. We call a rigid action which transforms the triangle to itself a symmetry. There is a trivial symmetry corresponding to "no action", which we have, in fact, for any geometrical object. A nontrivial symmetry is the rotation where A goes to B, B goes to C and C goes to A. We can describe it by the list $A \to B$, $B \to C$, $C \to A$, or by saying that we rotate counterclockwise by 120°. We have one more rotation for 240°. Then we have another type of symmetry—we can flip the triangle along its axes of symmetry. For instance, flipping along the axis going through A can be described as interchanging B with C while A does not move. Another natural way of representing the same group is by permutations of three elements. The six permutations

$$(A, B, C), \quad (C, A, B), \quad (B, C, A), \quad (B, A, C), \quad (C, B, A), \quad (A, C, B)$$

are the elements of the group. They correspond to the identity, the transformation that does not move anything, and the symmetries denoted by a, b, c, d, e in Fig. 1.2.

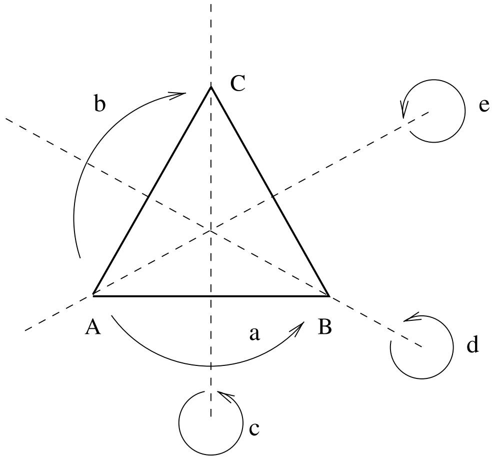

Fig. 1.2 The symmetries of an equilateral triangle

Fig. 1.3 The multiplication table of the group of symmetries of an equilateral triangle. The unit element of the group is denoted by 1

	1	a	b	c	d	e
1	1	a	b	c	d	e
a	a	b	1	e	c	d
b	b	1	a	d	e	c
c	c	d	e	1	a	b
d	d	e	c	b	1	a
e	e	c	d	a	b	1

The identity is the unit element of the group and is denoted by 1. The group operation is the composition of two permutations. For example, (C, A, B) is the transformation $A \mapsto B$, $B \mapsto C$, $C \mapsto A$ and (B, C, A) is the transformation $A \mapsto C$, $B \mapsto A$, $C \mapsto B$. Hence their composition is the identity (A, B, C).

These two representations use specific properties of the group. A general way by which we can represent any binary operation is the multiplication table. The multiplication table of the group of symmetries of an equilateral triangle is in Fig. 1.3.

Finally, we consider a way of representing groups that plays an important role in the study of finite groups—representations by matrices. In this way problems about finite groups can be translated into problems about matrices. Matrices form a very rich structure with a lot of interesting concepts and important theorems. The study of such representations is so useful that it forms a separate field called the *group representation theory*. Here is one such representation of the group of symmetries of an equilateral triangle.

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}, \quad \begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix}, \quad \begin{pmatrix} -1 & 1 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ 1 & -1 \end{pmatrix},$$
$$\begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}.$$

(For the definition of the matrix product see page 396.)

We have seen four representations of the same group. Each of them determines the structure of the group, but the group as an abstract object cannot be identified with any of them.

Why do we need various structures, why do we not just use numbers? The examples of graphs and groups show that there are practical situations which cannot be described only by numbers. We can think of structures as *models* of real and potentially realizable situations. Another possible view is that structures give us ways to classify objects. One useful way of classifying collections is to count the number of elements. We count our pieces of luggage to check that we have them all, which clearly does not ensure that we have all *our* luggage. But this test usually works. Numbers are not the only kind of structure used for such a classification. In particular groups are very good for this purpose. They are used in crystallography, to name a practical application. The symmetry group can be used to distinguish various objects, but it does not necessarily determine them completely.

In mathematics such a use of groups is almost ubiquitous. Returning to our example, we can distinguish the triangle from other geometrical objects by its group of symmetries. It is rather awkward here, as the triangle is much simpler than its symmetry group, but for larger objects it makes sense. In this case we would rather use the triangle to define the group.

One of the most beautiful pieces of mathematics, which I will consider in some detail in Chap. 4, is also based on this concept. This is the famous result that algebraic equations of degree 5 are not solvable using radicals. This means that there is no explicit formula using basic arithmetic operations and roots, expressing a solution to the equation in terms of the parameters. Here we have a natural scale given by the degree of the equation. But this gives us no clue why equations should be solvable up to degree 4, and unsolvable from degree 5 on. It was a great insight of Galois that one should assign groups to equations. The kind of groups that can be associated with equations of degree 5 and higher do not occur for equations of lower degree, and this gives the distinction between the solvable and the unsolvable.

Let me finally mention a result which belongs among the major achievements of twentieth century mathematics. The result is interesting also because it is a theorem with the longest proof ever written by mathematicians. It is called *the classification of simple finite groups*. The word “simple” is a little misleading; it is a technical term which specifies groups that are in some sense basic building blocks for constructing all finite groups. Naturally, having a description of them is very important, if we want to understand finite groups. The whole result is contained in a series of papers produced by a number of first rate mathematicians. The total number of pages amounts to several thousands. Some simple groups had to be described explicitly, the smallest one with $2^4 3^2 5 \cdot 11 = 7\ 920$ elements, the largest one having $2^{46} 3^{20} 5^9 7^6 11^2 13^3 17 \cdot 19 \cdot 23 \cdot 29 \cdot 31 \cdot 41 \cdot 47 \cdot 59 \cdot 71$ (approximately $8 \cdot 10^{53}$) elements, called *the Monster*. The enormous length of the proof and the huge size of the groups that it describes are certainly remarkable, but what is also interesting is a strange kind of irregularity. We are used to the fact that in mathematics things tend either to be very regular, or to look very random; if there is regularity with some exceptions then the exceptions are small. Here, in contrast, we have 26 exceptions that share very few common properties.

Types of Structures

In order to give a more precise meaning to the concept of a structure, we have to use more technical means of mathematics, some notation, and a few symbols. Formally, a structure is given by a list that consists of several sets. The first set is the universe, the set of objects of the structure. The remaining sets are relations, functions and operations on the universe. Let us denote by $\mathbb{N}$ the set of all natural numbers. Taking $\mathbb{N}$ as the universe, we can define various structures. The universe by no means determines the structure, however, there are some structures with this universe that we like more than the others. On the set $\mathbb{N}$ we usually take the following one $(\mathbb{N}; +, \cdot, \leq)$. To stress the special role of the universe, it is separated from the other sets by a semicolon. In this structure the binary relation $\leq$ is superfluous because we can define it from the operation $+$ (namely, $x \leq y$ if and only if there exists a z such that $x + z = y$), but we may have other reasons for keeping it. This structure has two binary operations and one binary relation—this information is what we call the *type of the structure.* Let $\mathbb{R}$ denote the set of all real numbers. We can define a structure of the same type as the natural numbers by taking $(\mathbb{R}; +, \cdot, \leq)$.

A different example is a directed graph. It is determined by a set of vertices and a general binary relation. Hence we can say that directed graphs are structures of the type consisting of one binary relation.

Structures with one binary operation also have a special name; they are called *magmas*, or *groupoids*.[2] Groups can be defined as those structures with one binary operation that satisfy the following axioms:

1. there exists a *unit element* (an element, usually denoted by 1, that satisfies $x \cdot 1 = 1 \cdot x = x$);
2. the operation is *associative* ($(x \cdot y) \cdot z = x \cdot (y \cdot z)$ for all elements x, y, z);
3. every element has its *inverse* (the inverse of x is usually denoted by x^{-1} and satisfies $x \cdot x^{-1} = x^{-1} \cdot x = 1$).

Groupoids and groups belong to a large class of structures, called *algebraic structures*, or *universal algebras*, which are structures that only have functions and operations, but no relations.

All the structures that we have considered so far are *first-order structures*. There are structures that use more complex objects; such structures are called *second order, third order*, etc. In second order structures we have sets of subsets of the universe and relations between such subsets. This can be explained as follows. In a second order structure we have two universes, one consists of the elements that we want to study, the other consists of *sets of elements*, which we call *second order elements*. In a second order structure we also have relations and functions defined on second order elements. In order to imagine second order elements, think of subsets of the universe as properties of elements and sets of these subsets as properties of properties.

[2] Not to be confused with groupoids in category theory.

Example Let us take the color *navy blue* as an example of a property of real objects. Then we can take *dark colors* as an example of a property of properties that contains navy blue as an element.

If we attempt to define second order structures in full generality things become quite complicated. We can consider not only relations between subsets, but also between subsets and elements. Furthermore we should allow talking about properties of relations. But that is still not enough, since functions are also first-order objects, so we should allow relations between functions and so on. It is rather complicated, but it is only a technicality. The essence is that we have certain levels: the zero level are elements, the first level are relations and operations. In a second order structure we can define relations and operations on all objects of the first two levels.

The simplest example of a class of second order structures is the class of topological spaces. Topological space consists of a set of *points* A (this is the universe), and a set of subsets of A, called *open sets* that must satisfy some laws. For instance, the real numbers as topological space (called *the real line*) have the universe $\mathbb{R}$ and the open sets are subsets of reals which are unions of open intervals. (An open interval is the set of numbers between r and s *not* including the endpoints r, s.) The empty set is defined to be open too. Intuitively an open set is a set which does not contain a point on its border.

Let us proceed to the third order. This essentially means that we allow subsets of all subsets of the universe. There is no reason to stop at the third order, but already there it is hard to find nice examples. Let us take the first-order structure of reals $(\mathbb{R}; +, \cdot, \leq)$. Extend it to a second order structure by adding the set of all continuous functions of one variable, denoted by $\mathcal{F}$. Then we would like to consider the limits of the continuous functions, so we add a topology on $\mathcal{F}$ by taking the set of all open subsets of the functions, denoted by $\mathcal{X}$. This results in a third order structure $(\mathbb{R}; +, \cdot, \leq, \mathcal{F}, \mathcal{X})$. In this structure $+$, $\cdot$, $\leq$ are first-order concepts, $\mathcal{F}$ is second order and $\mathcal{X}$ is third order.

The types of structures are associated with certain set-theoretical constructions. The first one is the *Cartesian product* of sets. The Cartesian product of two sets X and Y is denoted by $X \times Y$ and it is the set of all pairs of elements (x, y) where x is an element of X and y is an element of Y. The reason for using $\times$ is that the size of the Cartesian product is the product of the sizes of the two sets; otherwise this set operation shares very little with the corresponding operation on numbers. Clearly, we can iterate this operation to get the product of a finite number of sets. The name '*Cartesian*' is in honor of the French mathematician and philosopher René Descartes (1596–1650), to whom we attribute the invention of coordinates and analytic geometry (although some analytic methods in geometry had already been used in ancient Greece). In modern terms it means that one dimensional space can be identified with the set of real numbers, and higher dimensional spaces are simply the products of copies of one dimensional space. His invention was probably the first step in the process of formalization of mathematical objects by mathematical structures. Mathematicians very often use pictures to visualize structures that they are thinking about. In the case of the Cartesian product $X \times Y$ the picture is the familiar one with X drawn as the coordinate x, Y the coordinate y and the product being the points on the plane. The Cartesian product corresponds to relations, since we can define relations on a set A as subsets of the products of A with itself. Thus a subset of A is a unary relation, a subset of $A \times A$ is a binary relation, etc.

The second set operation is related to exponentiation and thus it is denoted by Y^X. It is the set of all functions f defined on X and having values in Y. Instead of saying that f is an element of Y^X, we prefer to express it by $f : X \to Y$. The Cartesian product enables us also to define functions with several variables, which we call operations. Thus, for example, a binary operation f on a set A is an element of $A^{A\times A}$, or using the other notation $f : A \times A \to A$ (which is read as '*f maps $A \times A$ into A*'). For higher order structures, we need yet another set operation. Let us denote by $\mathcal{P}(A)$ the *power set* of the set A, the set of all subsets of A. Thus, for example, relations between second order elements are subsets of $\mathcal{P}(A) \times \mathcal{P}(A)$.

This notation can be used to define types of structures, but for this book we do not need a formal definition. Moreover, there are types of structures that do not quite fit into this schema. In classical parts of mathematics real numbers play a key role, thus many structures are somehow connected with them. Consider, for instance, a *real vector space*. It is a set of vectors A and a binary operation on A, usually denoted by $+$, satisfying certain axioms (namely $(A; +)$ is a commutative group). Furthermore, for every real number r, we can multiply any vector a of A by r and thus obtain another vector of A. This does not fit into the above schema, as the real numbers are not in $(A; +)$, they are *external*. In order to define this structure we have to take the union of the two structures—the real numbers and the group of vectors. The resulting object can be denoted by $(\mathbb{R}, A; +_{\mathbb{R}}, \cdot_{\mathbb{R}}, +_A, \cdot_{\mathbb{R},A})$. I have distinguished the two additions and two multiplications by subscripts, (to be more precise, we should write specifications such as $\cdot_{\mathbb{R},A} : \mathbb{R} \times A \to A$ which is multiplication of a vector by a real number, etc.). So we have to generalize the concept of a structure further and allow more than one universe. Also notice that in this particular example the roles of the two universes $\mathbb{R}$ and A are different: while A may vary arbitrarily, $\mathbb{R}$ is fixed for all real vector spaces.

For understanding the foundations of mathematics we do not have to study the whole ramified variety of structures. The most interesting phenomena can be observed in simple first-order structures.

Structures of Structures

In order to understand structures, it is important to realize that only *the form* is important, not the content. This means that the nature of the elements is irrelevant. The word '*structure*' denoting this concept is chosen appropriately, as we would like to identify two objects that have *the same structure*. Thus to get the whole point we only need to define what '*the same structure*' means. Intuitively it means that we can move one structure so that it completely coincides with the other. To move the structure means to move the points of the universe, the rest, the relations and operations will move along because it is attached to the points. In mathematics structures do not live in space, so the transformations from one structure into another one are not continuous transitions (unless we incidentally study topology). Thus we only need to specify the beginning and the end of the movement. This is done by the concept of *mapping*. (A mapping and a function are the same things; we use different names only because of the different context.) Such a mapping should be *one-to-one*, which means that no two points are mapped onto one, and it should be *onto*, which means that every point of the universe of the second structure is an image of a point of the first structure. The mapping translates in a natural way relations and operations from one structure into the other. If the resulting image is identical with the second structure, we say that the structures are *isomorphic*. Isomorphism is the mathematical concept of having the same form. We often do not distinguish structures that are isomorphic and often say that '*two structures are the same, up to isomorphism*'.

To understand the above definition, think of the problem of comparing two pictures on a film in order to check if they are the same. First you have match them correctly. This means that you need some special points, in this case two are enough, which determine the correct position. If you put the pictures so that the points coincide, then it suffices to check if every line, every spot, etc. coincides.

The study of mappings of one structure into another is not restricted to isomorphisms. Given a class of structures one defines a more general concept, called *homomorphisms* or just *morphisms*, by using more general mappings. In particular, a homomorphism does not have to be a one-to-one mapping, hence it can map several elements on one. In this way some information about the structure on which it is defined may be lost in its image. Homomorphisms enable us to formalize the intuitive concept of similarity. The ability to recognize similarities is one of the most important features of human and animal thinking. Thus it is not surprising that in modern algebra many important results can be stated purely in terms of morphisms. A class of structures and morphisms is in some sense also a structure; it is called a *category*. We can study a class of structures by studying its category.

The Four Color Theorem

I will conclude this section with a couple of mathematical results that will be used as examples in the following chapters.

In 1852 an English mathematician, Francis Guthrie, conjectured that every map can be colored by four colors so that no two neighboring countries have the same color. This is, perhaps, the most famous problem in combinatorics, or at least it had been so until it was solved by Kenneth Appel and Wolfgang Haken in 1975 [5, 6]. The original statement talks about the topology of the plane, but it can be stated as a problem about certain graphs. Given a map, represent countries as vertices, say choose a point inside every country. Then connect by an arc every two vertices that come from neighboring countries. Then, instead of coloring countries, we will color vertices. The restriction is that two vertices connected by an arc must have different colors. This simple transformation shows why graphs are so useful. We can transform a rather complicated statement to a simple combinatorial one.

This reduction alone does not suffice to translate the problem to graph theory. Not every graph corresponds to a map and it is very easy to construct a graph that is not colorable by four colors (take five vertices and connect every pair of vertices). Thus we need a characterization of graphs that come from maps; these graphs are called *planar*, as they come from maps in the plane. Such a purely combinatorial characterization was found by Kazimierz Kuratowski (1896–1980), a Polish set theorist and topologist; thus the problem has been reduced to finite combinatorics.

Whether or not every map can be colored by four colors has no bearing on the foundations of mathematics. What has is the way the problem was solved. Appel and Haken did not write down a proof of the conjecture, they only tested by computer that a proof exists. Following some earlier results they reduced the problem to a finite number of cases that were possible then to check by computer. Each particular case can be checked "by hand", but the total number of cases is too large for a human, even with the more recent improvements that have reduced the number of cases. This raised a discussion as to whether such proofs are legitimate. Certainly, such a proof conveys less to a mathematician than a usual proof. Typically, a proof is based on a small number of ideas that one can memorize so that it is possible to reconstruct the formal proof when needed. The experience of mathematicians with long proofs is that they are very likely wrong if such a set of basic ideas cannot be extracted from them. I agree with that, as this concerns proofs that are written by people and such proofs are never completely formal. Once the things are done formally, computers are much more precise than people. By now the validity of the theorem has been verified by running the programs on different machines and by using alternative proofs written by different people. What remains a mystery is why we do not have a 'normal' proof, a proof sufficiently short to be understood by people. As we will see later, there are theorems that do not have short proofs. But our mathematical tools are still very limited and thus we are not able to prove for such concrete theorems almost anything about the lengths of their proofs.

Note that there is a generalization of this problem to all orientable surfaces. Interestingly enough, the generalization had been solved for surfaces of all genera, except for the plane, without using a computer and before the original problem was solved.

The Four Color Theorem was not the first case in which an infinite problem was reduced to a finite number of cases. The famous Goldbach Conjecture, probably the oldest unsolved problem in mathematics, says that every even natural number greater than 2 can be expressed as the sum of two prime numbers. A weaker conjecture states that every odd number greater than 7 is a sum of three odd primes. In the 1930s, the Russian number theorist Ivan M. Vinogradov proved the weaker conjecture for all odd numbers starting from some large number N_0 [299]. Thus, theoretically, it sufficed to check all odd numbers less than N_0 in order to complete the proof. Unfortunately the number N_0 was so large (the original estimate was $e^{e^{e^{42}}} \approx 10^{10^{10^{17}}}$) that there was no chance to check the remaining cases by computation. This is still the case, in spite of the bound on N_0 being substantially reduced and in spite of the possibility to use contemporary powerful computers.[3]

More recently another famous problem has been solved using a computer in a similar way as in the Four Color Theorem. It is the Kepler Conjecture that the densest arrangement of equal balls is, in fact, the one that people have always been using. In 1998 S.P. Ferguson and T.C. Hales announced a proof of the conjecture [112]. It is based on a reduction proposed by L. Fejes Tóth in the 1950s. Since the computations used computer arithmetic, some doubts about the completeness of the proof still persist.

One may expect that computer aided proofs would be quite widespread by now, but it is not so. It turns out that there is a very narrow window where computers may help mathematicians. If ever a proof can be reduced to a finite number of cases, then, usually, either the problem can be solved completely by a mathematician, or the number of cases is so huge that it cannot be checked even by a computer.

Ramsey's Theorem

Frank P. Ramsey (1903–1930), a British mathematician and philosopher, proved a lemma that he needed in order to solve a certain problem in logic (the decidability of a certain part of first order logic) [235]. The lemma was later rediscovered by Paul Erdős and Gyorgy Szekerés working in a totally different field [69]; since then it became one of the main parts of combinatorics. Today this lemma is called *Ramsey's Theorem* and plays an equally important role also in logic and set theory. Therefore this theorem is very useful when we want to illustrate the connections between various fields of mathematics.

The essence of the theorem can easily be explained to anybody. Suppose that you have a symmetric binary relation on a finite set. Such a relation is also called an 'undirected graph', or just a 'graph'. Traditionally, for this theorem, one takes a random group of people and the relation of knowing each other as an example of a graph. The question that this theorem addresses is to what degree the relation can be chaotic, or put positively, must there be at least some order in any such relation? There are many ways to define the degree of order, but the extreme cases are clear: if every pair is connected by the relation, then clearly it has the maximal order; by the same token, if no two are connected it also has the maximal order. Ramsey's theorem, roughly speaking, says that total chaos is impossible. More precisely, we can always find a small subset of vertices where either all elements are connected in the graph, or all elements are not connected. For example, if there are at least 6 people in the group, there must be at least 3 that all know each other or all do not know each other (see Fig. 1.4). Similarly, if the group has at least 18 people, then

[3] The very recent result of T. Tao [289] that every odd integer greater than 1 can be represented as a sum of 5 or fewer primes uses the fact that the Goldbach conjecture has been verified by computation for all numbers up to $4 \cdot 10^{14}$.

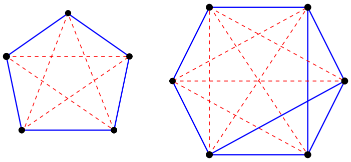

Fig. 1.4 Examples of a colorings of pairs of elements of a 5-element set and a 6-element set. The coloring of the 5-element set shows that $R(3) > 5$ because there is no 3-element monochromatic set. Since $R(3) = 6$, there must exist 3 points connected by lines of the same color in any coloring of a 6-element set. In the example there are two such triples, both form blue (solid line) triangles

there must be at least 4 that all know each other or all do not know each other. For 5, it suffices to have 46 people in the group.

In general, for every number n, we can find a number r, such that a graph on r vertices contains a subset of size n where either all elements are connected or no pair is. The *Ramsey number* $R(n)$ is defined as the least r such that every graph on r vertices contains a subset of size n where either all elements are connected or no pair is. The theorem says that this is a correct definition, such a number exists for every n.

The above examples can be stated as $R(3) \leq 6$, $R(4) \leq 18$ and $R(5) \leq 46$. In fact we know that $R(3) = 6$ and $R(4) = 18$, but we do not know the exact value of $R(5)$. We only know that $43 \leq R(5) \leq 46$. This is remarkable since to determine the value of $R(5)$ is a finite problem, one has "only" to check all the graphs on 43, 44 and 45 vertices. Testing a single graph is not so difficult (though it is quite a time consuming task—there are more than one million subsets of size 5 of a set of size 45), the problem is that there are too many graphs to be tested.

The classical infinite version of the theorem states that for every graph on the natural numbers, there is an *infinite* subset of the natural numbers such that either all elements in the subset are connected, or no pair is. A remarkable fact is that the finite version of the theorem can be derived from the infinite one. The advantage of such a proof of the finite version is that we do not have to bother with counting. The disadvantage, the price for the simplification, is that we do not get any bounds on the Ramsey numbers.

Notes

1. *General structures.* A general structure is defined by an *echelon construction.* The construction starts with base sets (universes) $A_1, \dots, A_n$. Then we can apply operations of the Cartesian product $\times$, the power set operation $\mathcal{P}$ and the operation of taking the set of all functions from one set into another set B^A. This means that we successively produce sets such that every new set is obtained from $A_1, \dots, A_n$ and the already produced ones by applying one of the three operations. We identify products of several sets if the order of the sets in them is the same; for instance, we do not distinguish between $(B_1 \times B_2) \times B_3$ and $B_1 \times (B_2 \times B_3)$. Thus we can omit parentheses in the products. A structure is a sequence of the form $(A_1, \dots, A_n; B_1, \dots, B_m)$ where $B_1, \dots, B_m$ are subsets of the sets obtained by the echelon construction or mappings between them.

For example, our third order structure considered above $(\mathbb{R}; +, \cdot, \leq, \mathcal{F}, \mathcal{X})$ is produced from the sequence $\mathbb{R}$, $\mathbb{R} \times \mathbb{R}$, $\mathbb{R}^{\mathbb{R}}$, $\mathcal{P}(\mathbb{R}^{\mathbb{R}})$, where the operations $+$ and $\cdot$ are mappings from $\mathbb{R} \times \mathbb{R}$ to $\mathbb{R}$, the relation $\leq$ is a subset of $\mathbb{R} \times \mathbb{R}$, the set $\mathcal{F}$ is a subset of $\mathbb{R}^{\mathbb{R}}$ (the set of all real functions) and $\mathcal{X}$ is a subset of $\mathcal{P}(\mathbb{R}^{\mathbb{R}})$ (the set of all subsets of real functions).

In a precise definition of a structure we have to associate a *type* to each of the sets. In particular, in first-order structures this means determining if it is a relations or an operation and then its arity. first-order structures are those where neither of the operations $\mathcal{P}(X)$, X^Y is used. In second order structures these operations can be applied, but not iterated, in third order structures they may be iterated once etc.

It is possible to simplify the matter by considering operations and functions as a special kind of relations (for example, a binary operation is a ternary relation). However, quite often, it is an advantage to have operations as a primitive concept.

2. *Higher type functionals.* General structures can use all three operations: Cartesian product, power set operation, and the operation of taking all functions from a given structure to another one. We can get, however, very interesting objects by considering only the last one. This means to concentrate on functions and not to use relations and sets. We start with elements as the basic type of objects; the set of elements is the universe of the structure. The next type consists of functions. A function is a mapping from the universe to itself. Then we can define functionals, which are mappings that map functions to elements. We can use also mappings that assign functions to elements and mappings that assign functions to functions and so on. We will simply call all such objects *functionals* and distinguish them by their *types*. As the types do not have linear structure, we cannot use numbers for denoting types, we need to introduce special notation. The type of elements will be denoted by o ('o' for 'objects'). Given types τ and σ, the type of functionals that map objects of type τ to objects of type σ will be denoted by $\tau \to \sigma$. Thus functions are functionals of type $o \to o$, the lowest level functionals are $(o \to o) \to o$, etc. Note that functionals of type $o \to (o \to o)$ can be identified with binary operations, that is, functions of two variables.

   Now we will consider some important classes of structures.

3. *Ordered sets.* An ordered set is a structure with one universe and one binary relation on it denoted usually by $\leq$ (ambiguously, because the relations in different structures are different). By an ordered set we usually mean a *partially* ordered set which means that there may be incomparable elements. The axioms of partially ordered sets are

   a. $x \leq x$—reflexivity,
   b. $x \leq y$ and $y \leq z$ implies $x \leq z$—transitivity,
   c. $x \leq y$ and $y \leq x$ implies $x = y$—antisymmetry.

   The ordered sets where every two elements are comparable are called *linear orderings*; they satisfy also

   d. $x \leq y$ or $y \leq x$.

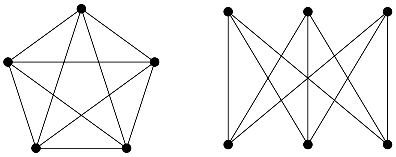

Fig. 1.5 The graphs K_5 and $K_{3,3}$

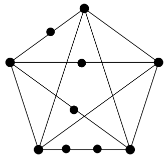

Fig. 1.6 A subdivision of K_5

4. *Graphs.* A general graph is a binary relation on the set of vertices. It is called a *directed graph* because we may have a directed edge (u, v) without having the opposite (v, u). Edges of the form (u, u) are called loops. For instance, partially ordered sets are a subclass of graphs. Graphs in the narrow sense are symmetric, which means (u, v) is an edge if and only if (v, u) is, and loops are prohibited. We denote by (u, v) an ordered pair. For symmetric graphs, we can take unordered pairs which are two-element sets. They are denoted by $\{u, v\}$. (Sometimes a more general concept is considered where there can be more than one arc between two vertices.)

   Kuratowski's characterization of *planar graphs* is based on forbidden subgraphs. He found a set of graphs such that planar graphs are exactly those that do not contain a graph from the set. The set of forbidden graphs consists of the two graphs in Fig. 1.5 *and all their subdivisions.* A *subdivision* of a graph is obtained by refining edges into paths; pictorially, we put several dots on an edge (see Fig. 1.6).
5. *Groups.* A group is usually considered as a structure with one binary operation, one unary operation (a function) and a constant. These are called *multiplication, the inverse element function* and *the unit element.* Thus we write $(G; \cdot, {}^{-1}, 1)$. The inverse element and the unit is definable from multiplication, but having these two additional primitives enables us to write axioms as equations:

   a. $1 \cdot x = x \cdot 1 = 1$,
   b. $x \cdot x^{-1} = x^{-1} \cdot x = 1$,
   c. $x \cdot (y \cdot x) = (x \cdot y) \cdot x$.

   Note that we do not postulate commutativity. You can check that the symmetry group of a triangle is not commutative. The groups where the commutative law $x \cdot y = y \cdot x$ holds are called *commutative* or *Abelian* groups. For commutative groups, one often uses additive notation, thus instead of calling the binary operation '*multiplication*' we call it '*addition*'.

   We will now define some concepts needed to explain the meaning of simple groups.