Universitext

Universitext

Editors (North America): S. Axler and K.A. Ribet

Aguilar/Gitler/Prieto: Algebraic Topology from a Homotopical Viewpoint
Aksoy/Khamsi: Nonstandard Methods in Fixed Point Theory
Andersson: Topics in Complex Analysis
Aupetit: A Primer on Spectral Theory
Bachman/Narici/Beckenstein: Fourier and Wavelet Analysis
Badescu: Algebraic Surfaces
Balakrishnan/Ranganathan: A Textbook of Graph Theory
Balser: Formal Power Series and Linear Systems of Meromorphic Ordinary Differential Equations
Bapat: Linear Algebra and Linear Models (2nd ed.)
Berberian: Fundamentals of Real Analysis
Blyth: Lattices and Ordered Algebraic Structures
Boltyanskii/Efremovich: Intuitive Combinatorial Topology. (Shenitzer, trans.)
Booss/Bleecker: Topology and Analysis
Borkar: Probability Theory: An Advanced Course
Böttcher/Silbermann: Introduction to Large Truncated Toeplitz Matrices
Carleson/Gamelin: Complex Dynamics
Cecil: Lie Sphere Geometry: With Applications to Submanifolds
Chae: Lebesgue Integration (2nd ed.)
Charlap: Bieberbach Groups and Flat Manifolds
Chern: Complex Manifolds Without Potential Theory
Cohn: A Classical Invitation to Algebraic Numbers and Class Fields
Curtis: Abstract Linear Algebra
Curtis: Matrix Groups
Debarre: Higher-Dimensional Algebraic Geometry
Deitmar: A First Course in Harmonic Analysis (2nd ed.)
DiBenedetto: Degenerate Parabolic Equations
Dimca: Singularities and Topology of Hypersurfaces
Edwards: A Formal Background to Mathematics I a/b
Edwards: A Formal Background to Mathematics II a/b
Farenick: Algebras of Linear Transformations
Foulds: Graph Theory Applications
Friedman: Algebraic Surfaces and Holomorphic Vector Bundles
Fuhrmann: A Polynomial Approach to Linear Algebra
Gardiner: A First Course in Group Theory
Gårding/Tambour: Algebra for Computer Science
Goldblatt: Orthogonality and Spacetime Geometry
Gustafson/Rao: Numerical Range: The Field of Values of Linear Operators and Matrices
Hahn: Quadratic Algebras, Clifford Algebras, and Arithmetic Witt Groups
Heinonen: Lectures on Analysis on Metric Spaces
Holmgren: A First Course in Discrete Dynamical Systems
Howe/Tan: Non-Abelian Harmonic Analysis: Applications of SL (2, R)
Howes: Modern Analysis and Topology
Hsieh/Sibuya: Basic Theory of Ordinary Differential Equations
Humi/Miller: Second Course in Ordinary Differential Equations
Hurwitz/Kritikos: Lectures on Number Theory
Jennings: Modern Geometry with Applications

(continued after index)

Wolfgang Rautenberg

A Concise Introduction to Mathematical Logic

Springer

Wolfgang Rautenberg
FB Mathematik und Informatik Inst.
Mathematik II
Freie Universität Berlin
14195 Berlin
Germany
raut@math.fu-berlin.de

Mathematics Subject Classification (2000): 03-XX 68N17

Library of Congress Control Number: 2005937016

ISBN-10: 0-387-30294-8
ISBN-13: 978-0387-30294-2

Printed on acid-free paper.

Printed in the United States of America. (SBA)

9 8 7 6 5 4 3 2

springer.com

Wolfgang Rautenberg

A Concise Introduction to Mathematical Logic

Textbook

Typeset and layout: The author
Version from December 2005

Foreword

by Lev Beklemishev, Utrecht

The field of mathematical logic—evolving around the notions of logical validity, provability, and computation—was created in the first half of the previous century by a cohort of brilliant mathematicians and philosophers such as Frege, Hilbert, Gödel, Turing, Tarski, Malcev, Gentzen, and some others. The development of this discipline is arguably among the highest achievements of science in the twentieth century: it expanded mathematics into a novel area of applications, subjected logical reasoning and computability to rigorous analysis, and eventually led to the creation of computers.

The textbook by Professor Wolfgang Rautenberg is a well-written introduction to this beautiful and coherent subject. It contains classical material such as logical calculi, beginnings of model theory, and Gödel's incompleteness theorems, as well as some topics motivated by applications, such as a chapter on logic programming. The author has taken great care to make the exposition readable and concise; each section is accompanied by a good selection of exercises.

A special word of praise is due for the author's presentation of Gödel's second incompleteness theorem in which the author has succeeded in giving an accurate and simple proof of the derivability conditions and the provable Σ_1-completeness, a technically difficult point that is usually omitted in textbooks of comparable level. This textbook can be recommended to all students who want to learn the foundations of mathematical logic.

Preface

This book is based on the second edition of my *Einführung in die Mathematische Logik* whose favorable reception facilitated the preparation of this English version. The book is aimed at students of mathematics, computer science, or linguistics. Because of the epistemological applications of Gödel's incompleteness theorems, this book may also be of interest to students of philosophy with an adequate mathematical background. Although the book is primarily designed to accompany lectures on a graduate level, most of the first three chapters are also readable by undergraduates. These first hundred pages cover sufficient material for an undergraduate course on mathematical logic, combined with a due portion of set theory. Some of the sections of Chapter **3** are partly descriptive, providing a perspective on decision problems, automated theorem proving, nonstandard models, and related topics.

Using this book for independent and individual study depends less on the reader's mathematical background than on his (or her) ambition to master the technical details. Suitable examples accompany the theorems and new notions throughout. To support a private study, the indexes have been prepared carefully. We always try to portray simple things simply and concisely and to avoid excessive notation, which could divert the reader's mind from the essentials. Linebreaks in formulas have been avoided. A special section at the end provides solution hints to most exercises, and complete solutions of exercises that are relevant for the text.

Starting from Chapter **4**, the demands on the reader begin to grow. The challenge can best be met by attempting to solve the exercises without recourse to the hints. The density of information in the text is pretty high; a newcomer may need one hour for one page. Make sure to have paper and pencil at hand when reading the text. Apart from a sufficient training in logical (or mathematical) deduction, additional prerequisites are assumed only for parts of Chapter **5**, namely some knowledge of classical algebra, and at the very end of the last chapter some acquaintance with models of axiomatic set theory.

On top of the material for a one-semester lecture course on mathematical logic, basic material for a course in logic for computer scientists is included in Chapter **4** on logic programming. An effort has been made to capture some of the interesting aspects of this discipline's logical foundations. The resolution theorem is proved constructively. Since all recursive functions are computable in PROLOG, it is not hard to get the undecidability of the existence problem for successful resolutions.

Chapter **5** concerns applications of mathematical logic in various methods of model construction and contains enough material for an introductory course on model theory. It presents in particular a proof of quantifier eliminability in the theory of real closed fields, a basic result with a broad range of applications.

A special aspect of the book is the thorough treatment of Gödel's incompleteness theorems. Since these require a closer look at recursive predicates, Chapter **6** starts with basic recursion theory. One also needs it for solving questions about decidability and undecidability. Defining formulas for arithmetical predicates are classified early, in order to elucidate the close relationship between logic and recursion theory. Along these lines, in **6.4** we obtain in one sweep Gödel's first incompleteness theorem, the undecidability of the tautology problem by Church, and Tarski's result on the nondefinability of truth. Decidability and undecidability are dealt with in **6.5**, and **6.6** includes a sketch of the solution to Hilbert's tenth problem.

Chapter **7** is devoted exclusively to Gödel's second incompleteness theorem and some of its generalizations. Of particular interest thereby is the fact that questions about self-referential arithmetical statements are algorithmically decidable due to Solovay's completeness theorem. Here and elsewhere, Peano arithmetic PA plays a key role, a basic theory for the foundations of mathematics and computer science, introduced already in **3.3**. The chapter includes some of the latest results in the area of self-reference not yet covered by other textbooks.

Remarks in small print refer occasionally to notions that are undefined or will be introduced later, or direct the reader toward the bibliography, which represents an incomplete selection only. It lists most English textbooks on mathematical logic and, in addition, some original papers, mainly for historical reasons. This book contains only material that will remain the subject of lectures in the future. The material is treated in a rather streamlined fashion, which has allowed us to cover many different topics. Nonetheless, the book provides only a selection of results and can at most accentuate certain topics. This concerns above all the Chapters **4**, **5**, **6**, and **7**, which go a step beyond the elementary. Philosophical and foundational problems of mathematics are not systematically discussed within the constraints of this book, but are to some extent considered when appropriate.

The seven chapters of the book consist of numbered sections. A reference like Theorem 5.4 is to mean Theorem 4 in Section **5** of a given chapter. In cross-referencing from another chapter, the chapter number will be adjoined. For instance, Theorem 6.5.4 is Theorem 5.4 in Chapter **6**. You may find additional information about the book or contact me on my website *www.math.fu-berlin.de/∼raut* .

I would like to thank the colleagues who offered me helpful criticism along the way; their names are too numerous to list here. Particularly useful for Chapter **7** were hints from Lev Beklemishev (Moscow) and Wilfried Buchholz (Munich). Thanks also to the publisher, in particular Martin Peters, Mark Spencer, and David Kramer.

Wolfgang Rautenberg
December 2005

Contents

Introduction

Traditional logic as a part of philosophy is one of the oldest scientific disciplines. It can be traced back to the Stoics and to Aristotle.[1] It is one of the roots of what is nowadays called philosophical logic. Mathematical logic, however, is a relatively young discipline, having arisen from the endeavors of Peano, Frege and Russell to reduce mathematics entirely to logic. It steadily developed during the twentieth century into a broad discipline with several subareas and numerous applications in mathematics, computer science, linguistics, and philosophy.

One of the features of modern logic is a clear distinction between object language and metalanguage. The latter is normally a kind of a colloquial language, although it differs from author to author and depends also on the audience the author has in mind. In any case, it is mixed up with semiformal elements, most of which have their origin in set theory. The amount of set theory involved depends on one's objectives. General semantics and model theory use stronger set-theoretical tools than does proof theory. But on average, little more is assumed than knowledge of the most common set-theoretical terminology, presented in almost every mathematical course for beginners. Much of it is used only as a façon de parler.

Since this book concerns mathematical logic, its language is similar to the language common to all mathematical disciplines. There is one essential difference though. In mathematics, metalanguage and object language strongly interact with each other and the latter is semiformalized in the best of cases. This method has proved successful. Separating object language and metalanguage is relevant only in special context, for example in axiomatic set theory, where formalization is needed to specify how certain axioms look like. Strictly formal languages are met more often in computer science. In analysing complex software or a programming language, like in logic, formal linguistic entities are the objects of consideration.

The way of arguing about formal languages and theories is traditionally called the *metatheory*. An important task of a metatheoretical analysis is to specify procedures of logical inference by so-called *logical calculi*, which operate purely syntactical. There are many different logical calculi. The choice may depend on the formalized language, on the logical basis, and on certain aims of the formalization. Basic metatheoretical tools are in any case the naive natural numbers and inductive proof procedures. We will sometimes call them proofs by *metainduction*, in particular when talking about formalized theories that may speak about natural numbers and induction themselves. Induction can likewise be carried out on certain sets of strings over a fixed alphabet, or on the system of rules of a logical calculus.

[1] The Aristotelian syllogisms are useful examples for inferences in a first-order language with unary predicate symbols. One of these serves as an example in Section **4.4** on logic programming.

The logical means of the metatheory are sometimes allowed or even explicitly required to be different from those of the object language. But in this book the logic of object languages, as well as that of the metalanguage, are classical, two-valued logic. There are good reasons to argue that classical logic is the logic of common sense. Mathematicians, computer scientists, linguists, philosophers, physicists, and others are using it as a common platform for communication.

It should be noticed that logic used in the sciences differs essentially from logic used in everyday language, where logic is more an art than a serious task of saying what follows from what. In everyday life, nearly every utterance depends on the context. In most cases logical relations are only alluded to and rarely explicitly expressed. Some basic assumptions of two-valued logic mostly fail, for instance, a context-free use of the logical connectives. Problems of this type are not dealt with in this book. To some extent, many-valued logic or Kripke semantics can help to clarify the situation, and sometimes intrinsic mathematical methods must be used in order to analyze and solve such problems. We shall use Kripke semantics here for a different goal though, the analysis of self-referential sentences in Chapter **7**.

Let us add some historical remarks, which, of course, a newcomer may find easier to understand *after* and not *before* reading at least parts of this book. In the relatively short period of development of modern mathematical logic in the last century, some highlights may be distinguished, of which we mention just a few.

The first was the axiomatization of set theory in various ways. The most important approaches are the ones of Zermelo (improved by Fraenkel and von Neumann) and the theory of types by Whitehead and Russell. The latter was to become the sole remnant of Frege's attempt to reduce mathematics to logic. Instead it turned out that mathematics can be based entirely on set theory as a first-order theory. Actually, this became more salient after the rest of the hidden assumptions by Russell and others were removed from axiomatic set theory[2] around 1915; see [Hej].

Right after these axiomatizations were completed, Skolem discovered that there are countable models of the set-theoretic axioms, a drawback for the hope for an axiomatic definition of a set. Just then, two distinguished mathematicians, Hilbert and Brouwer, entered the scene and started their famous quarrel on the foundations of mathematics. It is described in an excellent manner in [Kl2, Chapter IV] and need therefore not be repeated here.

As a next highlight, Gödel proved the completeness of Hilbert's rules for predicate logic, presented in the first modern textbook on mathematical logic, [HA]. Thus, to some extent, a dream of Leibniz became real, namely to create an *ars inveniendi* for mathematical truth. Meanwhile, Hilbert had developed his view on a foundation of

[2] For instance, the notion of an ordered pair is indeed a set-theoretical and not a logical one.

mathematics into a program. It aimed at proving the consistency of arithmetic and perhaps the whole of mathematics including its nonfinitistic set-theoretic methods by finitary means. But Gödel showed by his incompleteness theorems in 1931 that Hilbert's original program fails or at least needs thorough revision.

Many logicians consider these theorems to be the top highlights of mathematical logic in the twentieth century. A consequence of these theorems is the existence of consistent extensions of Peano arithmetic in which true and false sentences live in peaceful coexistence with each other, called "dream theories" in Section **7.2**. It is an intellectual adventure of holistic beauty to see wisdoms from number theory known for ages, like the Chinese remainder theorem, or simple properties of prime numbers and Euclid's characterization of coprimeness (page 193) unexpectedly assuming pivotal positions within the architecture of Gödel's proofs.

The methods Gödel developed in his paper were also basic for the creation of recursion theory around 1936. Church's proof of the undecidability of the tautology problem marks another distinctive achievement. After having collected sufficient evidence by his own investigations and by those of Turing, Kleene, and some others, Church formulated his famous thesis (Section **6.1**), although in 1936 no computers in the modern sense existed nor was it foreseeable that computability would ever play the basic role it does today.

As already mentioned, Hilbert's program had to be revised. A decisive step was undertaken by Gentzen, considered to be another groundbreaking achievement of mathematical logic and the starting point of contemporary proof theory. The logical calculi in **1.2** and **3.1** are akin to Gentzen's calculi of natural deduction.

We further mention Gödel's discovery that it is not the axiom of choice (AC) that creates the consistency problem in set theory. Set theory with AC and the continuum hypothesis (CH) is consistent provided set theory without AC and CH is. This is a basic result of mathematical logic that would not have been obtained without the use of strictly formal methods. The same applies to the independence proof of AC and CH from the axioms of set theory by P. Cohen in 1963.

The above indicates that mathematical logic is closely connected with the aim of giving mathematics a solid foundation. Nonetheless, we confine ourself to logic and its fascinating interaction with mathematics. History shows that it is impossible to establish a programmatic view on the foundations of mathematics that pleases everybody in the mathematical community. Mathematical logic is the right tool for treating the technical problems of the foundations of mathematics, but it cannot solve its epistemological problems.

Notation

We assume that the reader is familiar with basic mathematical terminology and notation, in particular with the elementary set-theoretical operations of *union*, *intersection*, *complemention*, and *cross product*, denoted by $\cup$, $\cap$, $\setminus$, and $\times$, respectively. Here we summarize only some notation that may differ slightly from author to author, or is specific for this book.

$\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$ denote the sets of natural numbers including 0, integers, rational, and real numbers, respectively. n, m, i, j, k denote always natural numbers unless stated otherwise. Hence, extended notation like $n \in \mathbb{N}$ is mostly omitted. $\mathbb{N}_+, \mathbb{Q}_+, \mathbb{R}_+$ denote the sets of positive members of the corresponding sets.

The *ordered pair* of elements a, b is denoted by (a, b). It should not be mixed up with the *pair set* $\{a, b\}$. Set inclusion is denoted by $M \subseteq N$, while $M \subset N$ means proper inclusion (i.e., $M \subseteq N$ and $M \neq N$). We write $M \subset N$ only if the circumstance $M \neq N$ has to be emphasized. If M is fixed in a consideration and N varies over subsets of M, then $M \setminus N$ may also be denoted by $\setminus N$ or $\neg N$. The *power set* (= set of all subsets) of M is denoted $\mathfrak{P}M$. $\emptyset$ denotes the *empty set*.

If one wants to emphasize that all elements of a set F are sets, F is also called a *family* or *system* of sets. $\bigcup F$ denotes the union of a set family F, that is, the set of elements belonging to at least one $M \in F$, and $\bigcap F$ stands for the intersection of F ($\neq \emptyset$), which is the set of elements belonging to all $M \in F$. If $F = \{M_i \mid i \in I\}$ then $\bigcup F$ and $\bigcap F$ are mostly denoted by $\bigcup_{i \in I} M_i$ and $\bigcap_{i \in I} M_i$, respectively.

A *relation* between M and N is a subset of $M \times N$. Such a relation, call it f, is said to be a *function* (or *mapping*) from M to N if for each $a \in M$ there is precisely one $b \in N$ with $(a, b) \in f$. This b is denoted by $f(a)$ or fa or a^f and called the *value of f at a*. We denote such an f also by $f : M \to N$, or by $f : x \mapsto t(x)$ provided $f(x) = t(x)$ for some term t (terms are defined in **2.2**). $id_M : x \mapsto x$ denotes the *identical function* on M. $\mathit{ran}\, f = \{fx \mid x \in M\}$ is called the *range* of f, while $\mathit{dom}\, f = M$ is called its *domain*. $f : M \to N$ is *injective* if $fx = fy \Rightarrow x = y$, for all $x, y \in M$, *surjective* if $\mathit{ran}\, f = N$, and *bijective* if f is both injective and surjective. The reader should basically be familiar with this terminology.

The set of all functions from M to N is denoted by N^M. The phrase "let f be a function from M to N" is sometimes shortened to "let $f : M \to N$." If f, g are mappings with $\mathit{ran}\, g \subseteq \mathit{dom}\, f$ then $h : x \mapsto f(g(x))$ is called their *composition*. It is sometimes denoted by $h = f \circ g$, but other notation is used as well.

Let I and M be sets, $f : I \to M$, and call I the *index set*. Then f will often be denoted by $(a_i)_{i \in I}$ and is named, depending on the context, a *family*, an *I-tuple*, or a *sequence*. If 0 is identified with $\emptyset$ and $n > 0$ with $\{0, 1, \ldots, n-1\}$, as is common in set theory, then M^n can be understood as the set of finite sequences or

n-tuples $(a_i)_{i<n} = (a_0, \dots, a_{n-1})$ *of length* n whose members are elements of M. In concatenating finite sequences which has an obvious meaning, the *empty sequence* (the only member of $M^0 = \{\emptyset\}$), plays the role of a neutral element. A sequence of the form $(a_1, \dots, a_n)$ will frequently be denoted by $\vec{a}$. This is for $n = 0$ the empty sequence, similar to $\{a_1, \dots, a_n\}$ for $n = 0$ being always the empty set.

If A is an *alphabet*, i.e., if the elements of A are symbols or at least called symbols, then the sequence $(a_1, \dots, a_n)$ is written as $a_1 \cdots a_n$ and called a *string* or a *word* over the alphabet A. The empty sequence is then called the *empty string* or the *empty word*. Let $\xi\eta$ denote the concatenation of the strings ξ and η. If $\xi = \xi_1\eta\xi_2$ for some strings ξ_1, ξ_2 and $\eta \neq \emptyset$ then η is called a *substring* or *segment* of ξ. If, in addition, $\xi_1 = \emptyset$ then η is called an *initial*, and if $\xi_2 = \emptyset$, a *terminal* segment of ξ.

Subsets $P, Q, R, \dots \subseteq M^n$ are called *n-ary predicates of M* or n-ary *relations*. A unary predicate will be identified with the corresponding subset of M. We may write $P\vec{a}$ instead of $\vec{a} \in P$, and $\neg P\vec{a}$ instead of $\vec{a} \notin P$. Metatheoretical predicates (or properties) cast in words will often be distinguished from the surrounding text by single quotes, for instance, if we speak of the syntactic predicate 'The variable x occurs in the formula α'. We can do so since quotes inside quotes will not occur. Single quoted predicates are often used in induction principles, or they are reflected in a theory, while ordinary ("double") quotes have a stylistic function only.

An n-ary *operation of M* is a function $f : M^n \to M$. Almost everywhere $f\vec{a}$ will be written instead of $f(a_1, \dots, a_n)$. Since $M^0 = \{\emptyset\}$, a 0-ary operation of M is of the form $\{(\emptyset, c)\}$ with $c \in M$; it is denoted by c for short and called a *constant*. Each operation $f : M^n \to M$ is uniquely described by the *graph of f*,

$$\operatorname{graph} f := \{(a_1, \dots, a_{n+1}) \in M^{n+1} \mid f(a_1, \dots, a_n) = a_{n+1}\}.$$

Both f and $\operatorname{graph} f$ are essentially the same, but in most situations it is more convenient to distinguish between f and $\operatorname{graph} f$.

If A, B are expressions of our metalanguage, $A \Leftrightarrow B$ stands for "*A iff B*," that is, "*A if and only if B*." Similarly, $A \Rightarrow B$, $A \,\&\, B$, and $A \vee B$ mean "*if A then B*," "*A and B*," and "*A or B*," respectively. This notation does not aim at formalizing the metalanguage but serves improved organization of metatheoretic statements. We agree that $\Rightarrow, \Leftrightarrow, \dots$ separate stronger than linguistic binding particles like "there is" or "for all." Hence, in $T \vDash \alpha \Leftrightarrow \alpha \in T$, for all $\alpha \in \mathcal{L}^0$ (definition page 64) the comma should not be omitted; otherwise some serious misunderstanding may arise, since '$\alpha \in T$ for all $\alpha \in \mathcal{L}^0$' has the meaning 'the theory T is inconsistent'.

$A :\Leftrightarrow B$ means that the expression A is defined by B. Similarly, $s := t$ means that the term s is defined by the term t, or whenever s is a variable, the allocation of the value of t to s. W.l.o.g. or w.l.o.g. abbreviates "Without loss of generality."

Chapter 1

Propositional Logic

Propositional logic, by which we here mean two-valued propositional logic, arises from analyzing connections of given sentences A, B, such as

$$A \text{ and } B, \quad A \text{ or } B, \quad \text{not } A, \quad \text{if } A \text{ then } B.$$

These connection operations can be approximately described by two-valued logic. There are other connections that have temporal or local features, for instance, *first A then B* or *here A there B*, as well as unary modal operators like *it is necessarily true that*, whose analysis goes beyond the scope of two-valued logic. These operators are the subject of temporal, modal, or other subdisciplines of many-valued or nonclassical logic. Furthermore, the connections that we began with may have a meaning in other versions of logic that two-valued logic only incompletely captures. This pertains in particular to their meaning in natural or everyday language, where meaning may strongly depend on context.

In two-valued propositional logic such phenomena are set aside. This approach not only considerably simplifies matters, but has the advantage of presenting many concepts, for instance those of consequence, rule induction, or resolution, on a simpler and more perspicuous level. This will in turn save a lot of writing in Chapter **2** when we consider the corresponding concepts in the framework of predicate logic.

We will not consider everything that would make sense in the framework of two-valued propositional logic, such as two-valued fragments and problems of definability and interpolation. The reader is referred instead to [KK] or [Ra1]. We will concentrate our attention more on propositional calculi. While there exist a multitude of applications of propositional logic, we will not consider technical applications such as the designing of Boolean circuits and problems of optimization. These topics have meanwhile been integrated into computer science. Rather, some useful applications of the propositional compactness theorem are described comprehensively.

1.1 Boolean Functions and Formulas

Two-valued logic is based on two foundational principles: the *principle of bivalence*, which allows only two truth values, namely *true* and *false*, and the *principle of extentionality*, according to which the truth value of a connected sentence depends only on the truth values of its parts, not on their meaning. Clearly, these principles form only an idealization of the actual relationships.

Questions regarding degrees of truth or the sense-content of sentences are ignored in two-valued logic. Despite this simplification, or indeed because of it, such a method is scientifically successful. One does not even have to know exactly what the truth values *true* and *false* actually are. Indeed, in what follows we will identify them with the two symbols 1 and 0. Of course, one could have chosen any other apt symbols such as $\top$ and $\bot$ or `t` and `f`. The advantage here is that all conceivable interpretations of *true* and *false* remain open, including those of a purely technical nature, for instance the two states of a gate in a Boolean circuit.

According to the meaning of the word *and*, the conjunction A *and* B of sentences A, B, in formalized languages written as $A \wedge B$ or $A \,\&\, B$, is true if and only if A, B are both true and is false otherwise. So conjunction corresponds to a binary function or operation over the set $\{0,1\}$ of truth values, named the *$\wedge$-function* and denoted by $\wedge$. It is given by its *value matrix* $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, where, in general, $\begin{pmatrix} 1\circ 1 & 1\circ 0 \\ 0\circ 1 & 0\circ 0 \end{pmatrix}$ represents the value matrix or *truth table* of a binary function $\circ$ with arguments and values in $\{0,1\}$. The delimiters of these small matrices will usually be omitted.

A function $f\colon\{0,1\}^n \to \{0,1\}$ is called an n-ary *Boolean* or *truth function*. Since there are 2^n n-tuples of $0, 1$, it is easy to see that the number of n-ary Boolean functions is 2^{2^n}. We denote their totality by $\boldsymbol{B}_n$. While $\boldsymbol{B}_2$ has $2^4 = 16$ members, there are only four unary Boolean functions. One of these is *negation*, denoted by $\neg$ and defined by $\neg 1 = 0$ and $\neg 0 = 1$. $\boldsymbol{B}_0$ consists just of the constants 0 and 1.

The first column of the table on the opposite page contains the common binary connections with examples of their instantiation in English. The second column lists some of its traditional symbols, which also denote the corresponding truth function, and the third its truth table. *Disjunction* is the *inclusive or* and is to be distinguished from the *exclusive disjunction*. The latter corresponds to addition modulo 2 and got therefore the symbol $+$. In Boolean circuits the functions $+, \downarrow, \uparrow$ are often denoted by *xor*, *nor*, and *nand*; the latter is also known as the *Sheffer function*.

A connected sentence and its corresponding truth function need not be denoted by the same symbol; for example one might take $\wedge$ for conjunction and *et* as the truth function. But in doing so one would only be creating extra notation, but no new insights. The meaning of a symbol will always be clear from the context: if α, β

are sentences of a formal language, then $\alpha \wedge \beta$ denotes their conjunction. If on the other hand, a, b are truth values, then $a \wedge b$ just denotes a truth value. Occasionally, we may want to refer to the symbols $\wedge$, $\vee$, $\neg, \ldots$ themselves, setting their meaning temporarily aside. Then we talk of the *connectives* or *truth functors* $\wedge$, $\vee$, $\neg, \ldots$

compound sentence	**symbol**	**truth table**
conjunction *A and B*; *A as well as B*	$\wedge$, &	1 0 0 0
disjunction *A or B*	$\vee$, $\vee$	1 1 1 0
implication *if A then B; B if A*	$\rightarrow$, $\Rightarrow$	1 0 1 1
equivalence (biconditional) *A if and only if B; A iff B*	$\leftrightarrow$, $\Leftrightarrow$	1 0 0 1
exclusive disjunction *either A or B but not both*	$+$	0 1 1 0
nihilition *neither A nor B*	$\downarrow$	0 0 0 1
incompatibility *not at once A and B*	$\uparrow$	0 1 1 1

Sentences formed using connectives given in the table are said to be logically equivalent if their corresponding truth tables are identical. This is the case, for example, for the sentences

$$A \text{ if } B, \quad A \text{ or not } B, \quad B \text{ only if } A,$$

which represent the *converse implication*, denoted by $A \leftarrow B$. It does not appear in the table since it arises by swapping A, B in the implication. This and similar reasons explain why only a few of the sixteen binary Boolean functions require notation. Amazingly, converse implication is used in the programming language PROLOG, dealt with in **4.4**. Recall our agreement in the section *Notation* that the symbols &, $\vee$, $\Rightarrow$, and $\Leftrightarrow$ will be used only on the metatheoretic level.

In order to recognize and describe logical equivalence of compound sentences it is useful to create a suitable formalism or a formal language. The idea is basically the same as in arithmetic, where general statements are more clearly expressed by means of certain formulas. As with arithmetical terms, we consider propositional formulas as strings of signs built in given ways from basic symbols. Among these basic symbols are variables, for our purposes called *propositional variables*, the set of which is denoted by *PV*. Traditionally, these signs are symbolized by $p_0, p_1, \ldots$ However, our numbering of the variables below begins with p_1 rather than with p_0,

enabling us later on to represent Boolean functions more conveniently. Further, we use certain logical signs such as $\wedge$, $\vee$, $\neg, \ldots$, similar to the signs $+, \cdot, \ldots$ of arithmetic. Finally, the parentheses (,) will serve as technical aids, although these two symbols are dispensable as will be seen later on.

Each time a propositional language is in question, the set of its logical symbols, called the *logical signature*, and the set of its variables must be given in advance. For instance, it is crucial in some applications of propositional logic in Section **1.5**, for PV to be an arbitrary set, and not a countably infinite one as indicated previously. Put concretely, we define a propositional language $\mathfrak{F}$ built up from the symbols $(\,,\,)\,,\wedge\,,\vee\,,\neg\,,p_1,p_2,\ldots$ inductively as follows:

(F1) The one-element strings $p_1, p_2, \ldots$ are formulas, called *prime formulas*.

(F2) If the strings α, β are formulas, then so too are $(\alpha \wedge \beta)$, $(\alpha \vee \beta)$, and $\neg\alpha$.

This is an inductive definition in the set of strings on the alphabet of the mentioned symbols, that is, only those strings gained using (F1) or (F2) are in this context formulas. Stated set-theoretically, $\mathfrak{F}$ is the smallest (that is, the intersection) of all sets of strings S built from the aforementioned symbols with the properties

$$\text{(f1) } p_1, p_2, \cdots \in S, \quad \text{(f2) } \alpha, \beta \in S \;\Rightarrow\; (\alpha \wedge \beta), (\alpha \vee \beta), \neg\alpha \in S.^{1}$$

Example. $(p_1 \wedge (p_2 \vee \neg p_1))$ is a formula. On the other hand, its initial segment $(p_1 \wedge (p_2 \vee \neg p_1)$ is not, because a closing parenthesis is missing. It is intuitively clear and will later be rigorously proved, that the number of left parentheses occurring in a formula coincides with the number of its right parentheses. Every proper initial segment of the example formula obviously fails to meet this condition.

The formulas so defined are called *Boolean formulas*, because they are obtained using the *Boolean signature* $\{\wedge\,,\vee\,,\neg\}$. It should be noticed that in the definition parentheses are needed only for binary connectives, not if a formula starts with the unary operator $\neg$. Should further connectives belong to the logical signature, for example $\rightarrow$ or $\leftrightarrow$, (F2) of the above definition must be augmented accordingly. But unless stated otherwise, $\alpha \rightarrow \beta$ and $\alpha \leftrightarrow \beta$ are here just abbreviations; namely $\alpha \rightarrow \beta := \neg(\alpha \wedge \neg\beta)$ and $\alpha \leftrightarrow \beta := (\alpha \rightarrow \beta) \wedge (\beta \rightarrow \alpha)$.

Occasionally, it is useful to have symbols in the logical signature for always true and always false, $\bot$ and $\top$ respectively, say, called *falsum* and *verum* and sometimes also denoted by 0 and 1. These are to be regarded as supplementary prime formulas, and clause (F1) should be altered accordingly. In the Boolean signature, $\bot$ and $\top$ are used as abbreviations for the formulas $p_1 \wedge \neg p_1$ and $p_1 \vee \neg p_1$, respectively.

[1] This is a set-theoretical translation of the above inductive definition. Some authors like to add a third condition to (F1), (F2), namely (F3): *No other strings than those obtained by* (F1) *and* (F2) *are formulas in this context.* But this at most underlines that (F1),(F2) are the only formula building rules; (F3) *follows* from our definition as its set-theoretical translation shows.

For the time being we let $\mathcal{F}$ be the set of all Boolean formulas, although in what follows, everything said about $\mathcal{F}$ holds correspondingly for any propositional language. Propositional variables will henceforth be denoted by the letters $p, q, \ldots$, formulas by $\alpha, \beta, \gamma, \delta, \varphi, \ldots$, prime formulas also by π, and sets of formulas by X, Y, Z, where these letters may also be indexed.

In order not to have to write down too many parentheses in formulas, we set some conventions similar to those used in writing arithmetical terms.

1. The outermost parentheses in a formula may be omitted (if there are any). For example, $(p \vee q) \wedge \neg p$ may be written in place of $((p \vee q) \wedge \neg p)$. Note that $(p \vee q) \wedge \neg p$ is not itself a formula but *denotes* the formula $((p \vee q) \wedge \neg p)$.

2. In the order $\neg, \wedge, \vee, \rightarrow, \leftrightarrow$, each connective binds more strongly than those following it. Thus, one may even write $p \vee q \wedge \neg p$ instead of $(p \vee (q \wedge \neg p))$.

3. By the multiple use of $\rightarrow$ we *associate to the right*. So $p \rightarrow q \rightarrow p$ is to mean $p \rightarrow (q \rightarrow p)$. Multiple occurrences of other binary connectives are associated to the left, for instance, $p \wedge q \wedge \neg p$ means $(p \wedge q) \wedge \neg p$. In place of $\alpha_0 \wedge \cdots \wedge \alpha_n$ and $\alpha_0 \vee \cdots \vee \alpha_n$ we may write $\bigwedge_{i \leqslant n} \alpha_i$ and $\bigvee_{i \leqslant n} \alpha_i$, respectively.

Also, in arithmetic, one normally associates to the left. An exception is x^{y^z} where traditionally association to the right is used, (that is, x^{y^z} equals $x^{(y^z)}$). Association to the right has some advantages in the writing of tautologies in which $\rightarrow$ plays a main role; for instance, in the logical axioms listed in **1.3**.

The above conventions are based on a reliable syntax in the framework of which intuitively clear facts, such as the identical number of left and right parentheses in a formula, are rigorously provable. These proofs are generally carried out using induction on the construction of a formula. To make this clearer we denote by $\mathcal{E}\varphi$ that a property $\mathcal{E}$ holds for a string φ. For example, let $\mathcal{E}$ mean the property 'φ is a formula with equally many right- and left-hand parentheses'. Obviously, $\mathcal{E}$ is valid for prime formulas, and if $\mathcal{E}\alpha$, $\mathcal{E}\beta$ then clearly also $\mathcal{E}(\alpha \wedge \beta)$, $\mathcal{E}(\alpha \vee \beta)$, and $\mathcal{E}\neg\alpha$. From this one may conclude that $\mathcal{E}$ applies to all formulas, our reasoning being a particularly simple instance of the following

Induction principle for formulas. *Let $\mathcal{E}$ be a property of strings such that*

(o) $\mathcal{E}\pi$ *for all prime formulas* π,

(s) $\mathcal{E}\alpha, \mathcal{E}\beta \Rightarrow \mathcal{E}(\alpha \wedge \beta), \mathcal{E}(\alpha \vee \beta), \mathcal{E}\neg\alpha$, *for all* $\alpha, \beta \in \mathcal{F}$.

Then $\mathcal{E}\varphi$ holds for all formulas φ.

The justification of this principle is straightforward. The set S of all strings with the property $\mathcal{E}$ satisfies, thanks to (o) and (s), the conditions (f1) and (f2) of page 4. But $\mathcal{F}$ is the smallest such set. Consequently, $\mathcal{F} \subseteq S$. In other words, $\mathcal{E}$ applies to all formulas φ.

It is intuitively clear that a compound formula φ (i.e., φ is not a prime formula) can be decomposed uniquely. For instance, a formula $\alpha \wedge \beta$ (outer parentheses omitted) cannot at the same time be written $\alpha' \vee \beta'$ with perhaps different formulas α', β'. Speaking more generally, compound formulas have the following basic property the proof of which is not as trivial as might be expected. Nonetheless, it is left as an exercise (Exercise 4) in order to maintain the flow of things.

Unique reconstruction property. *Each compound formula $\varphi \in \mathcal{F}$ is of the form $\neg\alpha$ or $(\alpha \circ \beta)$, where $\alpha, \beta \in \mathcal{F}$ and $\circ \in \{\wedge, \vee\}$ are uniquely determined by φ.*

It may be a surprise to the novice that for a unique reconstruction, parentheses are dispensable throughout. Indeed, propositional formulas, like arithmetical terms, can be written without any parentheses; this is realized in *Polish Notation* (= PN), also called *prefix notation*, once widely used in the logic literature. The idea consists in altering (F2) as follows: *if α, β are formulas then so too are $\wedge\alpha\beta$, $\vee\alpha\beta$, and $\neg\alpha$.*

Remark 1. Similar to PN is RPN (*Reverse Polish Notation*). It is used in some programming languages. RPN differs from PN only in that the connectives are placed *after* the arguments. For instance, $(p \wedge (q \vee \neg p))$ is written in RPN as $pqp\neg\vee\wedge$. Reading PN or RPN requires more effort due to the high density of information; but by the same token it can be processed very fast by a computer or a (high-tech) printer which gets its jobs as RPN-based PostScript-programs. The only advantage of the parenthesized version is that optical decoding is somewhat easier through the dilution of information.

Intuitively it is clear what a *subformula* of a formula φ is; for example, $(q \wedge \neg p)$ is a subformula of $(p \vee (q \wedge \neg p))$. All the same, for some purposes it is convenient to characterize the set $\operatorname{Sf}\varphi$ of all subformulas of φ inductively:

$$\operatorname{Sf}\pi = \{\pi\} \text{ for prime formulas } \pi; \quad \operatorname{Sf}\neg\alpha = \operatorname{Sf}\alpha \cup \{\neg\alpha\},$$
$$\operatorname{Sf}(\alpha \circ \beta) = \operatorname{Sf}\alpha \cup \operatorname{Sf}\beta \cup \{(\alpha \circ \beta)\} \text{ for a binary connective } \circ.$$

Thus, a formula is always regarded as a subformula of itself. The above is a typical example of a *recursive definition on the construction of formulas*. Another example of such a definition is the *rank* of a formula φ, $\operatorname{rk}\varphi$, which provides a sometimes more convenient measure of the complexity of φ than its length as a string and occasionally simplifies inductive arguments. Intuitively, $\operatorname{rk}\varphi$ is the highest number of nested pairs of parentheses or nested negation signs occurring in a formula φ. Let $\operatorname{rk}\pi = 0$ for prime formulas π, and if $\operatorname{rk}\alpha$ and $\operatorname{rk}\beta$ are given, then

$$\operatorname{rk}\neg\alpha = \operatorname{rk}\alpha + 1, \quad \operatorname{rk}(\alpha \circ \beta) = \max\{\operatorname{rk}\alpha, \operatorname{rk}\beta\} + 1 \text{ for a binary connective } \circ.$$

We will not give a general formulation of this definition procedure because it is so intuitive, and has been made sufficiently clear by the preceding examples. Its justification is essentially based on the unique reconstruction property, in contrast to justifying proofs by induction on formulas that immediately derive from the definition of formulas. The theoretical background of all this is that $\mathcal{F}$ forms an absolutely free algebra; see for instance [RS] for details.

If a property is to be proved by induction on the construction of formulas φ, we will say that it is a *proof by induction on* φ. Similarly, the recursive construction of a function f on $\mathcal{F}$ will generally be referred to as *defining* f *recursively on* φ, often not quite correctly paraphrased as *defining* f *by induction*. rk is an example.

Since the truth value of a connected sentence depends only on the truth values of its constituent parts, we may assign to every propositional variable of α a truth value rather than a sentence, thereby evaluating α, i.e., calculating a truth value. Similarly, terms are evaluated in, say, the arithmetic of real numbers, whose value is then a real (= real number). An arithmetical term t in the variables $x_1, \dots, x_n$ describes an n-ary function whose arguments and values are reals, while a formula φ in $p_1, \dots, p_n$ describes an n-ary Boolean function.

To be more precise, a propositional *valuation*, or alternatively a *realization* or (propositional) *model*, is a mapping $w : PV \to \{0, 1\}$. We can extend this to a mapping from the whole of $\mathcal{F}$ to $\{0, 1\}$ (also denoted by w) according to the stipulations

$$(*) \quad w(\alpha \wedge \beta) = w\alpha \wedge w\beta; \quad w(\alpha \vee \beta) = w\alpha \vee w\beta; \quad w\neg\alpha = \neg w\alpha.^{2}$$

By *the value* $w\varphi$ *of a formula* φ *under the valuation of variables* we mean the value given by this extension. We could denote the extended mapping by $\hat{w}$, say, but it is in fact not necessary to distinguish it symbolically from $w : PV \to \{0, 1\}$. Similarly, we keep the same symbol if an operation in $\mathbb{N}$ is extended to a larger domain. If the logical signature contains further connectives, for example $\to$, the conditions $(*)$ must be supplemented accordingly, with $w(\alpha \to \beta) = w\alpha \to w\beta$ in our example. However, if $\to$ is defined as in the Boolean case, then this equation must be provable. Indeed, it is provable, because from our definition of $\alpha \to \beta$ we get $w(\alpha \to \beta) = w\neg(\alpha \wedge \neg\beta) = \neg w(\alpha \wedge \neg\beta) = \neg(w\alpha \wedge \neg w\beta) = w\alpha \to w\beta$, for every w. A corresponding remark could be made with respect to $\leftrightarrow$. Similarly, always $w\top = 1$ and $w\bot = 0$ by our definition of $\top, \bot$, in accordance with the meaning of these symbols. However, if these or corresponding symbols belong to the logical signature, then the last two equations must be added to the definition of w.

Let $\mathcal{F}_n$ denote the set of all formulas of $\mathcal{F}$ in which at most the variables $p_1, \dots, p_n$ occur, $n > 0$. Then it can easily be seen that for $\alpha \in \mathcal{F}_n$, $w\alpha$ depends only on the truth values of the variables $p_1, \dots, p_n$. In other words,

$$(\star) \quad w\alpha = w'\alpha \text{ whenever } wp_i = w'p_i \text{ for } i = 1, \dots, n.$$

The simple proof follows from induction on the construction of formulas in $\mathcal{F}_n$: the property $(\star)$ holds for $p \in \mathcal{F}_n$, and if $(\star)$ is valid for $\alpha, \beta \in \mathcal{F}_n$, then also for $\neg\alpha$, $\alpha \wedge \beta$, and $\alpha \vee \beta$. It is then intuitively clear that a given $\varphi \in \mathcal{F}_n$ defines or represents an n-ary Boolean function according to the following definition.

[2] We often use $(*)$ or $(\star)$ as a temporary label for a condition (or property) that we refer back to in the text following the labeled condition.

Definition. A formula $\alpha \in \mathcal{F}_n$ *represents* the n-ary Boolean function f (or f is represented by α) if $w\alpha = fw\vec{p}$ for all valuations w, where $w\vec{p} := (wp_1, \ldots, wp_n)$.

Because $w\alpha$ for $\alpha \in \mathcal{F}_n$ is uniquely determined by $wp_1, \ldots, wp_n$, α represents precisely one function $f \in \boldsymbol{B}_n$, sometimes written as $\alpha^{(n)}$. For instance, both $p_1 \wedge p_2$ and $\neg(\neg p_1 \vee \neg p_2)$ represent the $\wedge$-function, as can easily be illustrated using a table. Similarly, $\neg p_1 \vee p_2$ and $\neg(p_1 \wedge \neg p_2)$ represent the $\rightarrow$-function, and $p_1 \vee p_2$, $\neg(\neg p_1 \wedge \neg p_2)$, $(p_1 \rightarrow p_2) \rightarrow p_2$ all represent the $\vee$-function. Incidentally, the last formula shows that the $\vee$-connective can be expressed using implication alone.

There is a caveat though: since $\alpha = p_1 \vee p_2$, say, belongs not only to $\mathcal{F}_2$ but to $\mathcal{F}_3$ as well, α also represents the Boolean function $f\colon (x_1, x_2, x_3) \mapsto x_1 \vee x_2$. However, the third argument is only "fictional," or put another way, the function f is not essentially ternary.

In general we say that an operation $f\colon M^n \to M$ is *essentially n-ary* if f has no fictional arguments, where the ith argument of f is called *fictional* whenever

$$f(x_1, \ldots, x_i, \ldots, x_n) = f(x_1, \ldots, x_i', \ldots, x_n),$$

for all $x_1, \ldots, x_i, \ldots, x_n, x_i' \in M$. Identity and the $\neg$-function are the essentially unary Boolean functions, and out of the sixteen binary functions, only ten are essentially binary, as is seen in scrutinizing the possible truth tables.

Remark 2. If a_n denotes temporarily the number of all n-ary Boolean functions and e_n the number of all essentially n-ary Boolean functions, it is not particularly difficult to prove that $a_n = \sum_{i \leqslant n} \binom{n}{i} e_i$. Solving for e_n results in $e_n = \sum_{i \leqslant n} (-1)^{n-i} \binom{n}{i} a_i$. However, we will not make use of these equations, which become important only in a more specialized study of Boolean functions; see any good textbook on discrete mathematics.

Exercises

1. $f \in \boldsymbol{B}_n$ is called *linear* if $f(x_1, \ldots, x_n) = a_0 + a_1 x_1 + \cdots + a_n x_n$ for suitable coefficients $a_0, \ldots, a_n \in \{0, 1\}$. Here $+$ denotes exclusive disjunction (addition modulo 2) and the not written multiplication is conjunction ($a_i x_i = x_i$ for $a_i = 1$ and $a_i x_i = 0$ for $a_i = 0$). (a) Show that the above representation of f is unique, (b) Determine the number of n-ary linear Boolean functions, (c) Prove that each formula α in $\neg, +$ (that is, α is a formula of the logical signature $\{\neg, +\}$) represents a linear Boolean function.

2. Show that a compound Boolean formula φ is of the form $\varphi = \neg\alpha$ or $\varphi = (\alpha \wedge \beta)$ or $\varphi = (\alpha \vee \beta)$ for some $\alpha, \beta \in \mathcal{F}$. Hence, if ξ is any string over the alphabet of $\mathcal{F}$ then $\neg\xi \in \mathcal{F} \Leftrightarrow \xi \in \mathcal{F}$. Similarly, $(\xi_1 \wedge \xi_2) \in \mathcal{F} \Leftrightarrow \xi_1, \xi_2 \in \mathcal{F}$, etc.

3. Prove that a proper initial segment of a formula φ is never a formula.

4. Prove (with Exercise 2 and 3) the unique reconstruction property.

1.2 Semantic Equivalence and Normal Forms

Throughout this chapter w will always denote a propositional valuation. Formulas α, β are called (logically or semantically) *equivalent*, and we write $\alpha \equiv \beta$, when $w\alpha = w\beta$ for all valuations w. For example $\alpha \equiv \neg\neg\alpha$. Obviously, $\alpha \equiv \beta$ iff for any n such that $\alpha, \beta \in \mathcal{F}_n$, both formulas represent the same n-ary Boolean function. It follows that at most 2^{2^n} formulas in $\mathcal{F}_n$ can be pairwise inequivalent, since there are no more than 2^{2^n} n-ary Boolean functions.

In arithmetics one writes simply $s = t$ to express the fact that the terms s, t represent the same function. For example, $(x + y)^2 = x^2 + 2xy + y^2$ expresses the equalitiy of values of the left- and right-hand terms for all values of x, y. This way of writing is permissible because formal syntax plays a minor role in arithmetics. In formal logic, however, as is always the case when syntactic considerations are to the fore, one uses the equality sign in $\alpha = \beta$ only for the syntactic identity of the strings α and β. Therefore, the equivalence of formulas must be denoted differently. Clearly, for all formulas α, β, γ the following equivalences hold:

$$\begin{array}{rcl@{\qquad}rcl@{\qquad}l}
\alpha\wedge(\beta\wedge\gamma) & \equiv & \alpha\wedge\beta\wedge\gamma, & \alpha\vee(\beta\vee\gamma) & \equiv & \alpha\vee\beta\vee\gamma & \text{(associativity)};\\
\alpha\wedge\beta & \equiv & \beta\wedge\alpha, & \alpha\vee\beta & \equiv & \beta\vee\alpha & \text{(commutativity)};\\
\alpha\wedge\alpha & \equiv & \alpha, & \alpha\vee\alpha & \equiv & \alpha & \text{(idempotency)};\\
\alpha\wedge(\alpha\vee\beta) & \equiv & \alpha, & \alpha\vee\alpha\wedge\beta & \equiv & \alpha & \text{(absorption)};\\
\alpha\wedge(\beta\vee\gamma) & \equiv & \alpha\wedge\beta\vee\alpha\wedge\gamma, & \alpha\vee\beta\wedge\gamma & \equiv & (\alpha\vee\beta)\wedge(\alpha\vee\gamma) & \text{(distributivity)};\\
\neg(\alpha\wedge\beta) & \equiv & \neg\alpha\vee\neg\beta, & \neg(\alpha\vee\beta) & \equiv & \neg\alpha\wedge\neg\beta & \text{(de Morgan rules)}.
\end{array}$$

Furthermore, $\alpha \vee \neg\alpha \equiv \top$, $\alpha \wedge \neg\alpha \equiv \bot$, and $\alpha \wedge \top \equiv \alpha \vee \bot \equiv \alpha$. It is also useful to list certain equivalences for formulas containing $\to$, for example the frequently used

$$\alpha \to \beta \equiv \neg\alpha \vee \beta; \quad \alpha \to \beta \to \gamma \equiv \alpha \wedge \beta \to \gamma \equiv \beta \to \alpha \to \gamma.$$

To generalize: $\alpha_1 \to \cdots \to \alpha_n \equiv \alpha_1 \wedge \cdots \wedge \alpha_{n-1} \to \alpha_n$. Further, we mention the "left distributivity" of implication with respect to $\wedge$ and $\vee$, namely

$$\alpha \to \beta\wedge\gamma \equiv (\alpha \to \beta)\wedge(\alpha \to \gamma); \quad \alpha \to \beta \vee \gamma \equiv (\alpha \to \beta) \vee (\alpha \to \gamma).$$

Should the symbol $\to$ lie to the right, then the following are valid:

$$\alpha\wedge\beta \to \gamma \equiv (\alpha \to \gamma) \vee (\beta \to \gamma); \quad \alpha \vee \beta \to \gamma \equiv (\alpha \to \gamma)\wedge(\beta \to \gamma).$$

Remark 1. These last two equivalences are responsible for a curious phenomenon in everyday language. For example, the two sentences

A: *Students and pensioners pay half price*, B: *Students or pensioners pay half price*

evidently have the same meaning. How do we explain this? Let the subjects *student* and *pensioner* be abbreviated by S, P, respectively, and *pay half price* by H. Then

$$\alpha: \quad (S \to H)\wedge(P \to H), \qquad \beta: \quad (S \vee P) \to H$$

express somewhat more precisely the factual content of A and B, respectively. Now, according to our truth tables, the formulas α and β are simply logically equivalent. The

everyday-language statements A and B of α and β obscure the structural difference of α and β through an apparently synonymous use of *and* and *or*.

Obviously, $\equiv$ is an equivalence relation, that is,

$$\begin{aligned} \alpha &\equiv \alpha && \text{(reflexivity)},\\ \alpha \equiv \beta \;&\Rightarrow\; \beta \equiv \alpha && \text{(symmetry)},\\ \alpha \equiv \beta, \beta \equiv \gamma \;&\Rightarrow\; \alpha \equiv \gamma && \text{(transitivity)}. \end{aligned}$$

Moreover, $\equiv$ is a *congruence relation*[3] on $\mathcal{F}$. This is to mean that for all $\alpha, \alpha', \beta, \beta'$,

$$\alpha \equiv \alpha', \beta \equiv \beta' \;\Rightarrow\; \alpha \circ \beta \equiv \alpha' \circ \beta', \neg\alpha \equiv \neg\alpha' \qquad (\circ \in \{\wedge, \vee\}).$$

For this reason the so-called *replacement theorem* holds: $\alpha \equiv \alpha' \;\Rightarrow\; \varphi \equiv \varphi'$, where φ' is obtained from φ by replacing one or several of the possible occurrences of the subformula α in φ by α'. For instance, by replacing the subformula $\neg p \vee \neg q$ by the equivalent formula $\neg(p \wedge q)$ in $\varphi = (\neg p \vee \neg q) \wedge (p \vee q)$ we obtain $\varphi' = \neg(p \wedge q) \wedge (p \vee q)$, which is equivalent to φ. A similar replacement theorem also holds for arithmetical terms and is constantly used in their manipulation. This procedure mostly goes unnoticed, because $=$ is written instead of $\equiv$, and the replacement is, consciously or not, usually correctly applied. The simple inductive proof of the replacement theorem will be given in a somewhat broader context in **2.4**.

Furnished with the equivalences $\neg(\alpha \wedge \beta) \equiv \neg\alpha \vee \neg\beta$, $\neg(\alpha \vee \beta) \equiv \neg\alpha \wedge \neg\beta$ and $\neg\neg\alpha \equiv \alpha$, and using the replacement theorem, it is easy to construct for each formula φ an equivalent formula in which the negation sign stands only immediately in front of variables. For example, $\neg(p \wedge q \vee r) \equiv \neg(p \wedge q) \wedge \neg r \equiv (\neg p \vee \neg q) \wedge \neg r$ is obtained in this way. Such manipulations lead also purely syntactically to conjunctive and disjunctive normal forms, considered below.

It is always something of a surprise to the newcomer that independent of its arity, *every* Boolean function can be represented by a Boolean formula. While this can be proved in various ways, we take the opportunity to introduce certain normal forms and therefore begin with the following

Definition. Prime formulas and negations of prime formulas are called *literals*. A disjunction $\alpha_1 \vee \cdots \vee \alpha_n$, where each α_i is a conjunction of literals, is called a *disjunctive normal form*, a DNF for short (also called an alternative normal form). A conjunction $\beta_1 \wedge \cdots \wedge \beta_n$, where every β_i is a disjunction of literals, is called a *conjunctive normal form*, a CNF for short.

Example 1. The formula $p \vee (q \wedge \neg p)$ is a DNF; $p \vee q$ is at once a DNF and a CNF; $p \vee \neg(q \wedge \neg p)$ is neither a DNF nor a CNF.

[3] This concept, stemming originally from geometry, is meaningfully defined in every algebraic structure and is one of the most important mathematical concepts; see **2.1**. The definition is equivalent to the condition $\alpha \equiv \alpha' \;\Rightarrow\; \alpha \circ \beta \equiv \alpha' \circ \beta, \beta \circ \alpha \equiv \beta \circ \alpha', \neg\alpha \equiv \neg\alpha'$, for all α, α', β.

Theorem 2.1 states that every Boolean function is represented by a Boolean formula, indeed by a DNF, and also by a CNF. It would suffice to show that for given n there are at least 2^{2^n} pairwise inequivalent DNFs (resp. CNFs). However, we present instead a constructive proof whereby for a Boolean function given in tabular form a representing DNF (resp. CNF) can explicitly be written down. In the formulation of Theorem 2.1 we temporarily use the following notation: $p^1 := p$ and $p^0 := \neg p$. With this stipulation, $w(p_1^{x_1} \wedge p_2^{x_2}) = 1$ iff $wp_1 = x_1$ and $wp_2 = x_2$. More generally, induction on $n \geqslant 1$ easily shows that for all $x_1, \ldots, x_n \in \{0,1\}$,

$$(*) \quad w(p_1^{x_1} \wedge \cdots \wedge p_n^{x_n}) = 1 \;\Leftrightarrow\; w\vec{p} = \vec{x} \;\; \text{(i.e., } wp_1 = x_1, \ldots, wp_n = x_n\text{)}.$$

Theorem 2.1. *Every Boolean function f with $f \in \boldsymbol{B}_n$ $(n > 0)$ is representable by a* DNF*, namely by*

$$\alpha_f := \bigvee_{f\vec{x}=1} p_1^{x_1} \wedge \cdots \wedge p_n^{x_n}.^4$$

At the same time, f is representable by the CNF

$$\beta_f := \bigwedge_{f\vec{x}=0} p_1^{\neg x_1} \vee \cdots \vee p_n^{\neg x_n}.$$

Proof. By the definition of α_f, the following holds for an arbitrary valuation w:

$$\begin{aligned} w\alpha_f = 1 \;&\Leftrightarrow\; \text{there is an } \vec{x} \text{ with } f\vec{x} = 1 \text{ and } w(p_1^{x_1} \wedge \cdots \wedge p_n^{x_n}) = 1 \\ &\Leftrightarrow\; \text{there is an } \vec{x} \text{ with } f\vec{x} = 1 \text{ and } w\vec{p} = \vec{x} \quad \big(\text{by } (*)\big) \\ &\Leftrightarrow\; fw\vec{p} = 1 \quad (\text{replace } \vec{x} \text{ by } w\vec{p}). \end{aligned}$$

Thus, $w\alpha_f = 1 \Leftrightarrow fw\vec{p} = 1$. From this equivalence, and because there are only two truth values, $w\alpha_f = fw\vec{p}$ follows immediately. The representability proof of f by β_f runs analogously; alternatively, Theorem 2.3 below may be used. ❑

Example 2. For the *exclusive-or function* $+$, the construction procedure of Theorem 2.1 gives the representing DNF $p_1 \wedge \neg p_2 \vee \neg p_1 \wedge p_2$, because $(1,0), (0,1)$ are the only pairs for which $+$ has the value 1. The CNF given by the theorem, on the other hand, is $(p_1 \vee p_2) \wedge (\neg p_1 \vee \neg p_2)$; the equivalent formula $(p_1 \vee p_2) \wedge \neg(p_1 \wedge p_2)$ makes the meaning of the exclusive-or compound particularly intuitive.

The DNF for the Boolean function $\rightarrow$ given by Theorem 2.1 is

$$p_1 \wedge p_2 \vee \neg p_1 \wedge p_2 \vee \neg p_1 \wedge \neg p_2.$$

It is longer than the formula $\neg p_1 \vee p_2$, which is also a representing DNF. But the former is distinctive in that each of its disjuncts contains each variable occurring in

[4] The disjuncts of α_f can be ordered, for instance according to the lexicographical order of the n-tuples $(x_1, \ldots, x_n) \in \{0,1\}^n$. If the disjunction is empty, in other words, if f does not take the value 1, define α_f to be $\bot$ $(= p_1 \wedge \neg p_1)$; similarly set the empty conjunction as $\top$ $(= \neg\bot)$. These conventions correspond to those in arithmetic, where the empty sum has the value 0 and the empty product the value 1.

the formula exactly once. A DNF of n variables with the analogous property is called *canonical*. The notion of canonical CNF is correspondingly explained. For instance, the $\leftrightarrow$-function is represented by the canonical CNF $(\neg p_1 \vee p_2) \wedge (p_1 \vee \neg p_2)$ according to Theorem 2.1. As a matter of fact, this theorem always provides canonical normal forms as representing formulas.

Functional completeness. A logical signature is called *functional complete* if every Boolean function is representable by a formula in this signature. Theorem 2.1 shows that $\{\neg, \wedge, \vee\}$ is functional complete. Because of $p \vee q \equiv \neg(\neg p \wedge \neg q)$ and $p \wedge q \equiv \neg(\neg p \vee \neg q)$, one can further leave aside $\vee$, or alternatively $\wedge$. This observation is the content of

Corollary 2.2. *Both* $\{\neg, \wedge\}$ *and* $\{\neg, \vee\}$ *are functional complete.*

Therefore, to show that a logical signature L is functional complete, it is enough to represent $\neg$, $\wedge$ or else $\neg$, $\vee$ by formulas in L. For example, because $\neg p \equiv p \to 0$ and $p \wedge q \equiv \neg(p \to \neg q)$, the signature $\{\to, 0\}$ is functional complete. On the other hand, $\{\to, \wedge, \vee\}$, and a fortiori $\{\to\}$, are not. Indeed, $w\varphi = 1$ for any formula φ in $\to, \wedge, \vee$ and any valuation w such that $wp = 1$ for all p. This can readily be confirmed by induction on φ. Thus, never $\neg p \equiv \varphi$ for any such formula φ.

It is noteworthy that the signature containing only $\downarrow$ is functional complete: from the truth table for $\downarrow$ we get $\neg p \equiv p \downarrow p$ as well as $p \wedge q \equiv \neg p \downarrow \neg q$. Likewise for $\{\uparrow\}$, because $\neg p \equiv p \uparrow p$ and $p \vee q \equiv \neg p \uparrow \neg q$. That $\{\uparrow\}$ must necessarily be functional complete once we know that $\{\downarrow\}$ is, will become obvious in the discussion of the duality theorem below. Even up to term equivalence, there exist still infinitely many signatures. Here signatures are called *term equivalent* if the formulas of these signatures represent the same Boolean functions as in Exercise 2, for instance.

Define inductively on the formulas from $\mathcal{F}$ a mapping $\delta : \mathcal{F} \to \mathcal{F}$ by

$$p^\delta = p, \quad (\neg\alpha)^\delta = \neg\alpha^\delta, \qquad (\alpha \wedge \beta)^\delta = \alpha^\delta \vee \beta^\delta, \quad (\alpha \vee \beta)^\delta = \alpha^\delta \wedge \beta^\delta.$$

α^δ is called the *dual formula* of α and is obtained from α simply by interchanging $\wedge$ and $\vee$. Obviously, for a DNF α, α^δ is a CNF, and vice versa. Define the *dual* of $f \in \boldsymbol{B}_n$ by $f^\delta \vec{x} := \neg f \neg \vec{x}$ with $\neg\vec{x} := (\neg x_1, \ldots, \neg x_n)$. Clearly $f^{\delta^2} := (f^\delta)^\delta = f$ since $(f^\delta)^\delta \vec{x} = \neg\neg f \neg\neg \vec{x} = f\vec{x}$. Note that $\wedge^\delta = \vee$, $\vee^\delta = \wedge$, $\leftrightarrow^\delta = +$, $\downarrow^\delta = \uparrow$, but $\neg^\delta = \neg$. In other words, $\neg$ is *self-dual*. One may check by going through all truth tables that essentially binary self-dual Boolean functions do not exist. But it was Dedekind who discovered the ternary self-dual function $d_3 : (x_1, x_2, x_3) \mapsto x_1 \wedge x_2 \vee x_1 \wedge x_3 \vee x_2 \wedge x_3$. The above notions of duality are combined in the following

Theorem 2.3 (The duality principle for two-valued logic). *If* α *represents the function* f *then the dual formula* α^δ *represents the dual function* f^δ.

Proof by induction on α. Trivial for $\alpha = p$. Let α, β represent f_1, f_2, respectively. Then $\alpha \wedge \beta$ represents $f : \vec{x} \mapsto f_1\vec{x} \wedge f_2\vec{x}$ and, in view of the induction hypothesis,

$(\alpha\wedge\beta)^\delta = \alpha^\delta \vee \beta^\delta$ represents $g : \vec{x} \mapsto f_1^\delta\vec{x} \vee f_2^\delta\vec{x}$. This is just the dual of f because

$$f^\delta\vec{x} = \neg f\neg\vec{x} = \neg(f_1\neg\vec{x}\wedge f_2\neg\vec{x}) = \neg f_1\neg\vec{x}\vee\neg f_2\neg\vec{x} = f_1^\delta\vec{x}\vee f_2^\delta\vec{x} = g\vec{x}.$$

The induction step for $\vee$ is similar. Now let α represent f. Then $\neg\alpha$ represents $\neg f : \vec{x} \mapsto \neg f\vec{x}$. By the induction hypothesis, α^δ represents f^δ. Thus $(\neg\alpha)^\delta = \neg\alpha^\delta$ represents $\neg f^\delta$ which coincides with $(\neg f)^\delta$ as is readily confirmed. ❑

We know, for example, that $\leftrightarrow$ is represented by $p\wedge q \vee \neg p\wedge\neg q$, hence $+ \ (= \leftrightarrow^\delta)$ by $(p\vee q)\wedge(\neg p\vee\neg q)$. More generally, if $f \in \boldsymbol{B}_n$ is represented by a canonical DNF α, then by the theorem, f^δ is represented by the canonical CNF α^δ. Thus, if every $f \in \boldsymbol{B}_n$ is representable by a DNF then every f must necessarily be representable by a CNF, because $f \mapsto f^\delta$ constitutes a bijection of $\boldsymbol{B}_n$, as follows directly from $f^{\delta^2} = f$. Note also that Dedekind's ternary self-dual function d_3 defined above shows that $p\wedge q \vee p\wedge r \vee q\wedge r \equiv (p\vee q)\wedge(p\vee r)\wedge(q\vee r)$ in view of Theorem 2.3.

Remark 2. $\{\wedge, \vee, 0, 1\}$ is *maximally functional incomplete*, that is, if f is any Boolean function not representable by a formula in $\wedge, \vee, 0, 1$, then $\{\wedge, \vee, 0, 1, f\}$ is functional complete (Exercise 4). As was shown by E. Post (1920), there are up to term equivalence only five maximally functional incomplete logical signatures: besides $\{\wedge, \vee, 0, 1\}$ only $\{\rightarrow, \wedge\}$, the dual of this, $\{\leftrightarrow, \neg\}$, and $\{d_3, \neg\}$. The formulas of the last one represent just the self-dual Boolean functions. Since $\neg p \equiv 1 + p$, the signature $\{0, 1, +, \cdot\}$ is functional complete, where $\cdot$ is written in place of $\wedge$. The deeper reason is that $\{0, 1, +, \cdot\}$ is also the extralogical signature of fields (see **2.1**). Functional completeness in the two-valued case just derives from the fact that for a finite field, each operation on its domain is represented by a suitable polynomial. We mention also that for any finite set M of truth values considered in many-valued logics there is a generalized two-argument Sheffer function, by which every operation on M can be obtained, similarly to $\uparrow$ in the two-valued case.

Exercises

1. Verify the logical equivalences

   $$(p\rightarrow q_1)\wedge(\neg p\rightarrow q_2) \equiv p\wedge q_1 \vee \neg p\wedge q_2, \quad p_1\wedge q_1\rightarrow p_2\vee q_2 \equiv (p_1\rightarrow p_2)\vee(q_1\rightarrow q_2).$$

2. Show that the signatures $\{+, 1\}$, $\{+, \neg\}$, $\{\leftrightarrow, 0\}$, and $\{\leftrightarrow, \neg\}$ are all term equivalent. The formulas of each of these signatures represent precisely the linear Boolean functions.

3. Set $0 \leqslant 0$, $0 \leqslant 1$, and $1 \leqslant 1$ as usual. Show that the formulas in $\wedge, \vee, 0, 1$ represent exactly the *monotonic* Boolean functions. These are the constants from $\boldsymbol{B}_0$ and for $n > 0$ the $f \in \boldsymbol{B}_n$ such that for $i = 1, \ldots, n$,

   $$f(x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_n) \leqslant f(x_1, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_n).$$

4. Show that the signature $\{\wedge, \vee, 0, 1\}$ is maximally functional incomplete.

1.3 Tautologies and Logical Consequence

Instead of $w\alpha = 1$ we prefer from now on to write $w \vDash \alpha$ and read this w *satisfies* α. Further, if X is a set of formulas, we write $w \vDash X$ if $w \vDash \alpha$ for all $\alpha \in X$ and say that w is a (propositional) *model for* X. A given α (resp. X) is called *satisfiable* if there is some w with $w \vDash \alpha$ (resp. $w \vDash X$). The relation $\vDash$, called the *satisfiability relation*, evidently has the following properties:

$$w \vDash p \;\Leftrightarrow\; wp = 1 \quad (p \in PV); \qquad w \vDash \neg\alpha \;\Leftrightarrow\; w \nvDash \alpha;$$
$$w \vDash \alpha \wedge \beta \;\Leftrightarrow\; w \vDash \alpha \text{ and } w \vDash \beta; \qquad w \vDash \alpha \vee \beta \;\Leftrightarrow\; w \vDash \alpha \text{ or } w \vDash \beta.$$

One can define the satisfiability relation $w \vDash \alpha$ for a given $w\colon PV \to \{0,1\}$ also inductively on α, according to the clauses just given. This approach is particularly useful for extending the satisfiability conditions in **2.3**.

It is obvious that $w\colon PV \to \{0,1\}$ will be uniquely determined by setting down in advance for which variables $w \vDash p$ should be valid. Likewise the notation $w \vDash \alpha$ for $\alpha \in \mathcal{F}_n$ is already meaningful when w is defined only for $p_1 \ldots, p_n$. One could extend such a w to a global valuation by setting, for example, $wp = 0$ for all not mentioned variables p.

For formulas containing other connectives the satisfaction conditions are to be formulated accordingly. For example, we would expect that

$$w \vDash \alpha \to \beta \quad\Leftrightarrow\quad \text{if } w \vDash \alpha \text{ then } w \vDash \beta.$$

If $\to$ is taken to be a primitive connective, this clause is required. However, we defined $\to$ in such a way that this satisfaction clause is provable.

Definition. α is called *logically valid* or a (two-valued) *tautology*, in short $\vDash \alpha$, if $w \vDash \alpha$ for all w. A formula not satisfiable at all is called a *contradiction*.

Examples. $p \vee \neg p$ is a tautology and so is $\alpha \vee \neg\alpha$ for every formula α, the so-called *law of the excluded middle* or the *tertium non datur*. On the other hand, $\alpha \wedge \neg\alpha$ and $\alpha \leftrightarrow \neg\alpha$ are always contradictions. The following tautologies in $\to$ are mentioned in most textbooks on logic (association to the right is applied only to some extend, to keep these formulas more easily in mind):

$$p \to p,$$
$$(p \to q) \to (q \to r) \to (p \to r),$$
$$(p \to q \to r) \to (q \to p \to r),$$
$$p \to q \to p \qquad \text{(premise charge)},$$
$$(p \to q \to r) \to (p \to q) \to (p \to r) \qquad \text{(Frege's formula)},$$
$$((p \to q) \to p) \to p \qquad \text{(Peirce's formula)}.$$

It will later turn out that all tautologies in $\to$ alone are derivable (in a sense still to be explained) from the last three named formulas.

Clearly, it is decidable whether a formula α is a tautology, in that one tries out the valuations of the variables of α. Unfortunately, no essentially more efficient method is known; such a method exists only for formulas of a certain form. We will have a somewhat closer look at this problem in **4.2**. Various questions like checking the equivalence of formulas can be reduced to a decision about whether a formula is a tautology. Observe, in particular, that $\alpha \equiv \beta \;\Leftrightarrow\; \vDash \alpha \leftrightarrow \beta$.

Definition. α is a *logical consequence* of X, written $X \vDash \alpha$, if $w \vDash \alpha$ for every model w of X. In short, $w \vDash X \Rightarrow w \vDash \alpha$, for all w.

While we use $\vDash$ both as the symbol for logical consequence (which is a relation between sets of formulas X and formulas α) and the satisfiability property, it will always be clear from the context what $\vDash$ actually means. Evidently, α is a tautology iff $\emptyset \vDash \alpha$, so that $\vDash \alpha$ can be regarded as an abbreviation for $\emptyset \vDash \alpha$.

In this book, $X \vDash \alpha, \beta$ will always mean '$X \vDash \alpha$ and $X \vDash \beta$'. More generally, $X \vDash Y$ is always to mean '$X \vDash \beta$ for all $\beta \in Y$'. We also write $\alpha_1, \ldots, \alpha_n \vDash \beta$ in place of $\{\alpha_1, \ldots, \alpha_n\} \vDash \beta$, and more briefly, $X, \alpha \vDash \beta$ in place of $X \cup \{\alpha\} \vDash \beta$.

Before giving examples, we note the following obvious properties:

$$\begin{array}{lll} \text{(R)} & \alpha \in X \Rightarrow X \vDash \alpha & (\textit{reflexivity}), \\ \text{(M)} & X \vDash \alpha \;\&\; X \subseteq X' \;\Rightarrow\; X' \vDash \alpha & (\textit{monotonicity}), \\ \text{(T)} & X \vDash Y \;\&\; Y \vDash \alpha \;\Rightarrow\; X \vDash \alpha & (\textit{transitivity}). \end{array}$$

Examples of logical consequence. (a) $\alpha, \beta \vDash \alpha \wedge \beta$ and $\alpha \wedge \beta \vDash \alpha, \beta$. This is evident from the truth table of $\wedge$. In view of (T), property (a) can also be stated as $X \vDash \alpha, \beta \Leftrightarrow X \vDash \alpha \wedge \beta$. (b) $\alpha, \alpha \to \beta \vDash \beta$, because $1 \to x = 1 \Rightarrow x = 1$ according to the truth table of $\to$. (c) $X \vDash \bot \;\Rightarrow\; X \vDash \alpha$ for each α, because $X \vDash \bot = p_1 \wedge \neg p_1$ clearly means that X is unsatisfiable (has no model). $X = \{p, \neg p\}$ is an example. (d) $X, \alpha \vDash \beta \;\&\; X, \neg\alpha \vDash \beta \;\Rightarrow\; X \vDash \beta$. Indeed, let $w \vDash X$. If $w \vDash \alpha$ then $X, \alpha \vDash \beta$ and hence $w \vDash \beta$; but if $w \vDash \neg\alpha$ then $w \vDash \beta$ follows from $X, \neg\alpha \vDash \beta$. Note that (d) reflects our case distinction made in the metatheory.

The property exemplified by (b) is also called *modus ponens* when formulated as a rule of inference, as will be done in **1.6**. Example (d) is another formulation of the often-used procedure of proof by cases: In order to conclude a sentence β from a set of premises X it suffices to show it to be a logical consequence both under an additional supposition and under its negation. This is generalized in Exercise 3.

Useful for many purposes is the closure of the logical consequence relation under substitution, which is a generalization of the fact that from $p \vee \neg p$ all tautologies of the form $\alpha \vee \neg\alpha$ arise from substituting α for p.

Definition. A *(propositional) substitution* is a mapping $\sigma : PV \to \mathcal{F}$ that can be extended in a natural way to a mapping $\sigma : \mathcal{F} \to \mathcal{F}$ as follows:

$$(\alpha \wedge \beta)^\sigma = \alpha^\sigma \wedge \beta^\sigma, \quad (\alpha \vee \beta)^\sigma = \alpha^\sigma \vee \beta^\sigma, \quad (\neg\alpha)^\sigma = \neg\alpha^\sigma.$$

Like valuations, substitutions can be considered as operating on the whole of $\mathcal{F}$. For example, if $p^\sigma = \alpha$ for some fixed p and $q^\sigma = q$ otherwise, then φ^σ arises from φ by substituting α for p at all occurrences of p in φ. For $X \subseteq \mathcal{F}$ let $X^\sigma := \{\varphi^\sigma \mid \varphi \in X\}$. The observation $\models \varphi \;\Rightarrow\; \models \varphi^\sigma$ turns out to be the special instance $X = \emptyset$ of the interesting property

(S) $\quad X \models \alpha \;\Rightarrow\; X^\sigma \models \alpha^\sigma \qquad$ (*substitution invariance*).

In order to verify (S), let w^σ for a given valuation w be defined by $w^\sigma p = wp^\sigma$. We first need to prove by induction on α that

$$(*) \quad w \models \alpha^\sigma \;\Leftrightarrow\; w^\sigma \models \alpha.$$

With α a prime formula, $(*)$ certainly holds. Further,

$$\begin{aligned} w \models (\alpha \wedge \beta)^\sigma \Leftrightarrow w \models \alpha^\sigma \wedge \beta^\sigma &\Leftrightarrow w \models \alpha^\sigma, \beta^\sigma \\ &\Leftrightarrow w^\sigma \models \alpha, \beta \quad \text{(induction hypothesis)} \\ &\Leftrightarrow w^\sigma \models \alpha \wedge \beta. \end{aligned}$$

The reasoning for $\vee$ and $\neg$ is analogous and so $(*)$ holds. To prove (S), let $X \models \alpha$ and $w \models X^\sigma$. By $(*)$, we get $w^\sigma \models X$. Thus $w^\sigma \models \alpha$, and again by $(*)$, $w \models \alpha^\sigma$. Another property of $\models$, important for applications, will be proved in **1.4**, namely

(F) $\quad X \models \alpha \;\Rightarrow\; X_0 \models \alpha$ for some finite subset $X_0 \subseteq X$.

$\models$ shares the properties (R), (M), (T), and (S) with almost all classical and non-classical (many-valued) logical systems. A relation $\vdash$ between sets of formulas and formulas of an arbitrary propositional language $\mathcal{F}$ is called a (propositional) *consequence relation* if $\vdash$ has the properties corresponding to (R), (M), (T), and (S). These properties are the starting point for a general and strong theory of logical systems created by Tarski, which underpins nearly all the logical systems considered in the literature. Should $\vdash$ satisfy the correspondingly formulated property (F) (which is not supposed, in general), then $\vdash$ is called *finitary*.

Remark. Notions such as tautology, consistency, maximal consistency (to be considered in **1.4**), and so on can be used with reference to *any* consequence relation $\vdash$. For instance, a set of formulas X is called *consistent* in $\vdash$ whenever $X \nvdash \alpha$ for some α, and $\vdash$ itself is consistent when $\nvdash \alpha$ for some α. If $\mathcal{F}$ contains $\neg$ then the consistency of X is often defined by $X \vdash \alpha, \neg\alpha$ for no α. But the aforementioned definition has the advantage of being completely independent on any assumption concerning the occurring connectives. Another example of a general definition is this: A formula set X is called *deductively closed in* $\vdash$ provided $X \vdash \alpha \Rightarrow \alpha \in X$, for all $\alpha \in \mathcal{F}$. Because of (R), this condition can be replaced by $X \vdash \alpha \Leftrightarrow \alpha \in X$. Examples in $\models$ are the set of all tautologies and the whole of $\mathcal{F}$. The intersection of a family of deductively closed sets is again deductively closed. Hence, each $X \subseteq \mathcal{F}$ is contained in a smallest deductively closed set, called the *deductive closure* of X. The notion of a consequence relation can also be defined in terms

of properties of the deductive closure. We mention that (F) holds not just for our $\vDash$ which is given by a two-valued matrix, but for the consequence relation of *any* finite logical matrix in *any* propositional language. This is stated and at once essentially generalized in Exercise 3 from **5.7** as an application of the ultraproduct theorem.

A special property of $\vDash$, easily provable, is

(D) $\quad X, \alpha \vDash \beta \;\Rightarrow\; X \vDash \alpha \to \beta.$

called the (semantic) *deduction theorem* for propositional logic. To see this suppose $X, \alpha \vDash \beta$ and let w be a model for X. If $w \vDash \alpha$ then by the supposition $w \vDash \beta$. If $w \nvDash \alpha$ then $w \vDash \alpha \to \beta$ as well. This proves $w \vDash \alpha \to \beta$ and hence (D).

As is immediately seen, the converse of (D) holds as well, that is, one may replace $\Rightarrow$ in (D) by $\Leftrightarrow$. Iterated application of this simple observation yields

$$\alpha_1, \ldots, \alpha_n \vDash \beta \;\;\Leftrightarrow\;\; \vDash \alpha_1 \to \alpha_2 \to \cdots \to \alpha_n \to \beta \;\Leftrightarrow\; \vDash \alpha_1 \wedge \alpha_2 \wedge \cdots \wedge \alpha_n \to \beta.$$

In this way, β's being a logical consequence of a finite set of premises is transformed into a tautology. Using (D) it is easy to obtain tautologies. For instance, to prove $\vDash p \to q \to p$, it is enough to verify $p \vDash q \to p$, for which it in turn suffices to show that $p, q \vDash p$, and this is trivial. By some simple applications of (D) each of the tautologies in the examples on page 14 can be obtained, except the formula of Peirce. As we shall see in Chapter **2**, all properties of $\vDash$ derived above and in the exercises will carry over to the consequence relation of a first-order language.

Exercises

1. Use the deduction theorem similar to its application in the text to prove

   (a) $\vDash (p \to q \to r) \to (p \to q) \to (p \to r)$ (b) $\vDash (p \to q) \to (q \to r) \to (p \to r)$.

2. Suppose that $X \vDash \alpha \to \beta$. Prove that $X \vDash (\gamma \to \alpha) \to (\gamma \to \beta)$.

3. Verify the (rule of) *disjunctive case distinction*: if $X, \alpha \vDash \gamma$ and $X, \beta \vDash \gamma$ then $X, \alpha \vee \beta \vDash \gamma$. This implication is written more suggestively as

$$\frac{X, \alpha \vDash \gamma \;\big|\; X, \beta \vDash \gamma}{X, \alpha \vee \beta \vDash \gamma}.$$

4. Verify the following *rules of contraposition*:

$$\frac{X, \alpha \vDash \beta}{X, \neg\beta \vDash \neg\alpha} \;\text{ and }\; \frac{X, \neg\beta \vDash \neg\alpha}{X, \alpha \vDash \beta}.$$

5. Let $\vdash$ be a consequence relation in $\mathcal{F}$ and $\bar{X} := \{\alpha \in \mathcal{F} \mid X \vdash \alpha\}$. Show that $\bar{X}$ is the smallest deductively closed set of formula containing X.

1.4 A Complete Calculus for $\models$

We will now define a derivability relation $\vdash$ by means of a calculus operating solely with some structural rules. $\vdash$ turns out to be identical to the consequence relation $\models$. The calculus $\vdash$ is of the so-called Gentzen type and its rules are given with respect to pairs (X, α) of sets of formulas X and formulas α. Another calculus for $\models$, of the Hilbert type, will be considered in **1.6**. In distinction to [Ge], we do not require that X be finite; our particular goals here make such a restriction dispensable. If $\vdash$ applies on the pair (X, α) then we write $X \vdash \alpha$ and say that α is *derivable* or *provable* from X (made precise below); otherwise we write $X \nvdash \alpha$.

Following [Kl1], Gentzen's name for (X, α), *Sequenz*, is translated as *sequent*. The calculus is formulated in terms of $\wedge, \neg$ and encompasses the following six rules, called the *basic rules*. Other rules derived from these are called *provable* or *derivable*. The choice of $\{\wedge, \neg\}$ as the basic signature is a matter of convenience and justified by its functional completeness. The other standard connectives are introduced by the definitions $\alpha \vee \beta := \neg(\neg\alpha \wedge \neg\beta)$, $\alpha \to \beta := \neg(\alpha \wedge \neg\beta)$, $\alpha \leftrightarrow \beta := (\alpha \to \beta) \wedge (\beta \to \alpha)$.

Of course, one could choose any other functional complete signature and change or adapt the basic rules correspondingly. But it should be observed that a complete calculus in $\neg, \wedge, \vee, \to$, say, must also include basic rules concerning $\vee$ and $\to$, which makes induction arguments on the basic rules of the calculus more lengthy.

Each of the basic rules below has certain premises and a conclusion. Only (IS) has no premises. It allows the derivation of all sequents $\alpha \vdash \alpha$. These are called the *initial* sequents, because each derivation must start with these. We mention that each of the six basic rules is really needed for proving the completeness of $\vdash$.

(IS) $\dfrac{}{\alpha \vdash \alpha}$ (initial sequent)	(MR) $\dfrac{X \vdash \alpha}{X' \vdash \alpha}$ $(X' \supseteq X,$ monotonicity$)$
$(\wedge 1)$ $\dfrac{X \vdash \alpha, \beta}{X \vdash \alpha \wedge \beta}$	$(\wedge 2)$ $\dfrac{X \vdash \alpha \wedge \beta}{X \vdash \alpha, \beta}$
$(\neg 1)$ $\dfrac{X \vdash \alpha, \neg\alpha}{X \vdash \beta}$	$(\neg 2)$ $\dfrac{X, \alpha \vdash \beta \;\mid\; X, \neg\alpha \vdash \beta}{X \vdash \beta}$

Here and in the following $X \vdash \alpha, \beta$ is to mean $X \vdash \alpha$ and $X \vdash \beta$. This convention is important since $X \vdash \alpha, \beta$ has another meaning in Gentzen calculi, which are given with respect to pairs of sets of formulas and which play a role in proof-theoretical investigations. Thus, the rules $(\wedge 1)$ and $(\neg 1)$ actually have two premises, just like $(\neg 2)$. Note further that $(\wedge 2)$ really consists of two subrules corresponding to the conclusions $X \vdash \alpha$ and $X \vdash \beta$. In $(\neg 2)$, X, α stands for $X \cup \{\alpha\}$, and this abbreviated form will always be used when there is no risk of misunderstanding.

$\alpha_1, \ldots, \alpha_n \vdash \beta$ stands for $\{\alpha_1, \ldots, \alpha_n\} \vdash \beta$; in particular $\alpha \vdash \beta$ for $\{\alpha\} \vdash \beta$ and $\vdash \alpha$ for $\emptyset \vdash \alpha$, just as with $\models$. The rule (MR) becomes provable if all (X, α) with $\alpha \in X$ are called initial sequents, that is, if (IS) is strengthened to $\dfrac{}{X \vdash \alpha}$ $(\alpha \in X)$.

$X \vdash \alpha$ (read "X *derivable* α") is to mean that the sequent (X, α) can be obtained though a stepwise application of the basic rules. We can make this idea of "stepwise application" of the basic rules rigorous and formally precise (intelligible to a computer, so to speak) in the following way: a *derivation* is to mean a finite sequence $(S_0; \ldots; S_n)$ of sequents such that every S_i is either an initial sequent or is obtained through the application of some basic rule to preceding elements in the sequence. So α is derivable from X when there is a derivation $(S_0; \ldots; S_n)$ with $S_n = (X, \alpha)$. An example with the end sequent $\alpha, \beta \vdash \alpha \wedge \beta$ is the derivation

$$(\alpha \vdash \alpha;\ \alpha, \beta \vdash \alpha;\ \beta \vdash \beta;\ \alpha, \beta \vdash \beta;\ \alpha, \beta \vdash \alpha \wedge \beta).$$

More interesting is the derivation of additional rules, which we will illustrate with the examples to follow. The second example, a generalization of the first, is the often-used proof method *reductio ad absurdum*: α is proved from X by showing that the assumption $\neg\alpha$ leads to a contradiction. The other examples are given with respect to the defined $\rightarrow$-connective. Hence, for instance, the $\rightarrow$-elimination mentioned below runs in the original language $\dfrac{X \vdash \neg(\alpha \wedge \neg\beta)}{X, \alpha \vdash \beta}$.

Examples of provable rules

		proof	*applied*
$\dfrac{X, \neg\alpha \vdash \alpha}{X \vdash \alpha}$			
($\neg$-elimination)	1	$X, \alpha \vdash \alpha$	(IS), (MR)
	2	$X, \neg\alpha \vdash \alpha$	supposition
	3	$X \vdash \alpha$	($\neg$2)
$\dfrac{X, \neg\alpha \vdash \beta, \neg\beta}{X \vdash \alpha}$			
(reductio ad absurdum)	1	$X, \neg\alpha \vdash \beta, \neg\beta$	supposition
	2	$X, \neg\alpha \vdash \alpha$	($\neg$1)
	3	$X \vdash \alpha$	$\neg$-elimination
$\dfrac{X \vdash \alpha \rightarrow \beta}{X, \alpha \vdash \beta}$			
($\rightarrow$-elimination)	1	$X, \alpha, \neg\beta \vdash \alpha, \neg\beta$	(IS), (MR)
	2	$X, \alpha, \neg\beta \vdash \alpha \wedge \neg\beta$	($\wedge$1)
	3	$X \vdash \neg(\alpha \wedge \neg\beta)$ $(= \alpha \rightarrow \beta)$	supposition
	4	$X, \alpha, \neg\beta \vdash \neg(\alpha \wedge \neg\beta)$	(MR)
	5	$X, \alpha, \neg\beta \vdash \beta$	($\neg$1) on 2 and 4
	6	$X, \alpha \vdash \beta$	$\neg$-elimination

$$\frac{X \vdash \alpha \;\big|\; X, \alpha \vdash \beta}{X \vdash \beta}$$
(cut rule)

	proof	*applied*
1	$X, \neg\alpha \vdash \alpha$	supposition, (MR)
2	$X, \neg\alpha \vdash \neg\alpha$	(IS), (MR)
3	$X, \neg\alpha \vdash \beta$	$(\neg 1)$
4	$X, \alpha \vdash \beta$	supposition
5	$X \vdash \beta$	$(\neg 2)$ on 4 and 3

$$\frac{X, \alpha \vdash \beta}{X \vdash \alpha \to \beta}$$
($\to$-introduction)

1	$X, \alpha \wedge \neg\beta, \alpha \vdash \beta$	supposition, (MR)
2	$X, \alpha \wedge \neg\beta \vdash \alpha$	(IS), (MR), $(\wedge 2)$
3	$X, \alpha \wedge \neg\beta \vdash \beta$	Cut rule
4	$X, \alpha \wedge \neg\beta \vdash \neg\beta$	(IS), (MR), $(\wedge 2)$
5	$X, \alpha \wedge \neg\beta \vdash \alpha \to \beta$	$(\neg 1)$
6	$X, \neg(\alpha \wedge \neg\beta) \vdash \alpha \to \beta$	(IS), (MR)
7	$X \vdash \alpha \to \beta$	$(\neg 2)$ on 5 and 6

The example of $\to$-introduction is nothing other than the syntactic form of the deduction theorem that was semantically formulated in the previous section.

Remark 1. The deduction theorem also holds for intuitionistic logic. However, it is not in general true for all logical systems dealing with implication, thus indicating that the deduction theorem is not an inherent property of every meaningful conception of implication. For instance, it is not valid for certain formal systems of relevance logic that attempt to model implication as a cause-and-effect relation.

A simple consequence of $\to$-elimination and the cut rule is the *detachment rule*

$$\frac{X \vdash \alpha, \alpha \to \beta}{X \vdash \beta}.$$

For notice that the premise $X \vdash \alpha \to \beta$ yields $X, \alpha \vdash \beta$ by $\to$-elimination, and since $X \vdash \alpha$, the cut rule yields $X \vdash \beta$. Applying detachment on $X = \{\alpha, \alpha \to \beta\}$, we obtain $\alpha, \alpha \to \beta \vdash \beta$. This collection of sequents is known as *modus ponens*, which will be more closely considered in **1.6**.

Many properties of $\vdash$ are proved through rule induction, which we describe after introducing some convenient terminology. We identify a property $\mathcal{E}$ of sequents with the set of all pairs (X, α) to which $\mathcal{E}$ applies. In this sense the logical consequence relation $\vDash$ is the property applying to all pairs (X, α) with $X \vDash \alpha$.

All the rules considered here are of the form

$$R: \quad \frac{X_1 \vdash \alpha_1 \;\big|\; \cdots \;\big|\; X_n \vdash \alpha_n}{X \vdash \alpha}$$

and are referred to as Gentzen-style rules. We say that $\mathcal{E}$ is *closed under* R when $\mathcal{E}(X_1, \alpha_1), \ldots, \mathcal{E}(X_n, \alpha_n)$ implies $\mathcal{E}(X, \alpha)$. For a rule without premises, i.e., $n = 0$,

this is just to mean $\mathcal{E}(X, \alpha)$. For instance, consider the property $\mathcal{E} : X \models \alpha$. Each basic rule of $\vdash$ is closed under $\mathcal{E}$. In detail this means

$$\alpha \models \alpha, \quad X \models \alpha \Rightarrow X' \models \alpha \text{ for } X' \supseteq X, \quad X \models \alpha, \beta \Rightarrow X \models \alpha \wedge \beta, \quad \text{etc.}$$

From the latter it will follow that $\mathcal{E}$ applies to all provable sequents; in other words, $\vdash$ is (semantically) *sound*. What we need here is the following easily justifiable

Principle of rule induction. *Let $\mathcal{E}$ ($\subseteq \mathfrak{P}\mathfrak{F} \times \mathfrak{F}$) be a property closed under all basic rules of $\vdash$. Then $X \vdash \alpha$ implies $\mathcal{E}(X, \alpha)$.*

Proof by induction on the length of a derivation of the sequent $S = (X, \alpha)$. If the length is 1, $\mathcal{E}S$ holds since S must be an initial sequent. Now let $(S_0; \ldots; S_n)$ be a derivation of the sequent $S := S_n$. By the induction hypothesis we have $\mathcal{E}S_i$ for all $i < n$. If S is an initial sequent then $\mathcal{E}S$ holds by assumption. Otherwise S has been obtained by the application of a basic rule on some of the S_i for $i < n$. But then $\mathcal{E}S$ holds, because $\mathcal{E}$ is closed under all basic rules. ❑

As already remarked, the property $X \models \alpha$ is closed under all basic rules. Therefore, the principle of rule induction immediately yields the *soundness* of the calculus, that is, $\vdash \subseteq \models$. More explicitly, $X \vdash \alpha \Rightarrow X \models \alpha$, for all X, α.

There are several equivalent definitions of $\vdash$. One that is purely set-theoretical is the following: $\vdash$ is the smallest of all relations $\subseteq \mathfrak{P}\mathfrak{F} \times \mathfrak{F}$ that are closed under all basic rules. The equivalence proofs of such definitions are wordy but not particularly contentful. We therefore do not elaborate further, especially because we henceforth only use rule induction and not the lengthy definition of $\vdash$. Using rule induction one can also prove $X \vdash \alpha \Rightarrow X^\sigma \vdash \alpha^\sigma$, and in particular the following theorem, for which the soundness of $\vdash$ is completely irrelevant.

Theorem 4.1 (Finiteness theorem for $\vdash$). *If $X \vdash \alpha$ then there is a finite subset $X_0 \subseteq X$ with $X_0 \vdash \alpha$.*

Proof. Let $\mathcal{E}(X, \alpha)$ be the property '$X_0 \vdash \alpha$ for some finite $X_0 \subseteq X$'. Certainly, $\mathcal{E}(X, \alpha)$ holds for $X = \{\alpha\}$, with $X_0 = X$. If X has a finite subset X_0 such that $X_0 \vdash \alpha$, then so too does every set X' such that $X' \supseteq X$. Hence $\mathcal{E}$ is closed under (MR). Let $\mathcal{E}(X, \alpha)$, $\mathcal{E}(X, \beta)$, with, say, $X_1 \vdash \alpha$, $X_2 \vdash \beta$ for finite $X_1, X_2 \subseteq X$. Then we also have $X_0 \vdash \alpha, \beta$ for $X_0 = X_1 \cup X_2$ by (MR). Hence $X_0 \vdash \alpha \wedge \beta$ by ($\wedge 1$). Thus $\mathcal{E}(X, \alpha \wedge \beta)$ holds, and $\mathcal{E}$ is closed under ($\wedge 1$). Analogously one shows the same for all remaining basic rules of $\vdash$. The claim then follows by rule induction. ❑

Of great significance is the notion of formal consistency. It fully determines the derivability relation, as the lemma to come shows. It will turn out that "consistent" formalizes adequately the notion "satisfiable." The proof of this adequacy is the clue to the completeness problem.

Definition. $X \subseteq \mathcal{F}$ is called *inconsistent* (in our calculus $\vdash$) if $X \vdash \alpha$ for all $\alpha \in \mathcal{F}$, and otherwise *consistent*. X is called *maximally consistent* if X is consistent but each $Y \supset X$ is inconsistent, or equivalently, $\alpha \notin X \Rightarrow X, \alpha \vdash \beta$ for all β.

The inconsistency of X can be identified by the derivability of a single formula, namely $\bot$ $(= p_1 \wedge \neg p_1)$. This is so because $X \vdash \bot$ implies $X \vdash p_1, \neg p_1$ by $(\wedge 2)$, hence $X \vdash \alpha$ for all α by $(\neg 1)$. Conversely, when X is inconsistent then in particular $X \vdash \bot$. Thus, $X \vdash \bot$ may be read as "X is inconsistent," and $X \nvdash \bot$ as "X is consistent." The most important is resumed by the following lemma in the properties $\mathtt{C}^+$ and $\mathtt{C}^-$, which can also each be understood as a pair of provable rules.

Lemma 4.2. *The derivability relation $\vdash$ has the properties*

$$\mathtt{C}^+: \quad X \vdash \alpha \;\Leftrightarrow\; X, \neg\alpha \vdash \bot, \qquad \mathtt{C}^-: \quad X \vdash \neg\alpha \;\Leftrightarrow\; X, \alpha \vdash \bot.$$

Proof. If $X \vdash \alpha$ holds then so too does $X, \neg\alpha \vdash \alpha$. Since certainly $X, \neg\alpha \vdash \neg\alpha$, we have $X, \neg\alpha \vdash \beta$ for all β by $(\neg 1)$, in particular $X, \neg\alpha \vdash \bot$. Conversely, let $X, \neg\alpha \vdash \bot$ be the case, so that in particular $X, \neg\alpha \vdash \alpha$, and thus $X \vdash \alpha$ by $\neg$-elimination on page 19. $\mathtt{C}^-$ is proved completely analogously. ❏

The claim $\models \subseteq \vdash$, not yet proved, is equivalent to $X \nvdash \alpha \Rightarrow X \not\models \alpha$, for all X and α. But so formulated it becomes apparent what needs to be done to obtain the proof. Since $X \nvdash \alpha$ is by $\mathtt{C}^+$ equivalent to the consistency of $X' := X \cup \{\neg\alpha\}$, and likewise $X \not\models \alpha$ to the satisfiability of X', we need only show that consistent sets are satisfiable. To this end we state the following lemma whose proof, exceptionally, jumps ahead of matters in that it uses Zorn's Lemma from **2.1** page 37.

Lemma 4.3 (Lindenbaum's theorem). *Every consistent set X can be extended to a maximally consistent set $X' \supseteq X$.*

Proof. Let H be the set of all consistent $Y \supseteq X$, partially ordered with respect to $\subseteq$. $H \neq \emptyset$, because $X \in H$. Let $K \subseteq H$ be a chain, i.e., $Y \subseteq Z$ or $Z \subseteq Y$, for all $Y, Z \in K$. Then $U = \bigcup K$ is an upper bound for K. Indeed, $Y \in K \Rightarrow Y \subseteq U$. Moreover, and this is here the point, U is consistent, so that $U \in H$. Assume $U \vdash \bot$. Then $U_0 \vdash \bot$ for some finite $U_0 = \{\alpha_0, \ldots, \alpha_n\} \subseteq U$. If, say, $\alpha_i \in Y_i \in K$, and Y is the biggest of the sets $Y_0, \ldots, Y_n$, then $\alpha_i \in Y$ for all $i \leqslant n$, hence also $Y \vdash \bot$ by (MR). This contradicts $Y \in H$. By Zorn's lemma, H therefore has a maximal element X', which is necessarily a maximally consistent extension of X. ❏

Remark 2. The advantage of this proof is that it is free of assumptions regarding the cardinality of the language. Lindenbaum's original construction was based, however, on countable languages $\mathcal{F}$ and runs as follows: Let $X_0 := X \subseteq \mathcal{F}$ be consistent and $\alpha_0, \alpha_1, \ldots$ be an enumeration of $\mathcal{F}$. Set $X_{n+1} = X_n \cup \{\alpha_n\}$ if this set is consistent and $X_{n+1} = X_n$ otherwise. Then $Y = \bigcup_{n\in\omega} X_n$ is a maximally consistent extension of X, as can be easily verified. Here Zorn's lemma, which is equivalent to the axiom of choice, is not required.

Lemma 4.4. *A maximally consistent set X has the property*

$$[\neg] \quad X \vdash \neg\alpha \;\Leftrightarrow\; X \nvdash \alpha, \text{ for arbitrary } \alpha.$$

Proof. If $X \vdash \neg\alpha$, then $X \vdash \alpha$ cannot hold due to the consistency of X. If on the other hand $X \nvdash \alpha$, then $X, \neg\alpha$ is by $\mathtt{C}^+$ a consistent extension of X. But then $\neg\alpha \in X$, because X is maximally consistent. Consequently $X \vdash \neg\alpha$. ❑

Only property $[\neg]$ from Lemma 4.4 and property $[\wedge]$ $X \vdash \alpha \wedge \beta \Leftrightarrow X \vdash \alpha, \beta$ are used in the simple model construction for maximally consistent sets in the following lemma, which reveals the requirements for proposional model construction in the logical base $\{\wedge, \neg\}$. If this base is changed, we need corresponding properties.

Lemma 4.5. *A maximally consistent set X is satisfiable.*

Proof. Define w by $w \models p \;\Leftrightarrow\; X \vdash p$. We are going to show that for all α,

$$(*) \quad X \vdash \alpha \;\Leftrightarrow\; w \models \alpha.$$

For prime formulas this is trivial. Further:

$$\begin{array}{lll} X \vdash \alpha \wedge \beta & \Leftrightarrow & X \vdash \alpha, \beta \quad (\text{rules } (\wedge 1), (\wedge 2)\,) \\ & \Leftrightarrow & w \models \alpha, \beta \quad (\text{induction hypothesis}) \\ & \Leftrightarrow & w \models \alpha \wedge \beta \quad (\text{definition}) \\ X \vdash \neg\alpha & \Leftrightarrow & X \nvdash \alpha \quad (\text{Lemma 4.4}) \\ & \Leftrightarrow & w \not\models \alpha \quad (\text{induction hypothesis}) \\ & \Leftrightarrow & w \models \neg\alpha \quad (\text{definition}). \end{array}$$

By $(*)$, w is a model for X, thereby completing the proof. ❑

The above shows the equivalence of the consistency and satisfiability of a set of formulas. From this fact we easily obtain the main result of the present section.

Theorem 4.6 (Completeness theorem). $X \vdash \alpha \;\Leftrightarrow\; X \models \alpha$, *for all* X, α.

Proof. The direction $\Rightarrow$ is the soundness of $\vdash$. Conversely, $X \nvdash \alpha$ implies that $X, \neg\alpha$ is consistent. Let Y be a maximally consistent extension of $X, \neg\alpha$, Lemma 4.3. By Lemma 4.5, Y is satisfiable, hence also $X, \neg\alpha$. Therefore $X \not\models \alpha$. ❑

An immediate consequence of Theorem 4.6 is the finiteness property (F) mentioned already in **1.3**, which is almost trivial for $\vdash$ but not for $\models$:

Theorem 4.7 (Finiteness theorem for $\models$). *If $X \models \alpha$, then so too $X_0 \models \alpha$ for some finite subset X_0 of X.*

This is clear because the finiteness theorem holds for $\vdash$ (Theorem 4.1). A further very important consequence of the completeness theorem is the following

Theorem 4.8 (Compactness theorem). *A set X is satisfiable provided each finite subset of X is satisfiable.*

This theorem holds because if X is unsatisfiable, i.e., $X \models \bot$, then by Theorem 4.7 we also know that $X_0 \models \bot$ for some finite $X_0 \subseteq X$, thus proving the claim. Conversely, one can easily obtain Theorem 4.7 from Theorem 4.8; that is, both theorems are directly derivable from one another.

Because Theorem 4.6 makes no assumptions regarding the cardinality of the set of variables, the compactness theorem following from it is likewise valid without the respective restrictions. That means that the theorem has many useful applications, as the next section illustrates.

Let us notice that direct proofs of Theorem 4.8 or appropriate reformulations of it can be given that have nothing to do with a calculus of logical rules. For example, the theorem is equivalent to $\bigcap_{\alpha\in X} \operatorname{Md}\alpha = \emptyset \;\Rightarrow\; \bigcap_{\alpha\in X_0} \operatorname{Md}\alpha = \emptyset$ for some finite $X_0 \subseteq X$, where $\operatorname{Md}\alpha$ denotes the set of all models of α. In this formulation the compactness of a certain naturally arising topological space is claimed; the points of this space are the valuations of the variables, hence the name "compactness theorem." More on this subject can be found in [RS].

Another approach to completeness (probably the simplest one) is provided by Exercises 3 and 4. This approach makes some elegant use of substitutions. It yields not only Theorems 4.6, 4.7, and 4.8 in one go, but also some further remarkable properties: Neither new tautologies nor new rules can consistently be adjoined to the consequence relation $\models$, the so-called Post completeness and structural completeness of $\models$, respectively; see for instance [Ra1] for details.

Exercises

1. Prove using Theorem 4.6: if $X \cup \{\neg\alpha \mid \alpha \in Y\}$ is inconsistent and $Y \neq \emptyset$, then there exist formulas $\alpha_0, \ldots, \alpha_n \in Y$ with $X \vdash \alpha_0 \vee \cdots \vee \alpha_n$.

2. Augment the signature $\{\neg, \wedge\}$ by $\vee$ and prove the completeness of the calculus obtained by supplementing the basic rules used so far with the rules
$$\frac{X \vdash \alpha}{X \vdash \alpha \vee \beta, \beta \vee \alpha}; \qquad \frac{X, \alpha \vdash \gamma \;\big|\; X, \beta \vdash \gamma}{X, \alpha \vee \beta \vdash \gamma}.$$

3. Let $\vdash$ be a finitary consequence relation in $\mathcal{F}\{\wedge, \neg\}$ with the properties $(\wedge 1)$ through $(\neg 2)$. Show that $\vdash$ is *maximal*, which is to mean $\vdash' \alpha$ for all α, for every proper extension $\vdash' \supset \vdash$. The latter is readily shown with the so-called *substitution method* explained in the Hints to the Exercises.

4. Show by referring to Exercise 3: there is exactly one consequence relation in $\mathcal{F}\{\wedge, \neg\}$ satisfying $(\wedge 1)$–$(\neg 2)$. This obviously implies the completeness of the calculus $\vdash$, for both $\models$ and $\vdash$ have these properties.

1.5 Applications of the Compactness Theorem

Theorem 4.8 is very useful in carrying over certain properties of finite structures to infinite ones. There follow some typical examples. While these could also be treated with the compactness theorem of predicate logic in **3.3**, the examples demonstrate how the consistency of certain sets of sentences in predicate logic can also be obtained in propositional logic. This approach is also useful for Chapter **4**.

1. Every set M can be (totally) ordered.[5]

This means that there is an irreflexive, transitive, and connex relation $<$ on M. For finite M this follows easily by induction on the number of elements of M. The claim is obvious when $M = \emptyset$ or is a singleton. Let now $M = N \cup \{a\}$ with an n-element set N and $a \notin N$, so that M has $n+1$ elements. Then we get an order on M from that for N by "setting a to the end," that is, defining $x < a$ for all $x \in N$.

Now let M be any set. We consider for every pair $(a, b) \in M \times M$ a propositional variable p_{ab}. Let X be the set consisting of the formulas

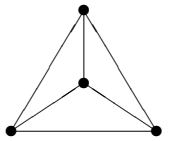
$$\begin{array}{ll} \neg p_{aa} & (a \in M), \\ p_{ab} \wedge p_{bc} \to p_{ac} & (a, b, c \in M), \\ p_{ab} \vee p_{ba} & (a \neq b). \end{array}$$

From a model w for X we obtain an order $<$, simply by putting $a < b \Leftrightarrow w \vDash p_{ab}$. $w \vDash \neg p_{aa}$ says the same thing as $a \not< a$. Analogously, the remaining formulas reflect transitivity and connexity. Thus, according to Theorem 4.8, it suffices to show that every finite subset $X_0 \subseteq X$ has a model. In X_0 only finitely many variables occur. Hence, there are finite sets $M_1 \subseteq M$ and $X_1 \supseteq X_0$, where X_1 is given exactly as X except that a, b, c now run through the finite set M_1 instead of M. But X_1 is satisfiable, because if $<$ is an order of the finite set M_1 and w is defined by $w \vDash p_{ab} \Leftrightarrow a < b$, then w is clearly a model for X_1, hence also for X_0.

2. The four-color theorem for infinite planar graphs.

A *simple graph* is a pair (V, E) with an irreflexive symmetrical relation $E \subseteq V^2$. The elements of V are called *points* or *vertices*. It is convenient to identify E with the set of all unordered pairs $\{a, b\}$ such that aEb and to call these pairs the *edges* of (V, E). If $\{a, b\} \in E$ then we say that a, b are *neighbors*. (V, E) is *k-colorable* if V can be decomposed into k *color classes* $C_i \neq \emptyset$, $V = C_1 \cup \cdots \cup C_k$, with $C_i \cap C_j = \emptyset$ for $i \neq j$, such that neighboring points do not carry the same color; in other words, if $a, b \in C_i$ then $\{a, b\} \notin E$ for $i = 1, \ldots, k$.

[5] Unexplained notions are defined in **2.1**. Our first application is interesting because in set theory the compactness theorem is weaker than the axiom of choice (AC) which is equivalent to the statement that every set can be well-ordered. Thus, the ordering principle is weaker than AC since it follows from the compactness theorem.



The figure shows the smallest four-colorable graph that is not three-colorable; all its points neighbor each other. We show that a graph (V, E) is k-colorable if every finite subgraph (V_0, E_0) is k-colorable. E_0 consists of the edges $\{a, b\} \in E$ with $a, b \in V_0$. To prove our claim consider the following set X of formulas built from the variables $p_{a,i}$ for $a \in V$ and $1 \leqslant i \leqslant k$:

$$\begin{array}{lll} p_{a,1} \vee \cdots \vee p_{a,k}, & \neg(p_{a,i} \wedge p_{a,j}) & (a \in V,\ 1 \leqslant i < j \leqslant k), \\ & \neg(p_{a,i} \wedge p_{b,i}) & (\{a, b\} \in E,\ i = 1, \ldots, k). \end{array}$$

The first formula states that every point belongs to at least one color class; the second ensures their disjointedness, and the third that no neighboring points have the same color. Once again it is enough to construct some $w \vDash X$. Defining then the C_i by $a \in C_i \Leftrightarrow w \vDash p_{a,i}$ proves that (V, E) is k-colorable. We must therefore satisfy each finite $X_0 \subseteq X$. Let (V_0, E_0) be the finite subgraph of (V, E) of all the points that occur as indices in the variables of X_0. The assumption on (V_0, E_0) obviously ensures the satisfiability of X_0 for reasons analogous to those given in Example 1, and this is all we need to show. The *four-colour theorem* says that every finite planar graph is four-colorable. Hence, the same holds for all graphs whose finite subgraphs are planar. These cover all planar graphs, embeddable in the real plane.

3. König's tree lemma.

There are several versions of this lemma. For simplicity, ours refers to a *directed tree*. This is a pair $(V, \lhd)$ with an irreflexive relation $\lhd \subseteq V^2$ such that for a certain point c, the *root* of the tree, and any other point a there is precisely one *path* connecting c with a. This is a sequence $(a_i)_{i \leqslant n}$ with $a_0 = c$, $a_n = a$, and $a_i \lhd a_{i+1}$ for all $i < n$. From the uniqueness of a path connecting c with any other point it follows that each $b \neq c$ has exactly one *predecessor* in $(V, \lhd)$, that is, a point a with $a \lhd b$.

The lemma in question then reads as follows: *If every $a \in V$ has only finitely many successors and V contains arbitrarily long finite paths, then there is an infinite path through V starting at c.* By such a path we mean an infinite sequence $(c_i)_{i \in \mathbb{N}}$ such that $c_0 = c$ and $c_i \lhd c_{i+1}$ for each i. In order to prove König's lemma we define inductively $S_0 = \{c\}$ and $S_{k+1} = \{b \in V \mid \text{there is some } a \in S_k \text{ with } a \lhd b\}$. Since every point has only finitely many successors, every "layer" S_k is finite, and since there are arbitrarily long paths starting in c, no S_k is empty. Now let p_a for every $a \in V$ be a propositional variable, and let X consist of the formulas

$$\begin{array}{lll} \text{(A)} & \bigvee_{a \in S_k} p_a, & \neg(p_a \wedge p_b) \quad \big(a, b \in S_k,\ a \neq b,\ k \in \mathbb{N}\big), \\ \text{(B)} & & p_b \to p_a \quad \big(a, b \in V,\ a \lhd b\big). \end{array}$$

Suppose that $w \vDash X$. Then by the formulas under (A), for every k there is precisely one $a \in S_k$ with $w \vDash p_a$, denoted by c_k. In particular, $c_0 = c$. Moreover, $c_k \lhd c_{k+1}$ for all k. Indeed, if a is the predecessor of $b = c_{k+1}$, then $w \vDash p_a$ in view of (B),

hence necessarily $a = c_k$. Thus, $(c_i)_{i\in\mathbb{N}}$ is a path of the type sought. Again, every finite subset $X_0 \subseteq X$ is satisfiable; for if X_0 contains variables with indices up to at most the layer S_n, then X_0 is a subset of a finite set of formulas X_1 that is defined as X, except that k runs only up to n, and for this case the claim is obvious.

4. The marriage problem (in linguistic guise).

Let N be a set of *words* or *names* (in speech) with meanings in a set M. A name $\nu \in N$ can be a *synonym* (i.e., it shares its meaning with other names in N), or a *homonym* (i.e., it can have several meanings), or even both. We proceed from the plausible assumption that each name ν has finitely many meanings and that k names have at least k meanings. It is claimed that a *pairing-off* exists; that is, an injection $f : N \to M$ that associates to each ν one of its original meanings.

For finite N, the claim will be proved by induction on the number n of elements of N. It is trivial for $n = 1$. Now let $n > 1$ and assume that the claim holds for all k-element sets of names whenever $k < n$.

Case 1: For each k ($< n$): k names in N have at least $k + 1$ distinct meanings. Then to an arbitrarily chosen ν from N, assign one of its meanings a to it so that from the names out of $N\setminus\{\nu\}$ any k names still have at least k meanings $\neq a$. By the induction hypothesis there is a pairing-off for $N\setminus\{\nu\}$ that together with the ordered pair (ν, a) yields a pairing-off for the whole of N.

Case 2: There is some k-element $K \subseteq N$ $(0 < k < n)$ such that the set M_K of all meanings of the $\nu \in K$ has only k members. Every $\nu \in K$ can be assigned its meaning from M_K by the induction hypothesis. From the names in $N\setminus K$ any i names ($i \leqslant n - k$) still have i meanings not in M_K, as is not hard to see. By the induction hypothesis there is also a pairing-off for $N\setminus K$ with a set of values disjoint from M_K. Joining the two obviously results in a pairing-off for the whole of N.

We will now prove the claim for arbitrary sets of names N: assign to each pair $(\nu, a) \in N \times M$ a variable $p_{\nu,a}$ and consider the set of formulas

$$X: \quad \begin{cases} p_{\nu,a} \vee \cdots \vee p_{\nu,e} & (\nu \in N,\ a, \ldots, e \text{ the meanings of } \nu), \\ \neg(p_{\nu,x} \wedge p_{\nu,y}) & (\nu \in N,\ x, y \in M,\ x \neq y). \end{cases}$$

If $w \vDash X$, then we obtain a pairing-off for N by $f(\nu) = b \Leftrightarrow w \vDash p_{\nu,b}$. But every finite $X_0 \subseteq X$ has a model, because only finitely many names appear in it as indices. This case was already covered, thus proving that X has a model.

5. The ultrafilter theorem.

This theorem is of fundamental significance in topology (from which it originally stems), model theory, set theory, and elsewhere. Let I be any nonempty set. A nonempty collection of sets $F \subseteq \mathfrak{P}I$ is called a *filter on* I if for all $M, N \subseteq I$,

$$\text{(a) } M, N \in F \Rightarrow M \cap N \in F, \qquad \text{(b) } M \in F \;\&\; M \subseteq N \Rightarrow N \in F.$$

As is easily verified, (a) and (b) are equivalent to just a single condition, namely

$$(\cap) \quad M \cap N \in F \;\Leftrightarrow\; M \in F \text{ and } N \in F.$$

Since $F \neq \emptyset$, (b) shows that always $I \in F$. For fixed $K \subseteq I$, $\{J \subseteq I \mid J \supseteq K\}$ is a filter, the *principal filter* generated by K. It is a *proper filter* provided $K \neq \emptyset$, which in general is to mean a filter with $\emptyset \notin F$. Another example on an infinite I is the set of all *cofinite* subsets $M \subseteq I$, i.e., $\neg M$ $(= I \setminus M)$ is finite. This holds because $M_1 \cap M_2$ is cofinite iff M_1, M_2 are both cofinite, so that $(\cap)$ is satisfied.

A filter F is said to be an *ultrafilter on* I provided it satisfies, in addition,

$$(\neg) \quad \neg M \in F \;\Leftrightarrow\; M \notin F.$$

Ultrafilters on an infinite set I containing all cofinite subsets are called *nontrivial*. That such ultrafilters exist will be shown below. It is nearly impossible to describe them more closely. Roughly speaking, "we know they exist but we cannot see them." A trivial ultrafilter on I contains at least one finite subset. $\{J \subseteq I \mid i_0 \in J\}$ is an example for each $i_0 \in I$, also called a *principal* ultrafilter. All trivial ultrafilters are of this form; Exercise 3. Thus, trivial and principal ultrafilters coincide. In particular, each ultrafilter on a finite set I is trivial in this sense.

Each proper filter F obviously satisfies the assumption of the following theorem and can thereby be extended to an ultrafilter.

Ultrafilter theorem. *Every subset $F \subseteq \mathfrak{P}I$ can be extended to an ultrafilter U on a set I, provided $M_0 \cap \cdots \cap M_n \neq \emptyset$ for all n and all $M_0, \ldots M_n \in F$.*

Proof. Consider along with the propositional variables p_J $(J \subseteq I)$ the formula set

$$X: \quad p_{M\cap N} \leftrightarrow p_M \wedge p_N, \quad p_{\neg M} \leftrightarrow \neg p_M, \quad p_J \quad (M, N \subseteq I,\ J \in F).$$

Let $w \vDash X$. Then $(\cap), (\neg)$ are valid for $U := \{J \subseteq I \mid w \vDash p_J\}$, hence U is an ultrafilter and also $F \subseteq U$. It therefore suffices to show that every finite subset of X has a model, for which it is in turn enough to prove the ultrafilter theorem for finite F. But this is easy: let $F = \{M_0, \ldots, M_n\}$, $D := M_0 \cap \cdots \cap M_n$, and $i_0 \in D$. Then $U = \{J \subseteq I \mid i_0 \in J\}$ is an ultrafilter with $U \supseteq F$. ❏

Exercises

1. Prove (using the compactness theorem) that every partial order $\leqslant_0$ on a set M can be extended to a total order $\leqslant$ on M.

2. Let F be a proper filter on I $(\neq \emptyset)$. Show that F is an ultrafilter if and only if $M \cup N \in F \;\Leftrightarrow\; M \in F$ or $N \in F$.

3. Show that an ultrafilter U on I is trivial iff there is an $i_0 \in I$ such that $U = \{J \supseteq I \mid i_0 \in J\}$. Thus, each ultrafilter on a finite set I is of this form.

1.6 Hilbert Calculi

In a certain sense the simplest logical calculi are so-called *Hilbert calculi*. They are based on tautologies selected to play the role of *logical axioms*; this selection is, however, rather arbitrary and depends considerably on the logical signature. They use rules of inference like, for example, modus ponens MP: $\alpha, \alpha \to \beta / \beta$[6]. An advantage of these calculi consists in the fact that formal proofs, defined below as certain finite sequences, are immediately rendered intuitive. This advantage will pay off above all in the arithmetization of proofs in **6.2**.

In the following we consider such a calculus with MP as the only rule of inference; we denote this calculus for the time being by $\vdash$, in order to distinguish it from the calculus $\vdash$ of **1.4**. The logical signature contains just $\neg, \wedge$. In the axioms of $\vdash$, however, we will also use implication defined by $\alpha \to \beta := \neg(\alpha \wedge \neg\beta)$, thus considerably shortening the writing down of the axioms.

The *logical axiom scheme* of our calculus consists of the set Λ of all formulas of the following form (not forgetting the right association of parentheses):

$\Lambda 1 \quad (\alpha \to \beta \to \gamma) \to (\alpha \to \beta) \to \alpha \to \gamma,$ $\qquad$ $\Lambda 2 \quad \alpha \to \beta \to \alpha \wedge \beta,$

$\Lambda 3 \quad \alpha \wedge \beta \to \alpha, \quad \alpha \wedge \beta \to \beta,$ $\qquad$ $\Lambda 4 \quad (\alpha \to \neg\beta) \to \beta \to \neg\alpha.$

Λ consists only of tautologies. Moreover, all formulas derivable from Λ using MP are tautologies as well, because $\vDash \alpha, \alpha \to \beta$ implies $\vDash \beta$. We will show that *all* 2-valued tautologies are provable from Λ by means of MP.

Definition. A *proof* from X (in $\vdash$) is a sequence $\Phi = (\varphi_0, \ldots, \varphi_n)$ such that for every $k \leqslant n$ either $\varphi_k \in X \cup \Lambda$ or there exist indices $i, j < k$ such that $\varphi_j = \varphi_i \to \varphi_k$ (i.e., φ_k results from applying MP to terms of Φ proceeding φ_k). A proof $(\varphi_0, \ldots, \varphi_n)$ with $\varphi_n = \alpha$ is called *a proof of* α *from* X of length n. When such a proof exists we write $X \vdash \alpha$ and say that α is *provable* or *derivable from* X.

Example. $(p, q, p \to q \to p \wedge q, q \to p \wedge q, p \wedge q)$ is a proof of $p \wedge q$ from $X = \{p, q\}$. The last two terms in this sequence derive with MP from the previous ones, which are members of $X \cup \Lambda$.

Since a proof contains only finitely many formulas, the preceding definition leads immediately to the finiteness theorem for $\vdash$, formulated correspondingly to Theorem 4.1. Every proper initial segment of a proof is obviously a proof itself. Moreover, concatenating proofs of α and $\alpha \to \beta$ and adjoining β to the resulting sequence will produce a proof for β, as is plain to see. This observation implies

$$(*) \quad X \vdash \alpha, \alpha \to \beta \;\Rightarrow\; X \vdash \beta.$$

[6] Putting it crudely, this notation should express the fact that β is held to be proved from a formula set X when α and $\alpha \to \beta$ are provable from X. Modus ponens is an example of a binary Hilbert-style rule; for a general definition of this type of rule see, for instance, [Ra1].

In short, the set of all formulas derivable from X is *closed under* MP. In applying the property $(*)$ we will often say "MP yields..." It is easily seen that $X \vdash\!\sim \alpha$ iff α belongs to the smallest set containing X and is closed under MP. For the arithmetization of proofs and for automated theorem proving, however, it is more appropriate to base derivability on the finitary notion of a proof. Fortunately, the following theorem relieves us of the necessity to verify a property of formulas α derivable from X each time by induction on the length of a proof of α from X.

Theorem 6.1 (Induction principle for $\vdash\!\sim$). *Let X be given and $\mathcal{E}$ be a property of formulas. Then $\mathcal{E}$ holds for all α with $X \vdash\!\sim \alpha$, provided*

(o) *$\mathcal{E}$ holds for all $\alpha \in X \cup \Lambda$,* (s) *$\mathcal{E}\alpha$ and $\mathcal{E}(\alpha \to \beta)$ imply $\mathcal{E}\beta$, for all α, β.*

Proof by induction on the length n of a proof Φ of α from X. If $\alpha \in X \cup \Lambda$ then $\mathcal{E}\alpha$ holds by (o), which applies in particular if $n = 0$. If $\alpha \notin X \cup \Lambda$ then $n > 1$ and Φ contains members φ_i and $\varphi_j = \varphi_i \to \alpha$ both having proofs of length $< n$. Hence $\mathcal{E}\varphi_i$ and $\mathcal{E}\varphi_j$ by the induction hypothesis, and so $\mathcal{E}\alpha$ by (s). ❑

An application of Theorem 6.1 is the proof that $\vdash\!\sim \subseteq \vDash$, or more explicitly

$$X \vdash\!\sim \alpha \;\Rightarrow\; X \vDash \alpha \qquad (\textit{soundness}).$$

To see this let $\mathcal{E}\alpha$ be the property '$X \vDash \alpha$' for fixed X. Certainly, $X \vDash \alpha$ holds for $\alpha \in X$. The same is true for $\alpha \in \Lambda$. Thus, $\mathcal{E}\alpha$ for all $\alpha \in X \cup \Lambda$, and (o) is confirmed. Now let $X \vDash \alpha, \alpha \to \beta$; then so too $X \vDash \beta$, thus confirming the inductive step (s). By Theorem 6.1, $\mathcal{E}\alpha$ (that is, $X \vDash \alpha$) holds for all α with $X \vdash\!\sim \alpha$.

Unlike the proof of completeness for $\vdash$, the one for $\vdash\!\sim$ requires a whole series of derivations to be undertaken. This is in accordance with the nature of things: to get Hilbert calculi up and running one must often begin with drawn-out derivations.

In the following, we shall use without further comment the evident monotonicity property $X' \supseteq X \vdash\!\sim \alpha \;\Rightarrow\; X' \vdash\!\sim \alpha$, where as usual, $\vdash\!\sim \alpha$ stands for $\emptyset \vdash\!\sim \alpha$.

Lemma 6.2. (a) $X \vdash\!\sim \alpha \to \neg\beta \;\Rightarrow\; X \vdash\!\sim \beta \to \neg\alpha$, (b) $\vdash\!\sim \alpha \to \beta \to \alpha$,
(c) $\vdash\!\sim \alpha \to \alpha$, (d) $\vdash\!\sim \alpha \to \neg\neg\alpha$, (e) $\vdash\!\sim \beta \to \neg\beta \to \alpha$.

Proof. (a): Clearly $X \vdash\!\sim (\alpha \to \neg\beta) \to \beta \to \neg\alpha$ by Axiom $\Lambda 4$. From this and from $X \vdash\!\sim \alpha \to \neg\beta$ the claim is derived by MP. (b): By $\Lambda 3$ $\vdash\!\sim \beta \wedge \neg\alpha \to \neg\alpha$, and so with (a) $\vdash\!\sim \alpha \to \neg(\beta \wedge \neg\alpha) = \alpha \to \beta \to \alpha$. (c): From $\gamma := \alpha$, $\beta := \alpha \to \alpha$ in $\Lambda 1$ we obtain $\vdash\!\sim (\alpha \to (\alpha \to \alpha) \to \alpha) \to (\alpha \to \alpha \to \alpha) \to \alpha \to \alpha$, which gives together with (b) the claim by applying MP twice; (d) then follows from (a) using $\vdash\!\sim \neg\alpha \to \neg\alpha$. (e): Due to $\vdash\!\sim \neg\beta \wedge \neg\alpha \to \neg\beta$ and (a), we get $\vdash\!\sim \beta \to \neg(\neg\beta \wedge \neg\alpha) = \beta \to \neg\beta \to \alpha$. ❑

Part (e) of this lemma immediately yields that $\vdash\!\sim$ satisfies the rule $(\neg 1)$ of **1.4**, and hence $X \vdash\!\sim \beta, \neg\beta \Rightarrow X \vdash\!\sim \alpha$. Because of $\Lambda 2, \Lambda 3$, $\vdash\!\sim$ also satisfies $(\wedge 1)$ and $(\wedge 2)$. After

some preparation we will show that $(\neg 2)$ holds for $\vdash$ as well, thereby obtaining the desired completeness result. A crucial step in the completeness proof is

Lemma 6.3 (Deduction theorem). *$X, \alpha \vdash \gamma$ implies $X \vdash \alpha \to \gamma$.*

Proof by induction in $\vdash$ with a given set of premises X, α. Let $X, \alpha \vdash \gamma$, and let $\mathcal{E}\gamma$ now mean '$X \vdash \alpha \to \gamma$'. To prove (o), let $\gamma \in \Lambda \cup X \cup \{\alpha\}$. If $\gamma = \alpha$ then clearly $X \vdash \alpha \to \gamma$ by Lemma 6.2(c). If $\gamma \in X \cup \Lambda$ then certainly $X \vdash \gamma$. Because also $X \vdash \gamma \to \alpha \to \gamma$ by Lemma 6.2(b), MP yields $X \vdash \alpha \to \gamma$, thus proving (o). To show (s) let $X, \alpha \vdash \beta$ and $X, \alpha \vdash \beta \to \gamma$, so that $X \vdash \alpha \to \beta, \alpha \to \beta \to \gamma$ by the induction hypothesis. Applying MP to $\Lambda 1$ twice yields $X \vdash \alpha \to \gamma$, thus confirming (s). Therefore, by Theorem 6.1, $\mathcal{E}\gamma$ for all γ, which completes the proof. ❏

Lemma 6.4. *$\vdash \neg\neg\alpha \to \alpha$.*

Proof. By $\Lambda 3$ and MP we have $\neg\neg\alpha \wedge \neg\alpha \vdash \neg\alpha, \neg\neg\alpha$. Choose any τ with $\vdash \tau$. The already-proved rule $(\neg 1)$ gives $\neg\neg\alpha \wedge \neg\alpha \vdash \neg\tau$, and Lemma 6.3 $\vdash \neg\neg\alpha \wedge \neg\alpha \to \neg\tau$. From Lemma 6.2(a) it follows that $\vdash \tau \to \neg(\neg\neg\alpha \wedge \neg\alpha)$. But $\vdash \tau$, so using MP we obtain $\vdash \neg(\neg\neg\alpha \wedge \neg\alpha)$ and the latter formula is the same as $\neg\neg\alpha \to \alpha$. ❏

Lemma 6.5. *$\vdash$ satisfies rule $(\neg 2)$, i.e., $X, \beta \vdash \alpha$ and $X, \neg\beta \vdash \alpha$ imply $X \vdash \alpha$.*

Proof. Suppose $X, \beta \vdash \alpha$ and $X, \neg\beta \vdash \alpha$; then also $X, \beta \vdash \neg\neg\alpha$ and $X, \neg\beta \vdash \neg\neg\alpha$ by Lemma 6.2(d). Hence, $X \vdash \beta \to \neg\neg\alpha, \neg\beta \to \neg\neg\alpha$ (Lemma 6.3), and so $X \vdash \neg\alpha \to \neg\beta$ and $X \vdash \neg\alpha \to \neg\neg\beta$ by Lemma 6.2(a). Thus, MP yields $X, \neg\alpha \vdash \neg\beta, \neg\neg\beta$, whence $X, \neg\alpha \vdash \neg\tau$ by $(\neg 1)$, with τ as in Lemma 6.4. Consequently $X \vdash \neg\alpha \to \neg\tau$, due to Lemma 6.3, and therefore $X \vdash \tau \to \neg\neg\alpha$ by Lemma 6.2(a). Since $X \vdash \tau$ it follows that $X \vdash \neg\neg\alpha$ and so eventually $X \vdash \alpha$ by Lemma 6.4. ❏

Theorem 6.6 (Completeness theorem). *$\vdash = \models$.*

Proof. By soundness, $\vdash \subseteq \models$. Since $\vdash$ satisfies all basic rules of $\vdash$, it follows that $\vdash \subseteq \vdash$. Since $\vdash$ and $\models$ coincide (Theorem 4.6), we get also $\models \subseteq \vdash$. ❏

From this follows in particular $\vdash \varphi \Leftrightarrow \models \varphi$. In short, using MP one obtains from the axiom system Λ exactly the two-valued tautologies.

Remark 1. It may be something of a surprise that $\Lambda 1$–$\Lambda 4$ are sufficient to obtain all propositional tautologies, because these axioms and all formulas derivable from them using MP are collectively valid in intuitionistic and minimal logic. That Λ permits the derivation of all tautologies is based on the fact that $\to$ was *defined*. Had $\to$ been considered as a primitive connective this would no longer have been the case. To see this, alter the interpretation of $\neg$ by setting $\neg 0 = \neg 1 = 1$. While one here indeed obtains the value 1 for every valuation of the axioms of Λ and formulas derived from them using MP, one does not do so for $\neg\neg p \to p$, which therefore cannot be derived. Modifying the two-valued matrix or using many-valued logical matrices is a widely applied method to obtain independence results for logical axioms.

Thus, there are various calculi to derive tautologies or other semantical properties of $\models$. Clearly, simple relations like $\alpha \to \beta \models (\gamma \to \alpha) \to (\gamma \to \beta)$ can be confirmed without recourse to $\vdash$ or $\mid\sim$, for instance with the semantical deduction theorem.

Using Hilbert calculi one can axiomatize other two- and many-valued logics, for example the functional incomplete fragment in Exercise 3. The fragment in $\wedge$, $\vee$ which, while having no tautologies, contains interesting Hilbert-style rules, is also axiomatizable through finitely many such rules. The proof is not as easy as might be expected; at least nine Hilbert rules are required. Exercise 4 treats the somewhat simpler case of the fragment in $\vee$ alone. This calculus is based on unary rules only which simplifies the matter, but the completeness proof is still nontrivial.

Remark 2. Each of the infinitely many fragments of two-valued logic with or without tautologies is axiomatizable by a Hilbert calculus using *finitely many* Hilbert-style rules of its respective language; cf. [HeR]. In some of these calculi the method of enlarging a consistent set to a maximally consistent one has to be modified, Exercise 2. Besides sequent and Hilbert-style calculi there are still other types of logical calculi; for example, various tableau calculi which are above all significant for their generalizations to nonclassical logical systems. Related to tableau calculi is the resolution calculus dealt with in **4.2**.

Exercises

1. Prove the completeness of the Hilbert calculus $\vdash$ in $\mathcal{F}\{\to, \bot\}$ with MP as the sole rule of inference, the definition $\neg\alpha := \alpha \to \bot$, and the axioms

   $$\text{A1: } \alpha \to \beta \to \alpha, \quad \text{A2: } (\alpha \to \beta \to \gamma) \to (\alpha \to \beta) \to \alpha \to \gamma, \quad \text{A3: } \neg\neg\alpha \to \alpha.$$

2. Let $\vdash$ be a finitary consequence relation and let $X \nvdash \varphi$. Use Zorn's lemma to prove that there is a *φ-maximal* $Y \supseteq X$, that is, $Y \nvdash \varphi$ but $Y, \alpha \vdash \varphi$ whenever $\alpha \notin Y$. Such a Y is deductively closed but need not be maximally consistent.

3. Let $\vdash$ denote the calculus in $\mathcal{F}\{\to\}$ with the rule of inference MP, the axioms A1, A2 from Exercise 1, and the Peirce axiom $((\alpha \to \beta) \to \alpha) \to \alpha$. Verify that (a) a φ-maximal set X is maximally consistent, (b) $\vdash$ is complete in $\mathcal{F}\{\to\}$.

4. Show the completeness of the calculus $\vdash$ in $\mathcal{F}\{\vee\}$ with the four unary Hilbert-style rules below. Since $\vee$ is the only connective, its writing has been omitted:

   $$(1)\ \alpha/\alpha\beta, \quad (2)\ \alpha\alpha/\alpha, \quad (3)\ \alpha\beta/\beta\alpha, \quad (4)\ \alpha(\beta\gamma)/(\alpha\beta)\gamma.$$

   Note that (5) $(\alpha\beta)\gamma/\alpha(\beta\gamma)$ is derivable because application of (3) and (4) yields $(\alpha\beta)\gamma \vdash \gamma(\alpha\beta) \vdash (\gamma\alpha)\beta \vdash \beta(\gamma\alpha) \vdash (\beta\gamma)\alpha \vdash \alpha(\beta\gamma)$. Crucial for completeness is the proof of "monotonicity" (m): $\alpha \vdash \beta \Rightarrow \alpha\gamma \vdash \beta\gamma$. (m) implies (M): $X, \alpha \vdash \beta \Rightarrow X, \alpha\gamma \vdash \beta\gamma$, proving first that a calculus $\vdash$ based solely on unary rules obeys $X \vdash \beta \Rightarrow \alpha \vdash \beta$ for some $\alpha \in X$.

Chapter 2

Predicate Logic

Mathematics and some other disciplines like computer science often consider domains of individuals in which certain relations and operations are singled out. When we use the language of propositional logic, our ability to talk about the properties of such relations and operations is very limited. Thus, it is necessary to refine our linguistic means of expression, in order to procure new possibilities of description. To this end, one needs not only logical symbols but also variables for the individuals of the domain being considered, as well as a symbol for equality and symbols for the relations and operations in question. Predicate logic is the part of logic that subjects properties of such relations and operations to logical analysis.

Linguistic particles as "for all" and "there exists" (called *quantifiers*), play a central role here whose analysis should be based on a well prepared semantical background. Hence, we first consider mathematical structures and classes of structures. Some of these are relevant both to logic (especially to model theory) and to computer science. Neither the newcomer nor the advanced student need to read all of Section **2.1** with its mathematical flavor at once. The first four pages should suffice. The reader may continue with **2.2** and later return to what is needed.

Next we home in on the most important class of formal languages, the *first-order* or *elementary languages.* Their main characteristic is a restriction of the quantification possibilities. We discuss in detail the semantics of these languages and arrive at a notion of *logical consequence* from arbitrary premises. In this context, the notion of a formalized theory is made more precise.

Finally, we treat the introduction of new notions by explicit definitions and other expansions of a language, for instance by Skolem functions. Not until Chapter **3** do we talk about methods of formal logical deduction. While a lot of technical details have to be considered in this chapter, nothing is especially profound. Anyway, most of it is important for the undertakings of the subsequent chapters.

2.1 Mathematical Structures

By a *structure* $\mathcal{A}$ we understand a nonempty set A together with certain distinguished relations and operations of A, as well as certain constants distinguished therein. The set A is also termed the *domain* of $\mathcal{A}$, or *universe*. The distinguished relations, operations, and constants are called the *(basic) relations, operations, and constants of* $\mathcal{A}$. A *finite structure* is one with a finite domain. An easy example is $(\{0,1\}, \wedge, \vee, \neg)$. Here $\wedge$, $\vee$, $\neg$ have their usual meanings on the domain $\{0,1\}$, and no distinguished relations or constants occur. An *infinite structure* has an infinite domain. $\mathcal{A} = (\mathbb{N}, <, +, \cdot, 0, 1)$ is an example with the domain $\mathbb{N}$; here $<$, $+$, $\cdot$, 0, 1 have again their ordinary meaning.

Without having to say so every time, for a structure $\mathcal{A}$ the corresponding letter A will always denote the domain of $\mathcal{A}$; similarly B denotes the domain of $\mathcal{B}$, etc. If $\mathcal{A}$ contains no operations or constants, then $\mathcal{A}$ is also called a *relational structure*. If $\mathcal{A}$ has no relations it is termed an *algebraic structure*, or simply an *algebra*. For example, $(\mathbb{Z}, <)$ is a relational structure, whereas $(\mathbb{Z}, +, 0)$ is an algebraic structure, the *additive group* $\mathbb{Z}$ (it is customary using here the symbol $\mathbb{Z}$ as well). Also the set of propositional formulas from **1.1** can be understood as an algebra, equipped with the operations $(\alpha, \beta) \mapsto (\alpha \wedge \beta)$, $(\alpha, \beta) \mapsto (\alpha \vee \beta)$, and $\alpha \mapsto \neg\alpha$. Thus, one may speak of the *formula algebra* $\mathcal{F}$ whenever wanted.

Despite our interest in specific structures, whole classes of structures are also often considered. For instance, the class of all groups, of rings, fields, vector spaces, Boolean algebras, and so on. Even when initially just a single structure is viewed, call it the paradigm structure, one often needs to talk about similar structures in the same breath, in *one* language, so to speak. This can be achieved by setting aside the concrete meaning of the relation and operation symbols in the paradigm structure and considering the symbols in themselves, creating thereby a formal language that enables one to talk at once about all structures relevant to a topic. Thus, one distinguishes in this context clearly between denotation and what is denoted. To emphasize this distinction, for instance for a structure $\mathcal{A} = (A, +, <, 0)$, one better writes $\mathcal{A} = (A, +^A, <^A, 0^A)$, where $+^A$, $<^A$ and 0^A mean the relation, operation, and constant denoted by $+, <, 0$ in A. Still more precise is writing $+^{\mathcal{A}}$, $<^{\mathcal{A}}$, $0^{\mathcal{A}}$ for $+^A$, $<^A$ and 0^A, respectively. In this way we are free to talk on the one hand about the structure $\mathcal{A}$ and on the other hand about the symbols $+, <, 0$.

A finite or infinite set L resulting in this way, consisting of relation, operation and constant symbols of given arity, is called an extralogical *signature*. For the class of all groups (see page 38), $L = \{\circ, e\}$ exemplifies a favored signature; that is, one often considers groups as structures of the form $(G, \circ, e)$, where $\circ$ denotes the group operation and e the unit element. But one can also define groups as structures of

the signature $\{\circ\}$, because e is definable in terms of $\circ$ as we shall see later. Of course, instead of $\circ$, the operation symbol could be chosen as $\cdot$, $*$, or $+$ (mainly used in connection with commutative groups and semigroups, page 38). In this sense, the actual appearance of a symbol is less important; what matters is its arity.

$r \in L$ always means that r is a relation symbol, and $f \in L$ that f is an operation symbol, each time of some arity $n > 0$, which of course depends on the symbols r and f, respectively.[1] An *L-structure* is a pair $\mathcal{A} = (A, L^{\mathcal{A}})$, where $L^{\mathcal{A}}$ contains for every $r \in L$ a relation $r^{\mathcal{A}}$ on A of the same arity as r, for every $f \in L$ an operation $f^{\mathcal{A}}$ on A of the arity of f, and for every $c \in L$ a constant $c^{\mathcal{A}} \in A$. We may omit the superscripts, provided it is clear from the context which operation or relation on A is meant. We occasionally abbreviate also the notation of certain structures. For instance, we sometimes speak of the ring $\mathbb{Z}$ or the field $\mathbb{R}$.

Every structure is an L-structure for a certain signature, namely that consisting of the symbols for its relations, functions, and constants. But this does not make the name L-structure superfluous. Basic concepts, such as isomorphism, substructure, etc. each refer to structures of the same signature. From **2.2** on, once the elementary language $\mathcal{L}$ belonging to L has been defined, L-structures will mostly be called $\mathcal{L}$-structures. We then also often say that r, f, or c belongs to $\mathcal{L}$ instead of L.

If $A \subseteq B$ and f is an n-ary operation on B then A is *closed* under f, briefly *f-closed*, if $f\vec{a} \in A$ for all $\vec{a} \in A^n$. If $n = 0$, i.e., if f is a constant c, this simply means $c \in A$. The intersection of any nonempty family of f-closed subsets of B is itself f-closed. Accordingly, we can talk of the smallest (the intersection) of all f-closed subsets of B that contain a given subset $E \subseteq B$. All of this extends in a natural way if f is here replaced by an arbitrary family of operations of B.

Example. For a given positive m, the set $m\mathbb{Z} := \{m \cdot n \mid n \in \mathbb{Z}\}$ of integers divisible by m is closed in $\mathbb{Z}$ under $+$, $-$, and $\cdot$, and is in fact the smallest such subset of $\mathbb{Z}$ containing m.

The *restriction* of an n-ary relation $r^B \subseteq B^n$ to a subset $A \subseteq B$ is $r^A = r^B \cap A^n$. For instance, the restriction of the standard order of $\mathbb{R}$ to $\mathbb{N}$ is the standard order of $\mathbb{N}$. Only because of this fact can the same symbol be used to denote these relations. The restriction f^A of an operation f^B on B to a set $A \subseteq B$ is defined analogously whenever A is f-closed. Simply let $f^A\vec{a} = f^B\vec{a}$ for $\vec{a} \in A^n$. For instance, addition in $\mathbb{N}$ is the restriction of addition in $\mathbb{Z}$ to $\mathbb{N}$, or addition in $\mathbb{Z}$ is an extension of this operation in $\mathbb{N}$. Again, only this state of affairs allows us to denote the two operations by the same symbol.

[1] Here r and f represent the general case and look differently in a concrete situation. They are sometimes also called predicate and function symbols respectively, in particular in the unary case. In special contexts, we also admit $n = 0$, regarding constants as 0-ary operations.

Let $\mathcal{B}$ be an L-structure and $A \subseteq B$ be nonempty and closed under all operations of $\mathcal{B}$; this will be taken to include $c^{\mathcal{B}} \in A$ for constant symbols $c \in L$. To such a subset A corresponds in a natural way an L-structure $\mathcal{A} = (A, L^{\mathcal{A}})$, where $r^{\mathcal{A}}$ and $f^{\mathcal{A}}$ for $r, f \in L$ are the restrictions of $r^{\mathcal{B}}$ respectively $f^{\mathcal{B}}$ to A. Finally, let $c^{\mathcal{A}} = c^{\mathcal{B}}$ for $c \in L$. The structure $\mathcal{A}$ so defined is then called a *substructure* of $\mathcal{B}$, and $\mathcal{B}$ is called an *extension* of $\mathcal{A}$, symbolically $\mathcal{A} \subseteq \mathcal{B}$. This notation is some abuse of the set-theoretical symbol $\subseteq$ but it does not cause confusion since the arguments indicate what is meant. $\mathcal{A} \subseteq \mathcal{B}$ implies $A \subseteq B$ but not conversely, in general.

For example, $\mathcal{A} = (\mathbb{N}, <, +, 0)$ is a substructure of $\mathcal{B} = (\mathbb{Z}, <, +, 0)$ since $\mathbb{N}$ is closed under addition in $\mathbb{Z}$ and 0 has the same meaning in $\mathcal{A}$ and $\mathcal{B}$. Similarly, if further relations or operations are considered. Note that we omitted the superscripts for $<$, $+$, and 0 since there is no risk of misunderstanding.

A nonempty subset G of the domain B of an L-structure $\mathcal{B}$ defines a smallest substructure $\mathcal{A}$ of $\mathcal{B}$ containing G, whose domain A is the smallest subset of B that contains G and is closed under all operations of B. $\mathcal{A}$ is called the substructure *generated from G in $\mathcal{B}$*. For instance, $3\mathbb{N}$ $(= \{3n \mid n \in \mathbb{N}\})$ is the domain of the substructure generated from $G = \{3\}$ in $(\mathbb{N}, +, 0)$, since $3\mathbb{N}$ contains 0 and 3, is closed under $+$, and is clearly the smallest such subset of $\mathbb{N}$. A structure $\mathcal{A}$ is called *finitely generated* if for some finite $G \subseteq A$ the substructure generated from G in $\mathcal{A}$ coincides with $\mathcal{A}$. For instance, $(\mathbb{Z}, +, -, 0)$ is finitely generated by $G = \{1\}$.

If $\mathcal{A}$ is an L-structure and $L_0 \subseteq L$ then the L_0-structure $\mathcal{A}_0$ with domain A and where $\zeta^{\mathcal{A}_0} = \zeta^{\mathcal{A}}$ for all symbols $\zeta \in L_0$ is termed the *L_0-reduct of $\mathcal{A}$*, and $\mathcal{A}$ is called an *L-expansion* of $\mathcal{A}_0$. For instance, the group $(\mathbb{Z}, +, 0)$ is the $\{+, 0\}$-reduct of the ordered ring $(\mathbb{Z}, <, +, \cdot, 0)$. The notions reduct and substructure must clearly be distinguished. A reduct of $\mathcal{A}$ has always the same domain as $\mathcal{A}$.

We now list some frequently cited properties of a binary relation R in a set A. It is convenient to write $a \lhd b$ and $a \not\lhd b$ instead of $(a, b) \in R$ and $(a, b) \notin R$, respectively. Also, $a \lhd b \lhd c$ stands for $a \lhd b \;\&\; b \lhd c$, just as $a < b < c$ is usually written in place of $a < b \;\&\; b < c$. In the listing, "for all a" and "there exists an a" more precisely mean "for all $a \in A$" and "there exists an $a \in A$," where A is a given set. Thus, everything below refers to a given A. The relation $\lhd \subseteq A^2$ is called

reflexive	if	$a \lhd a$ for all a,
irreflexive	if	$a \not\lhd a$ for all a,
symmetric	if	$a \lhd b \;\Rightarrow\; b \lhd a$, for all a, b,
antisymmetric	if	$a \lhd b \lhd a \;\Rightarrow\; a = b$, for all a, b,
transitive	if	$a \lhd b \lhd c \;\Rightarrow\; a \lhd c$, for all a, b, c,
connex	if	$a = b$ or $a \lhd b$ or $b \lhd a$, for all a, b.

Reflexive, transitive, and symmetric relations are called *equivalence relations*. These are often denoted by $\sim$, $\approx$, $\equiv$, $\simeq$, or similar symbols.

The following properties of a binary operation $\circ$ on a given set A will often be referred to. The operation $\circ$ is

commutative if $a \circ b = b \circ a$ for all a, b,
associative if $a \circ (b \circ c) = (a \circ b) \circ c$ for all a, b, c,
idempotent if $a \circ a = a$ for all a,
invertible if for all a, b there are $x, y \in A$ with $a \circ x = b$ and $y \circ a = b$.

We now present an overview of classes of structures that we will later refer back to, mainly in Chapter **5**. Hence, for the time being, the beginner may skip to **2.2**.

1. Graphs, partial orders, and orders. A relational structure $(A, \lhd)$ with some binary relation $\lhd$ on A is often termed a (directed) *graph*. If $\lhd$ is irreflexive and transitive we usually write $<$ for $\lhd$ and speak of a *partially ordered set* or a *strict* (= *irreflexive*) *partial order*. If we define $\leqslant$ by $x \leqslant y :\Leftrightarrow x < y$ or $x = y$, then $\leqslant$ is reflexive, transitive, and antisymmetric, called a *reflexive partial order* (the one corresponding to $<$). Starting with a reflexive partial order on A and defining $x < y :\Leftrightarrow x \leqslant y \;\&\; x \neq y$, then $<$ is a strict partial order on A as is easily seen.

A connex partial order $\mathcal{A} = (A, <)$ is called a *total* or *linear* order, mostly termed an *order* or *ordered set*. $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ are examples with respect to their standard orders. Here we follow the habit of referring to ordered sets by their domains only.

Let U be a nonempty subset of some ordered set A such that for all $a, b \in A$ with $a < b$ and $b \in U$ also $a \in U$, called an *initial segment* of A, and let $V := A \setminus U \neq \emptyset$. If U has no largest and V no smallest element we say that the pair (U, V) is a *gap* in A. If U *has* a largest element a, and V a smallest element b, then (U, V) is called a *jump*. b is then called *the immediate successor* of a, and a *the immediate predecessor* of b, for there is no element from A between a and b. An infinite ordered set without gaps and jumps like $\mathbb{R}$ is said to be *continuously ordered*. Such a set is easily seen to be *densely ordered*, i.e., between any two elements lies another one.

A totally ordered subset K of a partially ordered set H is called a *chain* in H. Such a K is said to be *bounded* if there is some $b \in H$ such that $a \leqslant b$ for all $a \in K$. Call $c \in H$ *maximal* in H if no $a \in H$ exists with $a > c$. An infinite partial order need not contain a maximal element, nor need all chains be bounded, as seen by the example $(\mathbb{N}, <)$. With these notions, an important mathematical tool can now be stated, used already in Theorem 1.4.8.

Zorn's lemma. *If every chain in a nonempty partially ordered set H is bounded then H has a maximal element.*

An ordered set A is *well-ordered* if every nonempty subset of A has a smallest element; equivalently, there are no infinite decreasing sequences $a_0 > a_1 > \cdots$ of elements from A. Clearly, every finite ordered set is well-ordered. The simplest example of an infinite well-ordered set is $\mathbb{N}$ together with its standard order.

2. Groupoids, semigroups, and groups. Algebras $\mathcal{A} = (A, \circ)$ with an operation $\circ\colon A^2 \to A$ are termed *groupoids*. If $\circ$ is associative then $\mathcal{A}$ is called a *semigroup*, and if $\circ$ is additionally invertible then $\mathcal{A}$ is said to be a *group*. It is provable that a group $(G, \circ)$ in this sense contains exactly one *unit element*, that is, an element e such that $x \circ e = e \circ x = x$ for all $x \in G$, also called a *neutral element*. A well-known example is the group of bijections of a set $M \neq \emptyset$. If $\circ$ is commutative in addition, then we speak of a *commutative* or *abelian* group, also called a *module*.

Here are some examples of semigroups that are not groups: (a) the set of strings on some alphabet A with respect to concatenation, the *word-semigroup* or *free semigroup generated from* A. (b) the set M^M of mappings from M to itself with respect to composition. (c) $(\mathbb{N}, +)$ and $(\mathbb{N}, \cdot)$; these two are commutative semigroups. With the exception of $(M^M, \circ)$, all mentioned examples of semigroups are *regular*, which is to mean $x \circ y = x \circ z \;\Rightarrow\; y = z$ and $x \circ z = y \circ z \;\Rightarrow\; x = y$, for all x, y, z.

Substructures of semigroups are again semigroups. Substructures of groups are in general only semigroups, as seen from $(\mathbb{N}, +) \subseteq (\mathbb{Z}, +)$. Not so in the signature $\{\circ, e, {}^{-1}\}$, where e denotes the unit element and x^{-1} the inverse of x. Here all substructures are subgroups. The reason is that in $\{\circ, e, {}^{-1}\}$, the group axioms can be written as universally quantified equations where, for brevity, we omit the writing of "for all x, y, z," namely as $x \circ (y \circ z) = (x \circ y) \circ z$, $x \circ e = x$, $x \circ x^{-1} = e$. These equations certainly retain their validity in the transition to substructures. We mention that from these three equations, $e \circ x = x$ and $x^{-1} \circ x = e$ are derivable, although $\circ$ is not supposed to be commutative.

Ordered semigroups and groups possess along with $\circ$ some order, with respect to which $\circ$ is monotonic in both arguments, like $(\mathbb{N}, +, 0, \leqslant)$. A commutative ordered semigroup $(A, +, 0, \leqslant)$ with zero element 0, which at the same time is the smallest element in A, and where $a \leqslant b$ iff there is some c with $a + c = b$, is called a *domain of magnitude*. Everyday examples are the domains of *length, mass, money*, etc.

3. Rings and fields. Because these are among the most commonly known structures, we do not repeat their definition here. The ring axioms are formalized in $+, -, \cdot, 0$ and include the axiom $x + (y - x) = y$. For fields, the constant symbol 1 is adjoined. Removing the last-mentioned axiom from the list of ring axioms and the minus symbol from the signature leaves us with the notion of a *semiring*.

Substructures of fields in the signature $\{0, 1, +, -, \cdot\}$ are *integral domains*. These are commutative rings without zero-divisors and with 1. Let $\mathcal{K}, \mathcal{K}'$ be fields with $\mathcal{K} \subset \mathcal{K}'$. We call $a \in K' \setminus K$ *algebraic* or *transcendental* on $\mathcal{K}$, depending on whether a is a zero of a polynomial with coefficients in K or not. If every polynomial of degree $\geqslant 1$ with coefficients in K breaks down into linear factors, as is the case for the field of complex numbers, then $\mathcal{K}$ is called *algebraically closed*, in short, $\mathcal{K}$ is a.c. These fields will be more closely inspected in **3.3** and Chapter **5**. Each field

$\mathcal{K}$ has a smallest subfield $\mathcal{P}$, called a *prime field*. One says that $\mathcal{K}$ *has characteristic* 0 or p (a prime number), depending on whether $\mathcal{P}$ is isomorphic to the field $\mathbb{Q}$ or the finite field of p elements. No other prime fields exist. It is not hard to show that $\mathcal{K}$ has the characteristic p iff the sentence $\mathtt{char}_p : \underbrace{1+\cdots+1}_{p} = 0$ holds in $\mathcal{K}$.

Semirings, rings, and fields can also be *ordered*, whereby the usual monotonicity laws are required. For example, $(\mathbb{Z}, <, +, \cdot, 0, 1)$ is the *ordered ring* of integers and $(\mathbb{N}, <, +, \cdot, 0, 1)$ the *ordered semiring* of natural numbers.

4. Semilattices and lattices. $\mathcal{A} = (A, \circ)$ is called a *semilattice* if $\circ$ is associative, commutative, and idempotent. An example is $(\{0,1\}, \wedge)$. In what follows, we omit the writing of the operation symbol. If we define $a \leqslant b :\Leftrightarrow ab = a$ then $\leqslant$ is a reflexive partial order on A. Reflexivity holds, since $aa = a$. As can be easily verified, ab is in fact the *infimum* of a, b with respect to $\leqslant$, $ab = \inf\{a, b\}$, that is, $ab \leqslant a, b$ and $c \leqslant a, b$ imply $c \leqslant ab$, for all $a, b, c \in A$.

$\mathcal{A} = (A, \cap, \cup)$ is called a *lattice* if $(A, \cap)$ and $(A, \cup)$ are both semilattices and the following so-called *absorption laws* hold: $a \cap (a \cup b) = a$ and $a \cup (a \cap b) = a$. These imply $a \cap b = a \Leftrightarrow a \cup b = b$. As above, $a \leqslant b :\Leftrightarrow a \cap b = a$ defines a partial order such that $a \cap b = \inf\{a, b\}$. In addition, $a \cup b = \sup\{a, b\}$ (the *supremum* of a, b), which is to mean $a, b \leqslant a \cup b$ and $a, b \leqslant c \;\Rightarrow\; a \cup b \leqslant c$, for all $c \in A$. If $\mathcal{A}$ satisfies, moreover, the *distributive laws* $x \cap (y \cup c) = (x \cap y) \cup (x \cap c)$ and $x \cup (y \cap c) = (x \cup y) \cap (x \cup c)$, then $\mathcal{A}$ is termed a *distributive* lattice. For instance, the power set $\mathfrak{P}M$ with the operations $\cap$ and $\cup$ for $\cap$ and $\cup$ respectively, is a distributive lattice, as is every nonempty family of subsets of M closed under $\cap$ and $\cup$, a so-called *lattice of sets*. Another important example is $(\mathbb{N}, \gcd, \mathrm{lcm})$. Here $\gcd(a, b)$ and $\mathrm{lcm}(a, b)$ denote the greatest common divisor and the least common multiple of $a, b \in \mathbb{N}$.

5. Boolean algebras. An algebra $\mathcal{A} = (A, \cap, \cup, \neg)$ where $(A, \cap, \cup)$ is a distributive lattice and in which at least the equations

$$\neg\neg x = x, \quad \neg(x \cap y) = \neg x \cup \neg y, \quad x \cap \neg x = y \cap \neg y$$

are valid is called a *Boolean algebra*. A paradigm structure is the two-element Boolean algebra $\mathcal{2} := (\{0,1\}, \wedge, \vee, \neg)$, with $\cap, \cup$ interpreted as $\wedge, \vee$, respectively. In the general case, one defines $0 := a \cap \neg a$ for any $a \in A$ and $1 := \neg 0$. There are many ways to characterize Boolean algebras $\mathcal{A}$, for instance, by saying that $\mathcal{A}$ satisfies all equations valid in $\mathcal{2}$. The signature can also be variously selected. For example, the signature $\wedge, \vee, \neg$ is well suited to deal algebraically with two-valued propositional logic. Terms of this signature are, up to the denotation of variables, precisely the Boolean formulas from **1.1**, and a logical equivalence $\alpha \equiv \beta$ corresponds to the equation $\alpha = \beta$, valid in all Boolean algebras. Further examples are the *Boolean algebras of sets* $\mathcal{A} = (A, \cap, \cup, \neg)$. Here A consists of a nonempty system of

subsets of a set I, closed under $\cap, \cup$ and $\neg$, where $\neg$ denotes complementation in I. These are the most general examples; a famous theorem, *Stone's representation theorem*, says that each Boolean algebra is isomorphic to an algebra of sets.

6. Logical L-matrices. These are structures $\mathcal{A} = (A, L^{\mathcal{A}}, D^{\mathcal{A}})$, where L contains just operation symbols (the "logical" symbols) and D denotes a unary predicate, *the set of distinguished values* of $\mathcal{A}$. Best known is the two-valued *Boolean matrix* $\mathcal{B} = (\mathit{2}, D^{\mathcal{B}})$ with $D^{\mathcal{B}} = \{1\}$. The consequence relation $\models_{\mathcal{A}}$ in the propositional language $\mathcal{F}$ with signature L is defined as in the two-valued case: Let $X \subseteq \mathcal{F}$ and $\alpha \in \mathcal{F}$. Then $X \models_{\mathcal{A}} \alpha$ if $w\alpha \in D^{\mathcal{A}}$ for every $w\colon PV \to A$ with $wX \subseteq D^{\mathcal{A}}$. In words, if the values of all $\varphi \in X$ are distinguished, then so too is the value of α.

Homomorphisms and isomorphisms. The following notions are important for both mathematical and logical investigations, mainly in Chapter **5**.

Definition. Let $\mathcal{A}, \mathcal{B}$ be L-structures and $h\colon \mathcal{A} \to \mathcal{B}$ (strictly speaking $h\colon A \to B$) a mapping such that for all $f, c, r \in L$ and $\vec{a} \in A^n$ ($n > 0$ is the arity of f or r),

$$(\star) \quad hf^{\mathcal{A}}\vec{a} = f^{\mathcal{B}}h\vec{a}, \quad hc^{\mathcal{A}} = c^{\mathcal{B}}, \quad r^{\mathcal{A}}\vec{a} \Rightarrow r^{\mathcal{B}}h\vec{a} \qquad \big(h\vec{a} := (ha_1, \dots, ha_n)\big).$$

Then h is called a *homomorphism*. If the third condition in $(\star)$ is replaced by the stronger condition $(\exists \vec{b} \in A^n)(h\vec{a}{=}h\vec{b}\ \&\ r^{\mathcal{A}}\vec{b}) \Leftrightarrow r^{\mathcal{B}}h\vec{a}$ [2] then h is said to be a *strong homomorphism* (for algebras, the word "strong" is dispensable). An injective strong homomorphism $h\colon \mathcal{A} \to \mathcal{B}$ is called an *embedding* of $\mathcal{A}$ into $\mathcal{B}$. If, in addition, h is bijective then h is called an *isomorphism*, and in case $\mathcal{A} = \mathcal{B}$, an *automorphism*.

An embedding or isomorphism $h\colon \mathcal{A} \to \mathcal{B}$ satisfies $r^{\mathcal{A}}\vec{a} \Leftrightarrow r^{\mathcal{B}}h\vec{a}$ as is easily seen. $\mathcal{A}, \mathcal{B}$ are said to be *isomorphic*, in symbols $\mathcal{A} \simeq \mathcal{B}$, if there is an isomorphism from $\mathcal{A}$ to $\mathcal{B}$. It is readily verified that $\simeq$ is reflexive, symmetric, and transitive.

Examples. (a) A valuation w considered in **1.1** can be regarded as a homomorphism of the propositional formula algebra $\mathcal{F}$ onto the two-element Boolean algebra $\mathit{2}$.
(b) Let $\mathcal{A} = (A, *)$ be a word semigroup with the concatenation operation $*$ and $\mathcal{B}$ the additive semigroup of natural numbers. These are L-structures for $L = \{\circ\}$ with $\circ^{\mathcal{A}} = *$ and $\circ^{\mathcal{B}} = +$. Let $\mathrm{lh}(w)$ denote the length of a word $w \in A$. Then $w \mapsto \mathrm{lh}(w)$ is a homomorphism since $\mathrm{lh}(w * w') = \mathrm{lh}(w) + \mathrm{lh}(w')$, for all $w, w' \in A$. If $\mathcal{A}$ is generated from just one letter, lh is evidently bijective, hence an isomorphism.
(c) The mapping $a \mapsto (a, 0)$ from $\mathbb{R}$ to $\mathbb{C}$ (= set of complex numbers, understood as ordered pairs of real numbers) is a paradigm of an embedding, here of the field $\mathbb{R}$ into the field $\mathbb{C}$. Nonetheless, in this and similar cases, we are used to saying that $\mathbb{R}$ is a subfield of $\mathbb{C}$, and that $\mathbb{R}$ is a subset of $\mathbb{C}$.

[2] $(\exists \vec{b} \in A^n)(h\vec{a}{=}h\vec{b}\ \&\ r^{\mathcal{A}}\vec{b})$ abbreviates 'there is some $\vec{b} \in A^n$ with $h\vec{a} = \vec{b}$ and $r^{\mathcal{A}}\vec{b}$'. If $h\colon \mathcal{A} \to \mathcal{B}$ is onto (and only this case will occur in our examples and applications) then the stronger condition is equivalent to the more suggestive condition $r^{\mathcal{B}} = \{h\vec{a} \mid r^{\mathcal{A}}\vec{a}\}$.

(d) Let $\mathcal{A} = (\mathbb{R}, +, <)$ be the ordered additive group of reals and $\mathcal{B} = (\mathbb{R}_+, \cdot, <)$ the multiplicative group of positive reals. Then for any $b \in \mathbb{R}_+ \setminus \{1\}$ there is precisely one isomorphism $\eta : \mathcal{A} \to \mathcal{B}$ such that $\eta 1 = b$, namely $\eta : x \mapsto b^x$, the exponential function $\exp_b$ to the base b. Indeed, η runs through every value in $\mathbb{R}_+$ exactly once, and $\eta(x + y) = \eta x \cdot \eta y$ holds for all $x, y \in \mathbb{R}$. One could even define $\exp_b$ as this isomorphism, by first proving that—up to isomorphism—there is only one continuously ordered abelian group (probably first noticed in [Ta4]).
(e) The algebras $\mathcal{A} = (\{0, 1\}, +)$ and $\mathcal{B} = (\{0, 1\}, \leftrightarrow)$ are only apparently different, but are in fact isomorphic, as shown by the isomorphism δ where $\delta 0 = 1$, $\delta 1 = 0$. Thus, since $\mathcal{A}$ is a group, $\mathcal{B}$ is a group as well, which is perhaps not so obvious (see the proof in **2.3**). By adjoining the unary predicate $D = \{1\}$, $\mathcal{A}$ and $\mathcal{B}$ become (nonisomorphic) logical matrices. These actually define the two "dual" fragmentary two-valued logics for the connectives *... if and only if ...* and *either ... or ...*

Congruences. A *congruence relation* (or simply a *congruence*) in a structure $\mathcal{A}$ of signature L is an equivalence relation $\approx$ in A such that for all $f \in L$ of arity n,

$(*)$ $\vec{a} \approx \vec{b} \Rightarrow f^A\vec{a} \approx f^A\vec{b}$, $\quad (\vec{a}, \vec{b} \in A^n;\ \vec{a} \approx \vec{b}$ means $a_i \approx b_i$ for $i = 1, \dots, n)$.

Let A' be the set of *equivalence classes* $a/\!\approx\ := \{x \in A \mid a \approx x\}$ for $a \in A$, also called the *congruence classes* of $\approx$, and set $\vec{a}/\!\approx\ := (a_1/\!\approx, \dots, a_n/\!\approx)$ for $\vec{a} \in A^n$. Define $f^{A'}(\vec{a}/\!\approx) := (f^A\vec{a})/\!\approx$ and let $r^{A'}\vec{a}/\!\approx :\Leftrightarrow (\exists \vec{b} \approx \vec{a}) r^A\vec{b}$. These definitions are *sound*, that is, independent of the choice of the n-tuple $\vec{a}$ of representatives. Then A' becomes an L-structure $\mathcal{A}'$, the *factor structure of* $\mathcal{A}$ *modulo* $\approx$, also denoted by $\mathcal{A}/\!\approx$. Interesting, in particular for Chapter **5**, is the following general

Homomorphism theorem. *Let $\mathcal{A}$ and $\mathcal{B}$ be L-structures and $\approx$ a congruence in $\mathcal{A}$. Then $k : a \mapsto a/\!\approx$ is a strong homomorphism from $\mathcal{A}$ onto $\mathcal{A}/\!\approx$, the canonical homomorphism. Conversely, if $h : \mathcal{A} \to \mathcal{B}$ is a strong homomorphism onto $\mathcal{B}$ then $\approx\ \subseteq A^2$, defined by $a \approx b \Leftrightarrow ha = hb$, is a congruence in $\mathcal{A}$, called the kernel of h; moreover, $\imath : a/\!\approx\ \mapsto ha$ is an isomorphism from $\mathcal{A}/\!\approx$ to $\mathcal{B}$, and $h = \imath \circ k$.*

Proof. We omit here the superscripts for f and r for the sake of faster legibility. Clearly, $kf\vec{a} = (f\vec{a})/\!\approx\ = f(\vec{a}/\!\approx) = fk\vec{a}$ $\big(= f(ka_1, \dots, ka_n)\big)$, and by our definitions, $(\exists \vec{b} \in A^n)(k\vec{a} = k\vec{b}\ \&\ r\vec{b}) \Leftrightarrow (\exists \vec{b} \approx \vec{a}) r\vec{b} \Leftrightarrow r\,\vec{a}/\!\approx\ \Leftrightarrow r\,k\vec{a}$. Hence k is what we claimed. The definition of $\imath$ is sound. $\imath$ is obviously bijective. Furthermore, the isomorphism conditions hold: $\imath f(\vec{a}/\!\approx) = hf\vec{a} = fh\vec{a} = f\imath(\vec{a}/\!\approx)$ and $r\,\vec{a}/\!\approx\ \Leftrightarrow r\,h\vec{a} \Leftrightarrow r\,\imath(\vec{a}/\!\approx)$. Finally, h is the composition $\imath \circ k$ according to the definitions of $\imath$ and k. ❑

For algebras $\mathcal{A}, \mathcal{B}$, this theorem is the usual homomorphism theorem of universal algebra. It covers groups, rings, etc. In groups, the kernel of a homomorphism is already determined by the congruence class of the unit element, called a *normal subgroup*, in rings by the congruence class of 0, called an *ideal*. Hence, in textbooks on basic algebra the homomorphism theorem may look somewhat differently.

Direct products. These provide the basis for many constructions of new structures, especially in **5.7**. A well-known example is the n-dimensional vector group $(\mathbb{R}^n, 0, +)$. This is the n-fold direct product of the group $(\mathbb{R}, 0, +)$ with itself. The addition in $\mathbb{R}^n$ is defined componentwise, as is also the case in the following

Definition. Let $(\mathcal{A}_i)_{i\in I}$ be a nonempty family of L-structures. The *direct product* $\mathcal{B} = \prod_{i\in I} \mathcal{A}_i$ is the following structure. Its domain is $B = \prod_{i\in I} A_i$, called the *direct product* of the sets A_i, whose elements $a = (a_i)_{i\in I}$ are functions defined on I with $a_i \in A_i$ for $i \in I$. Relations and operations are defined componentwise, that is,

$$r^{\mathcal{B}}\vec{a} \;\Leftrightarrow\; r^{\mathcal{A}_i}\vec{a}_i \text{ for all } i \in I, \quad f^{\mathcal{B}}\vec{a} = (f^{\mathcal{A}_i}\vec{a}_i)_{i\in I}, \quad c^{\mathcal{B}} = (c^{\mathcal{A}_i})_{i\in I},$$

where $\vec{a} = (a^1, \ldots, a^n) \in B^n$ (here the superscripts count the components of the n-tuple), $a^\nu := (a_i^\nu)_{i\in I}$ for $\nu = 1, \ldots, n$, and $\vec{a}_i := (a_i^1, \ldots, a_i^n) \in A_i^n$. The sequence $\vec{a}_i$ $(\in A_i^n)$ is called the *ith projection* of the n-tuple $\vec{a}$. For $I = \{1, \ldots, m\}$, the product $\prod_{i\in I} \mathcal{A}_i$ is also written as $\mathcal{A}_1 \times \cdots \times \mathcal{A}_m$. Whenever $\mathcal{A}_i = \mathcal{A}$ for all $i \in I$, then $\prod_{i\in I} \mathcal{A}_i$ is denoted by $\mathcal{A}^I$ and called a *direct power* of the structure $\mathcal{A}$.

If $I = \{0, \ldots, n-1\}$ one mostly writes $\mathcal{A}^n$ for $\mathcal{A}^I$. Note that $\mathcal{A}$ is embedded in $\mathcal{A}^I$ by $a \mapsto (a)_{i\in I}$, where $(a)_{i\in I}$ is the I-tuple with the constant value a.

Examples. (a) For $I = \{1, 2\}$ and $\mathcal{A}_i = (A_i, <^i)$, $a <^{\mathcal{B}} b \Leftrightarrow a_1 <^1 b_1 \;\&\; a_2 <^2 b_2$, for all $a, b \in B = A_1 \times A_2$. Note that if $\mathcal{A}_1, \mathcal{A}_2$ are orders then $\mathcal{B}$ is only a partial order. The deeper reason for this observation will become clear in Chapter **5**.
(b) Let $\mathcal{B} = \mathcal{2}^I$ be a direct power of the two-element Boolean algebra $\mathcal{2}$. The elements $a \in B$ are I-tuples of zeros and ones that uniquely correspond to the subsets of I via the mapping $\imath\colon a \mapsto I_a := \{i \in I \mid a_i = 1\}$. As a matter of fact, $\imath$ is an isomorphism from $\mathcal{B}$ to $(\mathfrak{P}I, \cap, \cup, \neg)$ as can readily be verified.

Exercises

1. Show that there are (up to isomorphism) exactly five two-element proper groupoids. Here a groupoid $(H, \cdot)$ is termed *proper* if $\cdot$ is essentially binary.

2. $\approx$ $(\subseteq A^2)$ is termed *Euclidean* if $a \approx b \;\&\; a \approx c \;\Rightarrow\; b \approx c$, for all $a, b, c \in A$. Show that $\approx$ is an equivalence relation in A iff $\approx$ is reflexive and Euclidean.

3. Prove that an equivalence relation $\approx$ on an algebraic L-structure $\mathcal{A}$ is already a congruence, if for all $f \in L$ of arity n and all $i = 1, \ldots, n$ holds

$$a \approx a' \;\Rightarrow\; f(a_1, \ldots, a_{i-1}, a, a_{i+1}, \ldots, a_n) \approx f(a_1, \ldots, a_{i-1}, a', a_{i+1}, \ldots, a_n).$$

4. Show that $h\colon \prod_{i\in I} \mathcal{A}_i \to \mathcal{A}_j$ with $ha = a_j$ is a homomorphism for each $j \in I$.

2.2 Syntax of Elementary Languages

Standard mathematical language enables us to talk precisely about structures, like the field of real numbers. However, for logical (and metamathematical) issues it is important to delimit the theoretical framework to be considered; this is achieved most simply by means of a formalization. In this way one obtains an *object language*; that is, the formalized elements of the language, like the components of a structure, are *objects* of our consideration. To formalize interesting properties of a structure in this language, one requires at least variables for the elements of its domain, also called *individual variables*. Further, a sufficient number of logical symbols, along with symbols for the relations, functions, and constants of the structure, which together constitute the *extralogical* signature L of the language to be defined.

In this manner one arrives at the *first-order languages*, also termed *elementary* languages. Nothing is lost in terms of generality if the set of variables is the same for all elementary languages; we denote this set by *Var* and take it to consist of the countably many symbols $\boldsymbol{v}_0, \boldsymbol{v}_1, \ldots$ Two such languages therefore differ only in the choice of their extralogical symbols. Variables for subsets of the domain are consciously excluded, since languages containing variables both for individuals and sets of these individuals—second-order languages, discussed in **3.7**—have different semantic properties than those investigated here.

We first determine the *alphabet*, the set of *basic symbols* of a first-order language determined by a signature L. It includes, of course, the variables $\boldsymbol{v}_0, \boldsymbol{v}_1, \ldots$ In what follows, the latter will mostly be denoted by x, y, z, u, v, though in some cases other letters with or without indices may serve the same purpose. The boldface printed variables are useful in writing down a formula in the variables $\boldsymbol{v}_{i_1}, \ldots, \boldsymbol{v}_{i_n}$, for these can then be denoted, for instance, by $v_1, \ldots, v_n$, or by $x_1, \ldots, x_n$.

Further, the logical symbols $\wedge$ (and), $\neg$ (not), $\forall$ (*for all*), the equality sign $\boldsymbol{=}$, and last but not least, all symbols of the extralogical signature L should belong to the alphabet.[3] Note that here the boldface symbol $\boldsymbol{=}$ is taken as a basic symbol; simply taking $=$ could lead to unintended mix-ups with the metamathematical use of the equality symbol $=$. Finally, the parentheses (,) are included in the alphabet. Additional logical symbols will be introduced later, including the symbols $\exists$ (*there exists*) and $\exists!$ (*there exists exactly one*).

From the set of all strings of these basic symbols we pick out the meaningful ones according to certain rules, beginning with terms. A term, under an interpretation of the language, will always denote an element of a domain, provided an assignment is given of the occurring variables to elements of that domain. In order to keep the syntax simple, terms will be parenthesis-free strings.

[3] Sometimes *identity-free* languages without $\boldsymbol{=}$ will be considered, for instance in Chapter **4**.

Terms in L:

(T1) Variables and constants are terms, called *prime terms*.

(T2) If $f \in L$ is n-ary and $t_1, \ldots, t_n$ are terms, then $ft_1 \cdots t_n$ is a term.

This is an inductive definition in the set of strings on the alphabet of $\mathcal{L}$, that is, any string that is not generated by (T1) and (T2) is not a term in this context (cf. the related definition of $\mathcal{F}$ in **1.1**). Parenthesis-free term notation simplifies the syntax, but for binary operations we proceed in practice otherwise. We write, for example, the term $\cdot + xyz$ as $(x + y) \cdot z$ because high density of information in the notation complicates reading. Our brain does not process information sequentially like a computer. Officially, terms are parenthesis-free, and the parenthesized notation is just an alternative way of rewriting terms. Similarly to the unique reconstruction property in **1.1**, here the *unique term reconstruction* holds (Exercise 2):

$$ft_1 \cdots t_n = fs_1 \cdots s_n \text{ implies } s_i = t_i \text{ for } i = 1, \ldots, n \quad (t_i, s_i \text{ terms}).$$

Let $\mathcal{T}$ $(= \mathcal{T}_L)$ denote the set of all terms of a given signature L. Variable-free terms, which can exist only with the availability of constant symbols, are called *constant terms* or *ground terms*, mainly in logic programming. With the operations given in $\mathcal{T}$ by $f^{\mathcal{T}}(t_1, \ldots, t_n) = ft_1 \cdots t_n$, $\mathcal{T}$ forms an algebra, the *term algebra*. From the definition of terms immediately arises the following useful

Principle of proof by term induction. *If $\mathcal{E}$ is a property of strings such that $\mathcal{E}t$ for all prime terms, and for each $n > 0$ and each n-ary function symbol f $\mathcal{E}t_1, \ldots, \mathcal{E}t_n$ implies $\mathcal{E}ft_1 \cdots t_n$, then all terms have the property $\mathcal{E}$.*

Indeed, $\mathcal{T}$ is by definition the smallest set of strings satisfying the conditions of this principle. Hence, $\mathcal{T}$ is a subset of the set of all strings with the property $\mathcal{E}$. It seems to be obvious that a compound term t is a *function term* in the sense that $t = ft_1 \cdots t_n$ for some n-ary function symbol f and some terms $t_1, \ldots, t_n$. But the critical reader may feel more comfortable after verifying this by term induction, considering the property $\mathcal{E}$: 't is either prime or a function term'.

We also have at our disposal a *definition principle* by term induction which, rather than defining it generally, we demonstrate through examples. The set $\mathit{var}\, t$ of variables occurring in a term t is inductively defined by

$$\mathit{var}\, c = \emptyset \ ; \ \mathit{var}\, x = \{x\} \ ; \ \mathit{var}\, ft_1 \cdots t_n = \mathit{var}\, t_1 \cup \cdots \cup \mathit{var}\, t_n.$$

Clearly, this definition makes sense only in view of the unique term reconstruction. $\mathit{var}\, t$ (and even $\mathit{var}\, \xi$ for any string ξ) can also easily be explicitly defined using concatenation, namely as $\mathit{var}\, t := \{x \in \mathit{Var} \mid \text{there are strings } \xi_0, \xi_1 \text{ with } t = \xi_0 x \xi_1\}$.

The notion of a *subterm* of a term can also inductively be defined. Again, we can do it more briefly using concatenation. We now define inductively those strings of the alphabet $\mathcal{L}$ to be denoted as *formulas*, also called (predicate logic) *expressions* or *well-formed formulas*. Certain formulas will later be called *sentences*.

Formulas in L:

(F1) If s, t are terms, then the string $s = t$ is a formula.

(F2) If $t_1, \ldots, t_n$ are terms and $r \in L$ is n-ary, then $rt_1 \cdots t_n$ is a formula.

(F3) If α, β are formulas and $x \in \mathit{Var}$, then $(\alpha \wedge \beta)$, $\neg\alpha$, and $\forall x\alpha$ are formulas.

Any string not generated according to (F1), (F2), (F3) is in this context not a formula. Other logical symbols serve throughout merely as abbreviations; namely $\exists x\alpha := \neg\forall x\neg\alpha$, and as in **1.1**, $\alpha \vee \beta := \neg(\neg\alpha \wedge \neg\beta)$, $\alpha \to \beta := \neg(\alpha \wedge \neg\beta)$, and $\alpha \leftrightarrow \beta := (\alpha \to \beta) \wedge (\beta \to \alpha)$.

Examples. (a) $\forall x \exists y\, x + y = 0$ (more explicitly, $\forall x \neg \forall y \neg\, x + y = 0$) is a formula. Here we assume tacitly that x, y denote distinct original variables. The same is assumed in all of the following whenever this can be made out from the context.
(b) $\forall x \forall x\, x = y$ is a formula, since repeated "quantification" of the same variable is not forbidden. $\forall z\, x = y$ is a formula, although z does not appear in $x = y$.

Example (b) indicates that the grammar of our formal language is more liberal as one might expect. This will spare us a lot of writing. The formula $\forall x \forall x\, x = y$, as well as $\exists x \forall x\, x = y$, both have the same meaning as $\forall x\, x = y$. These three formulas are logically equivalent (in a sense still to be defined), as are $\forall z\, x = y$ and $x = y$. It would be to our disadvantage to require any restriction here. In spite of this liberality, the formula syntax corresponds roughly to the syntax of natural language.

The formulas procured by (F1) and (F2) are called *prime formulas* (or simply *prime*, also called *atomic*). Similar to unique term reconstruction holds the *unique prime formula reconstruction* $rt_1 \cdots t_n = rs_1 \cdots s_n \;\Rightarrow\; t_i = s_i$ for $i = 1, \ldots, n$. Prime formulas of the form $s = t$ are called *equations*. These are the only prime formulas if L contains no relation symbols, in which case L is called an *algebraic* signature. For $\neg\, s = t$ we henceforth write $s \neq t$.

Prime formulas that are not equations always begin with a relation symbol. In the binary case the relation symbol tends to separate the two arguments as, for example, in $x \leqslant y$. The official notation is, however, that of clause (F2). As in propositional logic, prime formulas and their negations will be called *literals*.

The set of all formulas in L is denoted by $\mathcal{L}$, and if $L = \{\in\}$ then $\mathcal{L}$ is also denoted by $\mathcal{L}_\in$. Analogously for similarly simple signatures. The case $L = \emptyset$ is also permitted; it defines the *language of pure identity*, denoted by $\mathcal{L}_=$.

Formulas in which $\forall, \exists$ do not occur are termed *quantifier-free* or *open*. These are precisely the Boolean combinations of prime formulas. The *Boolean combinations* of the formulas from $X \subseteq \mathcal{L}$ are those that can be generated by $\wedge$ and $\neg$ from the formulas in X. The strings $\forall x$ and $\exists x$ (read "for all x" respectively "there is an x") are called *prefixes* and may occasionally occur also in the metalanguage.

Instead of terms, formulas, and structures of the signature L, we will talk of $\mathcal{L}$-terms, $\mathcal{L}$-formulas, and $\mathcal{L}$-structures respectively. We also omit the prefix $\mathcal{L}$- if $\mathcal{L}$ has been given earlier. In writing down formulas, we use the same conventions of parenthesis economy as in **1.1**. We will also allow ourselves other informal aids in order to increase readability. For instance, variously shaped parentheses may be used as in $\forall x \exists y \forall z [z \in y \leftrightarrow \exists u (z \in u \wedge u \in x)]$. Even verbal descriptions (partial or total) are permitted, as long as the intended formula is uniquely recognizable.

X, Y, Z always denote sets of formulas, $\alpha, \beta, \gamma, \delta, \pi, \varphi, \ldots$ denote formulas, and s, t terms, while Φ, Ψ are reserved to denote finite sequences of formulas and formal proofs. Substitutions (to be defined below) will be denoted by $\sigma, \tau, \omega, \rho$, and ι.

Principles of proof by formula induction and *of definition by formula induction* also exist for first-order (and other formal) languages. After the explanation of inductive proofs and definitions on formulas in Chapter **1**, we do without a general formulation, preferring instead to use examples, adhering to the maxim verba docent, exempla trahunt. For example, define $\operatorname{rk}\varphi$, the *rank* of the formula φ, by $\operatorname{rk}\pi = 0$ for prime formulas π and

$$\operatorname{rk}(\alpha \wedge \beta) = \max\{\operatorname{rk}\alpha, \operatorname{rk}\beta\} + 1, \quad \operatorname{rk}\neg\alpha = \operatorname{rk}\forall x\alpha = \operatorname{rk}\alpha + 1.$$

Useful for some purposes is the *quantifier rank*, $\operatorname{qr}\varphi$. It represents a measure of nested quantifiers in a formula. For prime formulas π let $\operatorname{qr}\pi = 0$ and

$$\operatorname{qr}\neg\alpha = \operatorname{qr}\alpha, \ \ \operatorname{qr}(\alpha \wedge \beta) = \max\{\operatorname{qr}\alpha, \operatorname{qr}\beta\}, \ \ \operatorname{qr}\forall x\alpha = \operatorname{qr}\alpha + 1.$$

Note that $\operatorname{qr}\exists x\varphi = \operatorname{qr}\neg\forall x\neg\varphi = \operatorname{qr}\forall x\varphi$. A *subformula* of a formula is defined analogously to the definition in **1.1**. Hence, we need say no more on this. We write $x \in \mathit{bnd}\,\varphi$ (or x occurs bound in φ) if φ contains the prefix $\forall x$. In subformulas of φ of the form $\forall x\alpha$, the formula α is also called the *scope* of $\forall x$. The same prefix can occur repeatedly and with nested scopes in a formula, as in $\forall x(\forall x\, x = 0 \wedge x < y)$. In practice we avoid this writing, though for a computer this would pose no problem.

Intuitively, the formulas (a) $\forall x \exists y\, x + y = 0$ and (b) $\exists y\, x + y = 0$ are different in that the former is in every context with a given meaning for $+$ and 0 either true or false, whereas in (b) the variable x is waiting to be assigned a value. One also says that all the variables occurring in (a) are bound. (b) contains the "free" variable x. The syntactic predicate 'x occurs free in φ', or '$x \in \mathit{free}\,\varphi$' is defined inductively: Let $\mathit{free}\,\alpha = \mathit{var}\,\alpha$ for prime formulas α ($\mathit{var}\,\alpha$ was defined on page 44), and

$$\mathit{free}\,(\alpha \wedge \beta) = \mathit{free}\,\alpha \cup \mathit{free}\,\beta, \quad \mathit{free}\,\neg\alpha = \mathit{free}\,\alpha, \quad \mathit{free}\,\forall x\alpha = \mathit{free}\,\alpha \setminus \{x\}.$$

For example, $\mathit{free}\,(\forall x \exists z\, x + y = 0) = \emptyset$, and $\mathit{free}\,(x \leqslant y \wedge \forall x \exists y\, x + y = 0) = \{x, y\}$. As the last formula shows, x can occur both free and bound in a formula. This too will be avoided in practice whenever possible. In some proof-theoretically oriented presentations, even different symbols are chosen for free and bound variables. Each of these approaches has its advantages and its disadvantages.

Formulas without free variables are called *sentences*, or *closed formulas*. $1+1=0$ and $\forall x\exists y\, x+y=0$ $(=\forall x\neg\forall y\neg x+y=0)$ are examples. Throughout take $\mathcal{L}^0$ to denote the set of all sentences of $\mathcal{L}$. More generally, let $\mathcal{L}^k$ be the set of all formulas φ such that $\mathit{free}\,\varphi \subseteq \mathit{Var}_k := \{\boldsymbol{v}_0, \ldots, \boldsymbol{v}_{k-1}\}$. Clearly, $\mathcal{L}^0 \subseteq \mathcal{L}^1 \subseteq \cdots$ and $\mathcal{L} = \bigcup_{k\in\mathbb{N}} \mathcal{L}^k$.

At this point we meet an important and for the remainder of the book valid

Convention. As long as not otherwise stated, the notation $\varphi = \varphi(x)$ means that the formula φ contains at most x as a free variable; more generally, $\varphi = \varphi(x_1, \ldots, x_n)$ or $\varphi = \varphi(\vec{x})$ is to mean $\mathit{free}\,\varphi \subseteq \{x_1, \ldots, x_n\}$, where $x_1, \ldots, x_n$ stand for arbitrary but distinct variables. Not all of these variables need actually occur in φ. Further, $t = t(\vec{x})$ for terms t is to be read completely analogously.

The term $ft_1\cdots t_n$ is often denoted by $f\vec{t}$, the prime formula $rt_1\cdots t_n$ by $r\vec{t}$. Note that $\vec{t}$ denotes here the concatenation $t_1\cdots t_n$ of terms. $\vec{t}$ behaves like a sequence as was pointed out already, and has the unique readability property.

Substitutions. We begin with the substitution $\frac{t}{x}$ of some term t for a single variable x. Put intuitively, $\varphi\,\frac{t}{x}$ (read "φ t for x," also denoted by $\alpha_x(t)$), is the formula that results from replacing all free occurrences of the variable x in φ by the term t. This intuitive characterization is made precise inductively, first for terms by

$$x\,\tfrac{t}{x} = t, \qquad y\,\tfrac{t}{x} = y \quad (x \neq y), \qquad c\,\tfrac{t}{x} = c, \qquad (ft_1\cdots t_n)\,\tfrac{t}{x} = ft_1'\cdots t_n',$$

where, for brevity, t_i' denotes the term $t_i\,\frac{t}{x}$, and next for formulas as follows:

$$(t_1=t_2)\,\tfrac{t}{x} = t_1'=t_2', \qquad (r\vec{t}\,)\,\tfrac{t}{x} = rt_1'\cdots t_n',$$

$$(\alpha\wedge\beta)\,\tfrac{t}{x} = \alpha\,\tfrac{t}{x}\wedge\beta\,\tfrac{t}{x}, \quad (\neg\alpha)\,\tfrac{t}{x} = \neg(\alpha\,\tfrac{t}{x}), \quad (\forall y\alpha)\,\tfrac{t}{x} = \begin{cases}\forall y\alpha \text{ in case } x=y,\\ \forall y(\alpha\,\frac{t}{x}) \text{ otherwise.}\end{cases}$$

Then also $(\alpha\to\beta)\,\frac{t}{x} = \alpha\,\frac{t}{x}\to\beta\,\frac{t}{x}$, and likewise for $\vee$ and $\exists$, as can easily be checked.

Along with these *simple* substitutions $\frac{t}{x}$, also *simultaneous substitutions*

$$\varphi\,\tfrac{t_1\cdots t_n}{x_1\cdots x_n} \qquad (x_1, \ldots, x_n \text{ distinct})$$

are useful. This will briefly be written $\varphi\frac{\vec{t}}{\vec{x}}$ or $\varphi_{\vec{x}}(\vec{t}\,)$ or just $\varphi(\vec{t}\,)$, provided there is no danger of misunderstanding. Here the variables x_i are simultaneously replaced by the terms t_i. Simple and simultaneous substitutions are special cases of what is called a *global* substitution σ. Such a σ assigns to *every* variable x a term $x^\sigma \in \mathcal{T}$. It is extended to the whole of $\mathcal{T}$ by the clauses $c^\sigma = c$ and $(f\vec{t}\,)^\sigma = ft_1^\sigma\cdots t_n^\sigma$, and subsequently to the formula set $\mathcal{L}$, so that σ is defined for the whole of $\mathcal{T}\cup\mathcal{L}$:

$$(t_1=t_2)^\sigma = t_1^\sigma=t_2^\sigma, \quad (r\vec{t}\,)^\sigma = rt_1^\sigma\cdots t_n^\sigma, \quad (\alpha\wedge\beta)^\sigma = \alpha^\sigma\wedge\beta^\sigma, \quad (\neg\alpha)^\sigma = \neg\alpha^\sigma,$$

and $(\forall x\varphi)^\sigma = \forall x\varphi^\tau$, where the global substitution τ is defined by $x^\tau = x$ and $y^\tau = y^\sigma$ for $y \neq x$. The *identical substitution*, always denoted by ι, is defined by $x^\iota = x$ for all x, hence $t^\iota = t$ and $\varphi^\iota = \varphi$ for all terms t and formulas φ.

A simultaneous substitution $\frac{\vec{t}}{\vec{x}}$ can be understood as the global substitution σ with $x_i^\sigma = t_i$ for $i = 1, \ldots, n$ and $x^\sigma = x$ otherwise. This can also be stated as follows: simultaneous substitutions are those global substitutions σ such that $x^\sigma = x$ for

almost all variables x, i.e., with the exception of finitely many. This way of putting things makes it immediately clear that the composition $\sigma_1\sigma_2$ of two simultaneous substitutions—let $x^{\sigma_1\sigma_2} = (x^{\sigma_1})^{\sigma_2}$—is again a simultaneous substitution. It is hence obvious that these constitute a semigroup with the neutral element ι.

It always holds that $\frac{t_1 t_2}{x_1 x_2} = \frac{t_2 t_1}{x_2 x_1}$, whereas the compositions $\frac{t_1}{x_1}\frac{t_2}{x_2}$ and $\frac{t_2}{x_2}\frac{t_1}{x_1}$ are distinct, in general. Let us elaborate by explaining the difference between $\varphi\frac{t_1 t_2}{x_1 x_2}$ and $\varphi\frac{t_1}{x_1}\frac{t_2}{x_2}$ $\big(= (\varphi\frac{t_1}{x_1})\frac{t_2}{x_2}\big)$. For example, if one wants to swap x_1, x_2 at their free occurrences in φ then this is $\varphi\frac{x_2 x_1}{x_1 x_2}$, but not, in general, $\varphi\frac{x_2}{x_1}\frac{x_1}{x_2}$; choose $\varphi := x_1{<}x_2$, for instance. Rather, $\varphi\frac{x_2 x_1}{x_1 x_2} = \varphi\frac{y}{x_2}\frac{x_2}{x_1}\frac{x_1}{y}$ for any $y \notin var\,\varphi$ distinct from x_1, x_2 as is shown by induction on φ. In the same way we readily obtain

(1) $\varphi\frac{\vec{t}}{\vec{x}} = \varphi\frac{y}{x_n}\frac{t_1\cdots t_{n-1}}{x_1\cdots x_{n-1}}\frac{t_n}{y}$ $\qquad (y \notin var\,\varphi \cup var\,\vec{x} \cup var\,\vec{t},\ n \geqslant 2)$.

Thus, a simultaneous and even a global substitution therefore yields *locally*, that is, with respect to individual formulas, just the same as a suitable composition of simple substitutions. In some cases (1) can be simplified. Useful, for example, is the following equation, which holds in particular when all terms t_i are variable-free:

(2) $\varphi\frac{\vec{t}}{\vec{x}} = \varphi\frac{t_1}{x_1}\cdots\frac{t_n}{x_n}$, $\qquad$ (provided $x_i \notin var\,t_j$ for $i \neq j$).

In Chapter **4** we intensively operate with substitutions. Getting on correctly with substitutions is not altogether simple; it requires practice, because our ability to regard complex strings is not especially trustworthy. A computer is not only much faster but more reliable in this respect.

Exercises

1. Show by term induction that a terminal segment of a term t is a concatenation $s_1\cdots s_m$ of terms s_i for some $m \geqslant 1$. Thus, a symbol in t is at each position of its occurrence in t the initial symbol of a subterm s of t which is unique by Exercise 2(c). The same then holds for a concatenation $t_1\cdots t_n$ of terms.

2. Prove (a) no term is a concatenation of two or more terms, (b) no proper initial segment of a term t is a term, (c) the subterm s of t in Exercise 1 is unique, (d) the unique term concatenation: $t_1\cdots t_n = t'_1\cdots t'_m \ \Rightarrow\ m = n\ \&\ t_i = t'_i$ for $i = 1, \ldots, n$. The latter obviously implies the unique term reconstruction and the unique prime formula reconstruction property.

3. Prove $\varphi\frac{t}{x} = \varphi$ for $x \notin free\,\varphi$, and $\varphi\frac{y}{x}\frac{t}{y} = \varphi\frac{t}{x}$ for $y \notin var\,\varphi$. Show by means of examples that these restrictions are indispensable provided $t \neq x$.

4. Let ξ, η be strings over the alphabet of $\mathcal{L}$. Verify (a) $\neg\xi \in \mathcal{L} \ \Rightarrow\ \xi \in \mathcal{L}$, (b) $\xi\wedge\eta \in \mathcal{L} \Rightarrow \xi, \eta \in \mathcal{L}$, (c) $\xi\to\eta \in \mathcal{L} \Rightarrow \xi, \eta \in \mathcal{L}$.

5. Let $\mathrm{qr}\,\varphi = n > 0$. Show that φ is a Boolean combination of formulas α with $\mathrm{qr}\,\alpha < n$ and at least one formulas $\forall x\beta$ with $\mathrm{qr}\,\beta = n - 1$.

2.3 Semantics of Elementary Languages

Intuitively it is clear that the formula $\exists y\, y + y = x$ can be allocated a truth value in the domain $(\mathbb{N}, +)$ only if to the free variable x there corresponds a value in $\mathbb{N}$. Thus, along with an interpretation of the extralogical symbols, a truth value allocation for a formula φ requires a valuation of at least the variables occurring free in φ. However, it is technically more convenient to work with a global assignment of values to all variables, even if in a concrete case only the values of finitely many variables are needed. We therefore begin with the following

Definition. A *model* $\mathcal{M}$ is a pair $(\mathcal{A}, w)$ consisting of an $\mathcal{L}$-structure $\mathcal{A}$ and a *valuation* $w\colon Var \to A$, $w\colon x \mapsto x^w$. We denote $r^{\mathcal{A}}, f^{\mathcal{A}}, c^{\mathcal{A}}$, and x^w also by $r^{\mathcal{M}}, f^{\mathcal{M}}, c^{\mathcal{M}}$, and $x^{\mathcal{M}}$, respectively. The domain of $\mathcal{A}$ is also called the *domain of* $\mathcal{M}$.

Models are also called *interpretations*, or $\mathcal{L}$*-models* if the connection to $\mathcal{L}$ is to be highlighted. Some authors identify models with structures from the outset. This also happens in **2.5**, where we talk about models of theories. The notion of a model is to be maintained flexible in logic, and adapted according to requirements.

A model $\mathcal{M}$ allocates in a natural way to every term t a value in A, denoted by $t^{\mathcal{M}}$ or $t^{\mathcal{A},w}$ or just by t^w. For prime terms the value is already given by $\mathcal{M}$. This evaluation extends to compound terms by term induction as follows:

$$(f\vec{t})^{\mathcal{M}} = f^{\mathcal{M}}\vec{t}^{\,\mathcal{M}},$$

where $\vec{t}^{\,\mathcal{M}}$ abbreviates the sequence of values $(t_1^{\mathcal{M}}, \ldots, t_n^{\mathcal{M}})$. If the context allows we neglect the superscripts and retain just an imaginary distinction between symbols and their interpretation. For instance, if $\mathcal{A} = (\mathbb{N}, +, \cdot, 0, 1)$ and $x^w = 2$, say, we write $(0 \cdot x + 1)^{\mathcal{A},w} = 0 \cdot 2 + 1 = 1$. The value of t under $\mathcal{M}$ depends only on the meaning of the symbols that effectively occur in t; using induction on t the following slightly more general claim is obtained: if $var\, t \subseteq V \subseteq Var$ and $\mathcal{M}, \mathcal{M}'$ are models with the same domain such that $x^{\mathcal{M}} = x^{\mathcal{M}'}$ for all $x \in V$ and $\zeta^{\mathcal{M}} = \zeta^{\mathcal{M}'}$ for all remaining symbols ζ occurring in t, then $t^{\mathcal{M}} = t^{\mathcal{M}'}$. Clearly, $t^{\mathcal{A},w}$ may simply be denoted by $t^{\mathcal{A}}$ provided the term t contains no variables.

We also consider models that differ from a given $\mathcal{M} = (\mathcal{A}, w)$ only in the values of one or more variables. Let $x_1, \ldots, x_n$ be distinct and $w' := w^{\vec{a}}_{\vec{x}}$ be defined by $x_i^{w'} = a_i$ for $i = 1, \ldots, n$ and $x^{w'} = x^w$, for any variable x distinct from $x_1, \ldots, x_n$. Then put $\mathcal{M}^{\vec{a}}_{\vec{x}} := (\mathcal{A}, w^{\vec{a}}_{\vec{x}})$. In particular, $\mathcal{M}^a_x$ denotes $(\mathcal{A}, w^a_x)$. This model differs from $\mathcal{M}$ only in the value of the fixed variable x.

We now define a satisfiability relation $\models$ between models $\mathcal{M} = (\mathcal{A}, w)$ and formulas φ, using induction on φ as in **1.3**. We read $\mathcal{M} \models \varphi$ as $\mathcal{M}$ *satisfies* φ, or $\mathcal{M}$ *is a model for* φ. Sometimes $\mathcal{A} \models \varphi\,[w]$ is written for $\mathcal{M} \models \varphi$. A similar notation, just as frequently encountered, is introduced later. Each of these notations has its

advantages, depending on the context. If $\mathcal{M} \vDash \varphi$ for all $\varphi \in X$ we write $\mathcal{M} \vDash X$ and call $\mathcal{M}$ a *model for* X. For the formulation of the satisfaction clauses below (taken from [Ta1]) we consider for given $\mathcal{M} = (\mathcal{A}, w)$, $x \in Var$, and $a \in A$ also the model $\mathcal{M}_x^a$. It differs from $\mathcal{M}$ only in that x receives the value a instead of $x^{\mathcal{M}}$.

$$\begin{array}{lll} \mathcal{M} \vDash s = t & \Leftrightarrow & s^{\mathcal{M}} = t^{\mathcal{M}}, \\ \mathcal{M} \vDash r\vec{t} & \Leftrightarrow & r^{\mathcal{M}}\vec{t}^{\,\mathcal{M}}, \\ \mathcal{M} \vDash \alpha \wedge \beta & \Leftrightarrow & \mathcal{M} \vDash \alpha \text{ and } \mathcal{M} \vDash \beta, \\ \mathcal{M} \vDash \neg\alpha & \Leftrightarrow & \mathcal{M} \nvDash \alpha, \\ \mathcal{M} \vDash \forall x\alpha & \Leftrightarrow & \mathcal{M}_x^a \vDash \alpha \text{ for all } a \in A. \end{array}$$

Example 1. Let $\mathcal{M}' := \mathcal{M}_x^{t^{\mathcal{M}}}$. We claim that $\mathcal{M}' \vDash x = t$ if $x \notin var\, t$. In this case namely $t^{\mathcal{M}'} = t^{\mathcal{M}}$. Since also $x^{\mathcal{M}'} = t^{\mathcal{M}}$ we get $x^{\mathcal{M}'} = t^{\mathcal{M}'}$. Thus $\mathcal{M}' \vDash x = t$.

Remark 1. The last satisfaction clause can be stated differently if a name for each $a \in A$, let's say $\boldsymbol{a}$, is available in the signature: $\mathcal{M} \vDash \forall x\alpha \;\Leftrightarrow\; \mathcal{M} \vDash \alpha\frac{\boldsymbol{a}}{x}$ for all $a \in A$. This assumption permits the definition of the satisfaction relation for sentences using induction on sentences while bypassing arbitrary formulas. If not every $a \in A$ has a name in L, one could "fill up" L in advance by adjoining to L a name $\boldsymbol{a}$ for each a. But expanding the language is not always wanted and does not really simplify the matter.

A natural, often-used generalization of the last satisfaction clause is

$$\mathcal{M} \vDash \forall\vec{x}\varphi \;\Leftrightarrow\; \mathcal{M}_{\vec{x}}^{\vec{a}} \vDash \varphi \text{ for all } \vec{a} \in A^n.$$

For $\wedge, \neg$ basically the same satisfaction clauses have been used as in **1.3**. Since the definitions of $\vee$, $\rightarrow$, and $\leftrightarrow$ have not been altered, the following equivalences are valid in the current approach:

$$\mathcal{M} \vDash \alpha \vee \beta \;\Leftrightarrow\; \mathcal{M} \vDash \alpha \text{ or } \mathcal{M} \vDash \beta, \quad \mathcal{M} \vDash \alpha \rightarrow \beta \;\Leftrightarrow\; \text{if } \mathcal{M} \vDash \alpha \text{ then } \mathcal{M} \vDash \beta,$$

and analogously for $\leftrightarrow$. Further, $\exists x\varphi$ was correctly defined in **2.2**, because

$$\mathcal{M} \vDash \exists x\varphi \;\Leftrightarrow\; \text{there exists some } a \in A \text{ such that } \mathcal{M}_x^a \vDash \varphi.$$

Indeed, if $\mathcal{M} \vDash \neg\forall x\neg\varphi$ then, by definition, $\mathcal{M}_x^a \vDash \neg\varphi$ does not hold for all a, hence there is some $a \in A$ such that $\mathcal{M}_x^a \nvDash \neg\varphi$, or equivalently, such that $\mathcal{M}_x^a \vDash \varphi$. And this chain of reasoning is obviously reversible.

We now introduce several fundamental notions that will be treated systematically in **2.4** and **2.5**, once certain necessary preparations have been completed.

Definition. A formula or set of formulas in $\mathcal{L}$ is termed *satisfiable* if it has a model. φ is called *generally valid, logically valid*, or a *tautology*, in short, $\vDash \varphi$, if $\mathcal{M} \vDash \varphi$ for every model $\mathcal{M}$. The formulas α, β are called (logically or semantically)*equivalent*, symbolically $\alpha \equiv \beta$, if $\mathcal{M} \vDash \alpha \;\Leftrightarrow\; \mathcal{M} \vDash \beta$, for all $\mathcal{L}$-models $\mathcal{M}$. Further, let $\mathcal{A} \vDash \varphi$ (read *in* $\mathcal{A}$ *holds* φ or $\mathcal{A}$ *satisfies* φ) if $(\mathcal{A}, w) \vDash \varphi$ for all $w\colon Var \rightarrow A$. One writes $\mathcal{A} \vDash X$ in case $\mathcal{A} \vDash \varphi$ for all $\varphi \in X$. Finally, let $X \vDash \varphi$ (*from* X *follows* φ or φ *is a consequence of* X) if every model of X also satisfies the formula φ.

As in Chapter **1**, $\models$ denotes both the satisfaction and the consequence relation. Here, as there, we also write $\varphi_1, \ldots, \varphi_n \models \varphi$ for $\{\varphi_1, \ldots, \varphi_n\} \models \varphi$ etc. In addition, $\models$ denotes the validity relation in structures which is illustrated by the following

Example 2. We show that $\mathcal{A} \models \forall x \exists y\, x \neq y$ where the domain of $\mathcal{A}$ contains at least two elements. Indeed, let $\mathcal{M} = (\mathcal{A}, w)$ and $a \in A$ be arbitrarily given. Then there is some $b \in A$ with $a \neq b$. Hence, $(\mathcal{M}_x^a)_y^b = \mathcal{M}_{xy}^{a\,b} \models x \neq y$ and so $\mathcal{M}_x^a \models \exists y\, x \neq y$. Since a was arbitrary, $\mathcal{M} \models \forall x \exists y\, x \neq y$. Clearly the actual values of w are irrelevant in this argument. Hence $(\mathcal{A}, w) \models \forall x \exists y\, x \neq y$ for all w, that is, $\mathcal{A} \models \forall x \exists y\, x \neq y$.

Here some care is needed. While $\mathcal{M} \models \varphi$ or $\mathcal{M} \models \neg\varphi$ for all formulas, $\mathcal{A} \models \varphi$ or $\mathcal{A} \models \neg\varphi$ (the law of the excluded middle for validity in structures) is in general correct only for sentences φ, as Theorem 3.1 will show. If $\mathcal{A}$ contains more than one element, then, for example, neither $\mathcal{A} \models x = y$ nor $\mathcal{A} \models x \neq y$. Indeed, $x = y$ is falsified by any w such that $x^w \neq y^w$, and $x \neq y$ by any w with $x^w = y^w$. This is one of the reasons why models were not simply identified with structures.

For $\varphi \in \mathcal{L}$ let $\varphi^{\mathtt{G}}$ be the sentence $\forall x_1 \cdots \forall x_m \varphi$, where $x_1, \ldots, x_m$ is an enumeration of *free*$\,\varphi$ according to index size, say. $\varphi^{\mathtt{G}}$ is called the *generalized of* φ, also called its *universal closure*. For $\varphi \in \mathcal{L}^0$ clearly $\varphi^{\mathtt{G}} = \varphi$. From this definition results

(1) $\quad \mathcal{A} \models \varphi \;\Leftrightarrow\; \mathcal{A} \models \varphi^{\mathtt{G}}$,

and more generally $\mathcal{A} \models X \Leftrightarrow \mathcal{A} \models X^{\mathtt{G}}$ $(:= \{\varphi^{\mathtt{G}} \mid \varphi \in X\})$. (1) explains why φ and $\varphi^{\mathtt{G}}$ are often notionally identified and the information that formally runs $\varphi^{\mathtt{G}}$ is often shortened to φ. It must always be clear from the context whether our eye is on validity in a structure or in a model with its fixed valuation. Only in the first case can a generalization (or globalization) of the free variables be thought of as carried out. However, independent of this discussion, $\models \varphi \;\Leftrightarrow\; \models \varphi^{\mathtt{G}}$ always holds.

Even after just these incomplete considerations it is already clear that numerous properties of structures and whole systems of axioms can adequately be described by first-order formulas and sentences. Thus, for example, the axiom system mentioned in **2.1** for groups in $\{\circ, e, {}^{-1}\}$ can be formulated as follows:

$$\forall x \forall y \forall z\; x \circ (y \circ z) = (x \circ y) \circ z; \quad \forall x\, x \circ e = x; \quad \forall x\, x \circ x^{-1} = e.$$

Precisely the sentences following from these three axioms are the theorems of the *elementary group theory in* $\circ, e, {}^{-1}$, denoted by $T_G^{=}$. In the sense elaborated in **2.6**, an equivalent formulation of the theory of groups in $\circ, e$, denoted by T_G, is obtained if the last $T_G^{=}$-axiom is replaced by $\forall x \exists y\, x \circ y = e$.

An axiom system for ordered sets can also easily be provided, in that one formalizes the properties of irreflexivity, transitivity, and connexity. Here and elsewhere $\forall x_1 \cdots x_n \varphi$ stands for $\forall x_1 \cdots \forall x_n \varphi$:

$$\forall x\, x \not< x; \quad \forall xyz(x < y \wedge y < z \to x < z); \quad \forall xy(x \neq y \to x < y \vee y < x).$$

In writing down these and other axioms (e.g. those for groups as done above) the outer $\forall$-prefixes are occasionally omitted so as to save on writing, and we think implicitly of the generalization of variables as having been carried out. This is also the case for the formulation of (1) above, which strictly speaking runs

$$\text{for all } \mathcal{A}, \varphi : \mathcal{A} \vDash \varphi \;\Leftrightarrow\; \mathcal{A} \vDash \varphi^{\mathfrak{G}}.$$

For sentences α of a given language it is intuitively clear that the values of the variables of w for the relation $(\mathcal{A}, w) \vDash \alpha$ are irrelevant. The precise proof is extracted from the following theorem for $V = \emptyset$. Thus, either $(\mathcal{A}, w) \vDash \alpha$ for all w and hence $\mathcal{A} \vDash \alpha$, or else $(\mathcal{A}, w) \vDash \alpha$ for no w, i.e., $(\mathcal{A}, w) \vDash \neg\alpha$ for all w, and hence $\mathcal{A} \vDash \neg\alpha$. Sentences therefore obey the already-cited tertium non datur.

Theorem 3.1 (Coincidence theorem). *Let $V \subseteq Var$, $free\,\varphi \subseteq V$ and $\mathcal{M}, \mathcal{M}'$ be models on the same domain A such that $x^{\mathcal{M}} = x^{\mathcal{M}'}$ for all $x \in V$, and $\zeta^{\mathcal{M}} = \zeta^{\mathcal{M}'}$ for all extralogical symbols ζ occurring in φ. Then $\mathcal{M} \vDash \varphi \;\Leftrightarrow\; \mathcal{M}' \vDash \varphi$.*

Proof by induction on φ. Let φ be the prime formula $r\vec{t}$. As was mentioned earlier, the value of a term t depends only on the meaning of the symbols occurring in t. But in view of the suppositions regarding $t_1, \ldots, t_n$, these symbols are just the same in $\mathcal{M}$ and $\mathcal{M}'$. Thus, $\vec{t}^{\,\mathcal{M}} = \vec{t}^{\,\mathcal{M}'}$ (i.e., $t_i^{\mathcal{M}} = t_i^{\mathcal{M}'}$ for $i = 1, \ldots, n$), and therefore $\mathcal{M} \vDash r\vec{t} \;\Leftrightarrow\; r^{\mathcal{M}}\vec{t}^{\,\mathcal{M}} \;\Leftrightarrow\; r^{\mathcal{M}'}\vec{t}^{\,\mathcal{M}'} \;\Leftrightarrow\; \mathcal{M}' \vDash r\vec{t}$. For equations $t_1 = t_2$ one reasons analogously. Further, the induction hypothesis for α, β yields

$$\mathcal{M} \vDash \alpha \wedge \beta \Leftrightarrow \mathcal{M} \vDash \alpha, \beta \Leftrightarrow \mathcal{M}' \vDash \alpha, \beta \Leftrightarrow \mathcal{M}' \vDash \alpha \wedge \beta.$$

In the same way one obtains $\mathcal{M} \vDash \neg\alpha \Leftrightarrow \mathcal{M}' \vDash \neg\alpha$. By the induction step on $\forall$ it becomes clear that the induction hypothesis needs to be skillfully formulated. It must be given with respect to any pair of models and any V. Therefore let $a \in A$ and $\mathcal{M}_x^a \vDash \varphi$. Since for $V' := V \cup \{x\}$ certainly $free\,\varphi \subseteq V'$ and the models $\mathcal{M}_x^a$, ${\mathcal{M}'}_x^a$ coincide for all $y \in V'$ (although in general $x^{\mathcal{M}} \neq x^{\mathcal{M}'}$), by the induction hypothesis we have $\mathcal{M}_x^a \vDash \varphi \Leftrightarrow {\mathcal{M}'}_x^a \vDash \varphi$. This clearly implies

$$\mathcal{M} \vDash \forall x\varphi \Leftrightarrow \mathcal{M}_x^a \vDash \varphi \text{ for all } a \Leftrightarrow {\mathcal{M}'}_x^a \vDash \varphi \text{ for all } a \Leftrightarrow \mathcal{M}' \vDash \forall x\varphi. \qquad \square$$

It follows from this theorem that an $\mathcal{L}$-model $\mathcal{M} = (\mathcal{A}, w)$ of φ for the case that $\varphi \in \mathcal{L} \subseteq \mathcal{L}'$ can be completely arbitrarily expanded to an $\mathcal{L}'$-model $\mathcal{M}' = (\mathcal{A}', w)$ of φ, i.e., arbitrarily fixing $\zeta^{\mathcal{A}'}$ for $\zeta \in L' \setminus L$ gives $\mathcal{M} \vDash \varphi \Leftrightarrow \mathcal{M}' \vDash \varphi$ by the above theorem with $V = Var$. This readily implies that the consequence relation $\vDash_{\mathcal{L}'}$ with respect to $\mathcal{L}'$ is a *conservative* extension of $\vDash_{\mathcal{L}}$ in that $X \vDash_{\mathcal{L}} \varphi \Leftrightarrow X \vDash_{\mathcal{L}'} \varphi$, for all sets $X \subseteq \mathcal{L}$ and all $\varphi \in \mathcal{L}$. Hence, there is no need here for using indices. In particular, the satisfiability or general validity of φ depends only on the symbols effectively occurring in φ.

Another application of Theorem 3.1 is the following fact, which justifies the already mentioned "omission of superfluous quantifiers."

(2) $\forall x\varphi \equiv \varphi \equiv \exists x\varphi$, *supposing that* $x \notin free\,\varphi$.

Indeed, $x \notin \mathit{free}\,\varphi$ implies $\mathcal{M} \vDash \varphi \;\Leftrightarrow\; \mathcal{M}^a_x \vDash \varphi$ (here $a \in A$ is arbitrary) according to Theorem 3.1; choose $\mathcal{M}' = \mathcal{M}^a_x$ and $V = \mathit{free}\,\varphi$. Therefore,

$\mathcal{M} \vDash \forall x\varphi \Leftrightarrow \mathcal{M}^a_x \vDash \varphi$ for all $a \Leftrightarrow \mathcal{M} \vDash \varphi \Leftrightarrow \mathcal{M}^a_x \vDash \varphi$ for some $a \Leftrightarrow \mathcal{M} \vDash \exists x\varphi$.

Very important for the next theorem and elsewhere is

(3) *If* $\mathcal{A} \subseteq \mathcal{B}$, $\mathcal{M} = (\mathcal{A}, w)$, $\mathcal{M}' = (\mathcal{B}, w)$ *and* $w\colon \mathrm{Var} \to A$ *then* $t^{\mathcal{M}} = t^{\mathcal{M}'}$.

This is clear for prime terms, and the induction hypothesis $t_i^{\mathcal{M}} = t_i^{\mathcal{M}'}$ for $i = 1, \dots, n$ implies $(f\vec{t})^{\mathcal{M}} = f^{\mathcal{M}}(t_1^{\mathcal{M}}, \dots, t_n^{\mathcal{M}}) = f^{\mathcal{M}'}(t_1^{\mathcal{M}'}, \dots, t_n^{\mathcal{M}'}) = (f\vec{t})^{\mathcal{M}'}$.

By Theorem 3.1 the satisfaction of φ in $(\mathcal{A}, w)$ depends only on the values of the $x \in \mathit{free}\,\varphi$ under w. Let $\varphi = \varphi(\vec{x})$[4] and $\vec{a} = (a_1, \dots, a_n) \in A^n$. Then the statement

$$(\mathcal{A}, w) \vDash \varphi \text{ for a valuation } w \text{ with } x_1^w = a_1, \dots, x_n^w = a_n$$

can more suggestively be expressed by writing

$$(\mathcal{A}, \vec{a}) \vDash \varphi \quad \text{or} \quad \mathcal{A} \vDash \varphi\,[a_1, \dots, a_n] \quad \text{or} \quad \mathcal{A} \vDash \varphi\,[\vec{a}]$$

without mentioning w as a global valuation. Such notation also makes sense if w is restricted to a valuation on $\{x_1, \dots, x_n\}$. One may accordingly extend the concept of a model and call a pair $(\mathcal{A}, \vec{a})$ a model for a formula $\varphi(\vec{x})$ whenever $(\mathcal{A}, \vec{a}) \vDash \varphi(\vec{x})$, in particular if $\varphi \in \mathcal{L}^n$. We return to this extended concept in **4.1**. Until then we use it only for $n = 0$. That is, besides $\mathcal{M} = (\mathcal{A}, w)$ also the structure $\mathcal{A}$ itself is occasionally called a model for a set $S \subseteq \mathcal{L}^0$ of sentences provided $\mathcal{A} \vDash S$.

Corresponding to the above let $t^{\mathcal{A},\vec{a}}$, or more suggestively $t^{\mathcal{A}}(\vec{a})$, denote the value of $t = t(\vec{x})$. Then (3) can somewhat more simply be written as

(4) $\mathcal{A} \subseteq \mathcal{B}$ *and* $t = t(\vec{x})$ *imply* $t^{\mathcal{A}}(\vec{a}) = t^{\mathcal{B}}(\vec{a})$ $\quad (\vec{a} \in A^n)$.

Thus, along with the basic functions also the so-called *term functions* $\vec{a} \mapsto t^{\mathcal{A}}(\vec{a})$ are the restrictions to their counterparts in $\mathcal{B}$. Clearly, if $n = 0$ or t is variable-free, one may write $t^{\mathcal{A}}$ for $t^{\mathcal{A}}(\vec{a})$. Note that in these cases $t^{\mathcal{A}} = t^{\mathcal{B}}$ provided $\mathcal{A} \subseteq \mathcal{B}$, by (4).

As above let $\varphi = \varphi(\vec{x})$. Then $\varphi^{\mathcal{A}} := \{\vec{a} \in A^n \mid \mathcal{A} \vDash \varphi\,[\vec{a}]\}$ is called *the predicate defined by the formula* φ *in the structure* $\mathcal{A}$. For instance, the $\leqslant$-predicate in $(\mathbb{N}, +)$ is defined by $\varphi(x, y) = \exists z\, z + x = y$, but also by several other formulas.

More generally, $P \subseteq A^n$ is termed (elementarily or first order) *definable in* $\mathcal{A}$ if there is some $\varphi = \varphi(\vec{x})$ with $P = \varphi^{\mathcal{A}}$. Analogously, $f\colon A^n \to A$ is called definable in $\mathcal{A}$ if $\varphi^{\mathcal{A}} = \operatorname{graph} f$ for some $\varphi(\vec{x}, y)$. We also talk in all these cases of *explicit* definability in $\mathcal{A}$, to distinguish this from recursive definability. Much information on a structure can be gained from the knowledge which predicates, or at least which sets, are definable. For instance, the sets definable in $(\mathbb{N}, 0, 1, +)$ are the eventually periodic ones (periodic from some number upwards). Thus, $\cdot$ cannot explicitly be defined by $+, 0, 1$ because the set of square numbers is not eventually periodic.

[4] Since this is to mean $\mathit{free}\,\varphi \subseteq \{x_1, \dots, x_n\}$, $\vec{x}$ is not uniquely determined by φ. Hence, the phrase "Let $\varphi = \varphi(\vec{x}) \dots$" implicitly includes along with a given φ also a tuple $\vec{x}$ given in advance. The notation $\varphi = \varphi(\vec{x})$ does not even state that φ contains free variables at all.

$\mathcal{A} \subseteq \mathcal{B}$ and $\varphi = \varphi(\vec{x})$ do not imply $\varphi^{\mathcal{A}} = \varphi^{\mathcal{B}} \cap A^n$, in general. For instance, let $\mathcal{A} = (\mathbb{N}, +)$, $\mathcal{B} = (\mathbb{Z}, +)$, and $\varphi = \exists z\, z + x = y$. Then $\varphi^{\mathcal{A}} = \leqslant^{\mathcal{A}}$, while $\varphi^{\mathcal{B}}$ contains all pairs $(a, b) \in \mathbb{Z}^2$. As the next theorem will show, $\varphi^{\mathcal{A}} = \varphi^{\mathcal{B}} \cap A^n$ holds in general only for open formulas φ, and is even characteristic for $\mathcal{A} \subseteq \mathcal{B}$ provided $A \subseteq B$. Clearly, $A \subseteq B$ is much weaker a condition than $\mathcal{A} \subseteq \mathcal{B}$:

Theorem 3.2 (Substructure theorem). *For structures $\mathcal{A}, \mathcal{B}$ such that $A \subseteq B$ the following conditions are equivalent:*

(i) $\mathcal{A} \subseteq \mathcal{B}$,

(ii) $\mathcal{A} \vDash \varphi\,[\vec{a}] \Leftrightarrow \mathcal{B} \vDash \varphi\,[\vec{a}]$, *for all open* $\varphi = \varphi(\vec{x})$ *and all* $\vec{a} \in A^n$,

(iii) $\mathcal{A} \vDash \varphi\,[\vec{a}] \Leftrightarrow \mathcal{B} \vDash \varphi\,[\vec{a}]$, *for all prime formulas* $\varphi(\vec{x})$ *and all* $\vec{a} \in A^n$.

Proof. (i)⇒(ii): It suffices to prove that $\mathcal{M} \vDash \varphi \Leftrightarrow \mathcal{M}' \vDash \varphi$, with $\mathcal{M} = (\mathcal{A}, w)$ and $\mathcal{M}' = (\mathcal{B}, w)$ where $w\colon Var \to A$. In view of (3) the claim is obvious for prime formulas, and the induction steps for $\wedge, \neg$ are carried out just as in Theorem 3.1. (ii)⇒(iii): Trivial. (iii)⇒(i): By (iii), $r^{\mathcal{A}}\vec{a} \Leftrightarrow \mathcal{A} \vDash r\vec{x}\,[\vec{a}] \Leftrightarrow \mathcal{B} \vDash r\vec{x}\,[\vec{a}] \Leftrightarrow r^{\mathcal{B}}\vec{a}$. Analogously $f^{\mathcal{A}}\vec{a} = b \Leftrightarrow \mathcal{A} \vDash f\vec{x} = y\,[\vec{a}, b] \Leftrightarrow \mathcal{B} \vDash f\vec{x} = y\,[\vec{a}, b] \Leftrightarrow f^{\mathcal{B}}\vec{a} = b$, for all $\vec{a} \in A^n$ and $b \in A$. These conclusions state precisely that $\mathcal{A} \subseteq \mathcal{B}$. ❏

Let α be of the form $\forall\vec{x}\beta$ with open β, where $\forall\vec{x}$ may also be the empty prefix. Then α is a *universal* or $\forall$*-formula* (spoken "A-formula"), and for $\alpha \in \mathcal{L}^0$ also a *universal* or $\forall$*-sentence*. A simple example is $\forall x \forall y\ x = y$, which holds in $\mathcal{A}$ iff A contains precisely one element. Dually, $\exists\vec{x}\beta$ (β open) is termed an $\exists$*-formula*, and an $\exists$*-sentence* whenever $\exists\vec{x}\beta \in \mathcal{L}^0$. Examples are the "how-many sentences"

$$\exists_1 := \exists \boldsymbol{v}_0\, \boldsymbol{v}_0 = \boldsymbol{v}_0; \quad \exists_n := \exists \boldsymbol{v}_0 \ldots \exists \boldsymbol{v}_{n-1} \bigwedge_{i<j<n} \boldsymbol{v}_i \neq \boldsymbol{v}_j \quad (n > 1).$$

$\exists_n$ states 'there exist at least n elements', $\neg\exists_{n+1}$ thus that 'there exist at most n elements', and $\exists_{=n} := \exists_n \wedge \neg\exists_{n+1}$ says 'there exist exactly n elements'. Since $\exists_1$ is a tautology, it is convenient to set $\top := \exists_1$, and $\exists_0 := \bot := \neg\top$.

Corollary 3.3. *Let $\mathcal{A} \subseteq \mathcal{B}$. Then every $\forall$-sentence $\forall\vec{x}\alpha$ valid in $\mathcal{B}$ is also satisfied in $\mathcal{A}$. Dually, every $\exists$-sentence $\exists\vec{x}\beta$ valid in $\mathcal{A}$ is also valid in $\mathcal{B}$.*

Proof. Let $\mathcal{B} \vDash \forall\vec{x}\beta$ and $\vec{a} \in A^n$. Then $\mathcal{B} \vDash \beta\,[\vec{a}]$; hence $\mathcal{A} \vDash \beta\,[\vec{a}]$ by Theorem 3.2. $\vec{a}$ was arbitrary, so $\mathcal{A} \vDash \forall\vec{x}\beta$. Now let $\mathcal{A} \vDash \exists\vec{x}\beta$. Then $\mathcal{A} \vDash \beta\,[\vec{a}]$ for some $\vec{a} \in A^n$, hence $\mathcal{B} \vDash \beta\,[\vec{a}]$ by Theorem 3.2, and consequently $\mathcal{B} \vDash \exists\vec{x}\beta$. ❏

We formulate a generalization of certain individual often-used arguments, namely

Theorem 3.4 (Invariance theorem). *Let $\mathcal{A}, \mathcal{B}$ be isomorphic L-structures and let $\imath\colon \mathcal{A} \to \mathcal{B}$ be an isomorphism. Then for all $\varphi = \varphi(\vec{x})$ and all $\vec{a} \in A^n$,*

$$\mathcal{A} \vDash \varphi\,[\vec{a}] \;\Leftrightarrow\; \mathcal{B} \vDash \varphi\,[\imath\vec{a}] \quad \big(\imath\vec{a} = (\imath a_1, \ldots, \imath a_n)\big).$$

In particular $\mathcal{A} \vDash \alpha \Leftrightarrow \mathcal{B} \vDash \alpha$, for all sentences α of $\mathcal{L}$.

Proof. It is convenient to reformulate the claim as

$$\mathcal{M} \vDash \varphi \;\Leftrightarrow\; \mathcal{M}' \vDash \varphi \qquad \big(\mathcal{M} = (\mathcal{A}, w),\ \mathcal{M}' = (\mathcal{B}, w'),\ w' : x \mapsto \imath x^w\big).$$

It is easy to confirm this inductively on φ after one has first proved that $\imath(t^{\mathcal{M}}) = t^{\mathcal{M}'}$ inductively on t. The particular case for sentences results from the case $n = 0$. ❑

Thus, for example, it is once and for all clear that the isomorphic image of a group is a group even if we know at first only that it is a groupoid. Simply let α in the theorem run through all axioms of group theory. Here is another application. Let $\imath$ be an isomorphism of the group $\mathcal{A} = (A, \circ)$ onto the group $\mathcal{A}' = (A', \circ)$ and let e and e' denote their unit elements, not named in the signature. We claim that nonetheless $\imath e = e'$, using the easily provable fact that the unit element of a group is the only solution of the equation $x \circ x \mathbin{=\!\!=} x$ (Example 2, page 65). Thus, since $\mathcal{A} \vDash e \circ e \mathbin{=\!\!=} e$, we get $\mathcal{A}' \vDash \imath e \circ \imath e \mathbin{=\!\!=} \imath e$ by Theorem 3.4, hence $\imath e = e'$. Theorem 3.4, incidentally, holds for formulas of higher order as well; see **3.7**. For instance, that a set is continuously ordered is likewise invariant under isomorphism.

$\mathcal{L}$-structures $\mathcal{A}, \mathcal{B}$ are termed *elementary equivalent* if $\mathcal{A} \vDash \alpha \Leftrightarrow \mathcal{B} \vDash \alpha$, for all $\alpha \in \mathcal{L}^0$. One then writes $\mathcal{A} \equiv \mathcal{B}$. We consider this important notion in **3.3** and more closely in **5.1**. Theorem 3.4 states in particular that $\mathcal{A} \simeq \mathcal{B} \;\Rightarrow\; \mathcal{A} \equiv \mathcal{B}$. The question immediately arises whether the converse of this also holds. For infinite structures the answer is negative (see **3.3**), for finite structures affirmative; a finite structure of a finite signature can, up to isomorphism, even be described by a single sentence. For example, the 2-element group $(\{0, 1\}, +)$ is up to isomorphism well determined by the following sentence, which tells us precisely how + operates:

$$\exists \boldsymbol{v}_0 \exists \boldsymbol{v}_1 [\boldsymbol{v}_0 \neq \boldsymbol{v}_1 \wedge \forall x (x \mathbin{=\!\!=} \boldsymbol{v}_0 \vee x \mathbin{=\!\!=} \boldsymbol{v}_1) \wedge \boldsymbol{v}_0 + \boldsymbol{v}_0 \mathbin{=\!\!=} \boldsymbol{v}_1 + \boldsymbol{v}_1 \mathbin{=\!\!=} \boldsymbol{v}_0 \wedge \boldsymbol{v}_0 + \boldsymbol{v}_1 \mathbin{=\!\!=} \boldsymbol{v}_1 + \boldsymbol{v}_0 \mathbin{=\!\!=} \boldsymbol{v}_1].$$

We now investigate the behavior of the satisfaction relation under substitution. The definition of $\varphi \frac{t}{x}$ in **2.2** pays no attention to *collision of variables*, which is taken to mean that certain variables of the substitution term t after application of the substitution fall into the scope of quantifiers. In this case $\mathcal{M} \vDash \forall x \varphi$ does not necessarily imply $\mathcal{M} \vDash \varphi \frac{t}{x}$, although this might have been expected. In other words, $\forall x \varphi \vDash \varphi \frac{t}{x}$ is not unrestrictedly correct. For instance, if $\varphi = \exists y\, x \neq y$ then certainly $\mathcal{M} \vDash \forall x \varphi$ $(= \forall x \exists y\, x \neq y)$, provided $\mathcal{M}$ has at least two elements, but $\mathcal{M} \vDash \varphi \frac{y}{x}$ $(= \exists y\, y \neq y)$ is certainly false. Analogously $\varphi \frac{t}{x} \vDash \exists x \varphi$ is not correct, in general. Choose, for example, $\forall y\, x \mathbin{=\!\!=} y$ for φ and y for t.

One could forcibly obtain $\forall x \varphi \vDash \varphi \frac{t}{x}$ without any limitation by renaming bound variables by a suitable modification of the inductive definition of $\varphi \frac{t}{x}$ in the quantifier step. However, such measures are rather unwieldy for the arithmetization of proof method in **6.2**. It is therefore preferable to put up with minor restrictions when we are formulating rules of deduction later. The restrictions we will use are somewhat stronger than they need to be but can easier be handled; they look as follows:

$\varphi, \frac{t}{x}$ are called *collision-free* if $y \notin bnd\,\varphi$ for all $y \in var\,t \setminus \{x\}$. We need not to require $x \notin bnd\,\varphi$ because t is substituted only at free occurrences of x, that is, even if $x \in var\,t$, x cannot fall after substitution within the scope of a prefix $\forall x$. For collision-free $\varphi, \frac{t}{x}$ we will then get $\forall x\varphi \vDash \varphi\,\frac{t}{x}$ by Corollary 3.6 below.

If σ is a global substitution (see **2.2**) then φ, σ are termed *collision-free* if $\varphi, \frac{x^\sigma}{x}$ are collision-free for every $x \in Var$. In the special case $\sigma = \frac{\vec{t}}{\vec{x}}$, this condition clearly need be checked only for the pairs $\varphi, \frac{x_i^\sigma}{x_i}$ $(i = 1, \ldots, n)$.

For $\mathcal{M} = (\mathcal{A}, w)$ put $\mathcal{M}^\sigma := (\mathcal{A}, w^\sigma)$ where $x^{w^\sigma} := (x^\sigma)^{\mathcal{M}}$ for all $x \in Var$. This equation reproduces itself inductively to $t^{\mathcal{M}^\sigma} = t^{\sigma\mathcal{M}}$ for all t. Indeed, it is correct for prime terms. Now let $t_i^{\mathcal{M}^\sigma} = t_i^{\sigma\mathcal{M}}$ for $i = 1, \ldots, n$ by the induction hypothesis. Then the claim for $t = ft_1 \cdots t_n$ follows from

$$t^{\mathcal{M}^\sigma} = f^{\mathcal{M}}(t_1^{\mathcal{M}^\sigma}, \ldots, t_n^{\mathcal{M}^\sigma}) = f^{\mathcal{M}}(t_1^{\sigma\mathcal{M}}, \ldots, t_n^{\sigma\mathcal{M}}) = t^{\sigma\mathcal{M}}.$$

Note that $\mathcal{M}^\sigma$ coincides with $\mathcal{M}_{\vec{x}}^{\vec{t}^{\mathcal{M}}}$ for the case $\sigma = \frac{\vec{t}}{\vec{x}}$.

Theorem 3.5 (Substitution theorem). *Suppose $\mathcal{M}$ is a model and σ a global substitution. Then for all formulas φ such that φ, σ are collision-free,*

$$\mathcal{M} \vDash \varphi^\sigma \;\Leftrightarrow\; \mathcal{M}^\sigma \vDash \varphi.$$

In particular, $\mathcal{M} \vDash \varphi\,\frac{\vec{t}}{\vec{x}} \;\Leftrightarrow\; \mathcal{M}_{\vec{x}}^{\vec{t}^{\mathcal{M}}} \vDash \varphi$, provided $\varphi, \frac{\vec{t}}{\vec{x}}$ are collision-free.

Proof by induction on φ. In view of $t^{\sigma\mathcal{M}} = t^{\mathcal{M}^\sigma}$, we obtain

$$\mathcal{M} \vDash (t_1 \mathbf{=} t_2)^\sigma \;\Leftrightarrow\; t_1^{\sigma\mathcal{M}} = t_2^{\sigma\mathcal{M}} \;\Leftrightarrow\; t_1^{\mathcal{M}^\sigma} = t_2^{\mathcal{M}^\sigma} \;\Leftrightarrow\; \mathcal{M}^\sigma \vDash t_1 \mathbf{=} t_2.$$

Prime formulas of the form $r\vec{t}$ are treated analogously. The induction steps for $\wedge, \neg$ are harmless. Only the $\forall$-step $\varphi = \forall x\alpha$ is interesting, and is achieved as follows:

$$\begin{aligned}
\mathcal{M} \vDash (\forall x\alpha)^\sigma &\Leftrightarrow \mathcal{M} \vDash \forall x\,\alpha^\tau && \text{(where } x^\tau = x \text{ and } y^\tau = y^\sigma \text{ otherwise)}\\
&\Leftrightarrow \mathcal{M}_x^a \vDash \alpha^\tau \text{ for all } a && \text{(definition)}\\
&\Leftrightarrow (\mathcal{M}_x^a)^\tau \vDash \alpha \text{ for all } a && \text{(induction hypothesis; } \alpha, \tau \text{ collision-free)}\\
&\Leftrightarrow (\mathcal{M}^\sigma)_x^a \vDash \alpha \text{ for all } a && \big(\text{since } (\mathcal{M}_x^a)^\tau = (\mathcal{M}^\sigma)_x^a, \text{ see below}\big)\\
&\Leftrightarrow \mathcal{M}^\sigma \vDash \forall x\alpha.
\end{aligned}$$

We show $(\mathcal{M}_x^a)^\tau = (\mathcal{M}^\sigma)_x^a$. Since $\forall x\alpha, \sigma$ (hence $\forall x\alpha, \frac{y^\sigma}{y}$ for every y) are collision-free, we have $x \notin var\,y^\sigma$ provided $y \neq x$, and since $y^\tau = y^\sigma$ we get in this case

$$y^{(\mathcal{M}_x^a)^\tau} = y^{\tau\mathcal{M}_x^a} = y^{\sigma\mathcal{M}_x^a} = y^{\sigma\mathcal{M}} = y^{\mathcal{M}^\sigma} = y^{(\mathcal{M}^\sigma)_x^a}.$$

But also in the case $y = x$ we have $x^{(\mathcal{M}_x^a)^\tau} = x^{\tau\mathcal{M}_x^a} = x^{\mathcal{M}_x^a} = a = x^{(\mathcal{M}^\sigma)_x^a}$. ❏

Corollary 3.6. *For all φ and $\frac{\vec{t}}{\vec{x}}$ such that $\varphi, \frac{\vec{t}}{\vec{x}}$ are collision-free, the following hold:*

(a) $\forall\vec{x}\varphi \vDash \varphi\,\frac{\vec{t}}{\vec{x}}$, *in particular* $\forall x\varphi \vDash \varphi\,\frac{t}{x}$, (b) $\varphi\,\frac{\vec{t}}{\vec{x}} \vDash \exists\vec{x}\varphi$,

(c) $\varphi\,\frac{s}{x}, s \mathbf{=} t \vDash \varphi\,\frac{t}{x}$ *if $\varphi, \frac{s}{x}, \frac{t}{x}$ are collision-free.*

Proof. Let $\mathcal{M} \vDash \forall \vec{x}\varphi$, so that $\mathcal{M}^{\vec{a}}_{\vec{x}} \vDash \varphi$ for all $\vec{a} \in A^n$. In particular, $\mathcal{M}^{\vec{t}^{\mathcal{M}}}_{\vec{x}} \vDash \varphi$, so that $\mathcal{M} \vDash \varphi\, {}^{\vec{t}}_{\vec{x}}$ by the theorem. (b) is equivalent to $\neg\exists\vec{x}\varphi \vDash \neg\varphi\, {}^{\vec{t}}_{\vec{x}}$. This holds by (a), for $\neg\exists\vec{x}\varphi \equiv \forall\vec{x}\neg\varphi$ and $\neg(\varphi\, {}^{\vec{t}}_{\vec{x}}) \equiv (\neg\varphi)\, {}^{\vec{t}}_{\vec{x}}$. (c): Let $\mathcal{M} \vDash \varphi\, {}^{s}_{x}, s = t$, so that $s^{\mathcal{M}} = t^{\mathcal{M}}$ and $\mathcal{M}^{s^{\mathcal{M}}}_{x} \vDash \varphi$ by the theorem, hence also $\mathcal{M}^{t^{\mathcal{M}}}_{x} \vDash \varphi$. Thus $\mathcal{M} \vDash \varphi\, {}^{t}_{x}$. ❑

Remark 2. Since the identical substitution ι is obviously collision-free with every formula; $\forall x\varphi \vDash \varphi\ (= \varphi^{\iota})$ is always the case. Moreover, $\forall x\varphi \vDash \varphi\, {}^{t}_{x}$ is correct without any restriction provided t contains at most the variable x, since $\varphi, {}^{t}_{x}$ are then collision-free. Theorem 3.5 and Corollary 3.6 are easily strengthened. Define inductively a ternary predicate 't is free for x in φ', which intuitively is to mean that no free occurrence in φ of the variable x lies within the scope of a prefix $\forall y$ whenever $y \in var\, t$. Theorem 3.5 holds then for $\sigma = {}^{t}_{x}$ as well, so that nothing needs to be changed in the proofs based on this theorem if one works with 't is free for x in φ', or simply reads "$\varphi, {}^{t}_{x}$ are collision-free" as "x is free for t in φ." Though collision-freeness is somewhat cruder, it is for all that more wieldy, which will pay off, for example, in **6.2** where proofs will be gödelized. Once one has become accustomed to the required caution, it is allowable not always to state explicitly the restrictions caused by collisions of variables, but rather to assume them tacitly.

Theorem 3.5 also shows that the quantifier "there exists exactly one," denoted by $\exists!$, is correctly defined by $\exists!x\varphi := \exists x\varphi \wedge \forall x\forall y(\varphi \wedge \varphi\, {}^{y}_{x} \to x = y)$ with $y \notin var\,\varphi$. Indeed, $\mathcal{M} \vDash \forall x\forall y(\varphi \wedge \varphi\, {}^{y}_{x} \to x = y)$ means just $\mathcal{M}^{a}_{x} \vDash \varphi\ \&\ \mathcal{M}^{b}_{y} \vDash \varphi\ \Rightarrow\ a = b$, or equivalently, $\mathcal{M}^{a}_{x} \vDash \varphi$ for at most one a. Anyone who would like to verify this to the utmost precision should observe that $\mathcal{M}^{b}_{y} \vDash \varphi\ \Leftrightarrow\ \mathcal{M} \vDash \varphi$ whenever $y \notin var\,\varphi$. Putting together, $\mathcal{M} \vDash \exists!x\varphi$ iff there is precisely one $a \in A$ such that $\mathcal{M}^{a}_{x} \vDash \varphi$. A particularly simple example is $\mathcal{M} \vDash \exists!x\, x = y$, for arbitrary $\mathcal{M}$. In other words, $\exists!x\, x = y$ is a tautology. These will be discussed in more detail in **2.4**.

There are various equivalent definitions of $\exists!x\varphi$. For example, a short and catchy formula is $\exists x\forall y(\varphi\, {}^{y}_{x} \leftrightarrow x = y)$, where $y \notin var\,\varphi$.

Exercises

1. Prove $\exists x\exists y(\varphi \wedge \varphi\, {}^{y}_{x} \wedge x \neq y) \vDash \forall x\exists y(\varphi\, {}^{y}_{x} \wedge x \neq y)$ provided $y \notin var\,\varphi$.

2. Verify $\exists x\forall y(\varphi\, {}^{y}_{x} \leftrightarrow x = y) \vDash \exists!x\varphi \qquad (y \notin var\,\varphi)$.

3. Suppose $\mathcal{A}'$ results from $\mathcal{A}$ by adjoining a constant symbol $\boldsymbol{a}$ for some $a \in A$. Prove $t(x)^{\mathcal{A},a} = t(\boldsymbol{a})^{\mathcal{A}'}$ and $\mathcal{A} \vDash \alpha\,[a]\ \Leftrightarrow\ \mathcal{A}' \vDash \alpha(\boldsymbol{a})\ (= \alpha\, {}^{\boldsymbol{a}}_{x})$ for $\alpha = \alpha(x)$. This is easily generalized to the case of more than one variable.

4. Show that (a) a conjunction of the $\exists_i$ and their negations is equivalent to $\exists_n$, $\neg\exists_n$, or $\exists_n \wedge \neg\exists_m$ for suitable n, m, (b) a Boolean combination of the $\exists_i$ is equivalent to $\bigvee_{\nu \leqslant n} \exists_{=k_\nu} \vee \exists_m$, where $0 \leqslant k_0 < \cdots < k_n$, $n < m$, and the disjunction term $\exists_m$ may actually be absent.

2.4 General Validity and Logical Equivalence

From the perspective of predicate logic $\alpha \vee \neg\alpha$ ($\alpha \in \mathcal{L}$) is a trivial example of a tautology, because it results by inserting α for p from the propositional tautology $p \vee \neg p$. Every propositional tautology provides generally valid $\mathcal{L}$-formulas by the insertion of $\mathcal{L}$-formulas for the propositional variables. But there are also tautologies not arising in this way, for example $\forall x(x < x \vee x \not< x)$. This tautology is the result of generalizing $x < x \vee x \not< x$. However, the tautologies $\exists x\, x = x$ and $\exists x\, x = t$ for $x \notin var\, t$ are not generated in this way. The former arises from the convention that structures are always nonempty, the latter from that all basic operations are totally defined. A particularly interesting tautology is presented by the following

Example 1 (Russell's antinomy). We will show that $\vDash \neg\exists u\forall x(x \in u \leftrightarrow x \notin x)$, the nonexistence of the "Russellean set" u, consisting of all sets not containing themselves as a member (see also **3.4**). Remarkably, the proof does not assume that $\in$ means membership. By Corollary 3.6(a), $\forall x(x \in u \leftrightarrow x \notin x) \vDash u \in u \leftrightarrow u \notin u$. Since $u \in u \leftrightarrow u \notin u$ is obviously unsatisfiable, the same holds for $\forall x(x \in u \leftrightarrow x \notin x)$, hence also for $\exists u\forall x(x \in u \leftrightarrow x \notin x)$. Thus, $\neg\exists u\forall x(x \in u \leftrightarrow x \notin x)$ is a tautology. This inference need not at all be related to set theory! The antinomy arises here from that the (unsatisfiable) $\exists u\forall x(x \in u \leftrightarrow x \notin x)$ *should* hold in set theory if Cantor's definition of a set as an arbitrary collection of objects is taken literally.

The satisfaction clause for $\alpha \to \beta$ easily yields $\alpha \vDash \beta \;\Leftrightarrow\; \vDash \alpha \to \beta$, a special case of $X, \alpha \vDash \beta \;\Leftrightarrow\; X \vDash \alpha \to \beta$. This can be useful in checking whether formulas given in implicative form are tautologies, as was mentioned already in **1.3**. Thus, from $\forall x\alpha \vDash \alpha\,\frac{t}{x}$ one immediately obtains $\vDash \forall x\alpha \to \alpha\,\frac{t}{x}$ for collision-free $\alpha, \frac{t}{x}$.

As in propositional logic, $\alpha \equiv \beta$ is again equivalent to $\vDash \alpha \leftrightarrow \beta$. By inserting $\mathcal{L}$-formulas for the variables of a propositional equivalence one automatically procures one of predicate logic. Thus, for instance, $\alpha \to \beta \equiv \neg\alpha \vee \beta$, because certainly $p \to q \equiv \neg p \vee q$. Since every $\mathcal{L}$-formula results from the insertion of propositionally irreducible $\mathcal{L}$-formulas in a formula of propositional logic, one also sees that every $\mathcal{L}$-formula can equivalently be converted into a conjunctive normal form. But there are also numerous other equivalences, for example

$$\neg\forall x\alpha \equiv \exists x\neg\alpha \text{ and } \neg\exists x\alpha \equiv \forall x\neg\alpha.$$

The first of these means just $\neg\forall x\alpha \equiv \neg\forall x\neg\neg\alpha$ ($= \exists x\neg\alpha$), obtained by replacing α on the left by the equivalent formula $\neg\neg\alpha$. This is a simple application of Theorem 4.1 below with $\equiv$ for $\approx$. As in propositional logic, semantical equivalence is an equivalence relation in $\mathcal{L}$ and, moreover, a *congruence in* $\mathcal{L}$. Speaking more generally, an equivalence relation $\approx$ in $\mathcal{L}$ that satisfies the congruence property

$$\text{CP:}\quad \alpha \approx \alpha',\ \beta \approx \beta' \;\Rightarrow\; \alpha \wedge \beta \approx \alpha' \wedge \beta',\ \neg\alpha \approx \neg\alpha',\ \forall x\alpha \approx \forall x\alpha'$$

is termed a *congruence in* $\mathcal{L}$. Its most important property is expressed by

Theorem 4.1 (Replacement theorem). *Let $\approx$ be a congruence in $\mathcal{L}$ and $\alpha \approx \alpha'$. If φ' results from φ by replacing the formula α at one or more of its occurrences in φ by the formula α', then $\varphi \approx \varphi'$.*

Proof by induction on φ. Suppose φ is a prime formula. Both for $\varphi = \alpha$ and $\varphi \neq \alpha$, $\varphi \approx \varphi'$ clearly holds. Now let $\varphi = \varphi_1 \wedge \varphi_2$. In case $\varphi = \alpha$ holds trivially $\varphi \approx \varphi'$. Otherwise $\varphi' = \varphi'_1 \wedge \varphi'_2$, where φ'_1, φ'_2 result from φ_1, φ_1 by possible replacements. By the induction hypothesis $\varphi_1 \approx \varphi'_1$ and $\varphi_2 \approx \varphi'_2$. Hence, $\varphi = \varphi_1 \wedge \varphi_2 \approx \varphi'_1 \wedge \varphi'_2 = \varphi'$ according to CP. The induction steps for $\neg$, $\forall$ follow analogously. ❑

This theorem will constantly be used, mainly with $\equiv$ for $\approx$, without actually specifically being cited, just as in the arithmetical rearrangement of terms, where the laws of arithmetic used are hardly ever named explicitly. The theorem readily implies that CP is provable for all defined connectives like $\rightarrow$ and $\exists$. For example, $\alpha \approx \alpha' \Rightarrow \exists x\alpha \approx \exists x\alpha'$, because $\alpha \approx \alpha' \Rightarrow \exists x\alpha = \neg\forall x\neg\alpha \approx \neg\forall x\neg\alpha' = \exists x\alpha'$.

Predicate logical languages have a finer structure than those of propositional logic. There are consequently further interesting congruences in $\mathcal{L}$. Thus, formulas α, β are *equivalent in an L-structure* $\mathcal{A}$, symbolized $\alpha \equiv_{\mathcal{A}} \beta$, if $\mathcal{A} \vDash \alpha\,[w] \Leftrightarrow \mathcal{A} \vDash \beta\,[w]$, for all w. For instance, in $\mathcal{A} = (\mathbb{N}, <, +, 0)$ the formulas $x < y$ and $\exists z\,(z \neq 0 \wedge x + z = y)$ are equivalent. The proof of the congruence property CP for $\equiv_{\mathcal{A}}$ is very simple, hence is left to the reader.

Clearly, $\alpha \equiv_{\mathcal{A}} \beta$ is equivalent to $\mathcal{A} \vDash \alpha \leftrightarrow \beta$. Because of $\equiv \subseteq \equiv_{\mathcal{A}}$, properties such as $\neg\forall x\alpha \equiv \exists x\neg\alpha$ carry over from $\equiv$ to $\equiv_{\mathcal{A}}$. But there are often new interesting equivalences in certain structures. For instance, there are structures in which every formula is equivalent to an open one, as we will see in **5.6**.

A very important fact with an almost trivial proof is that the intersection of a family of congruences is itself a congruence. Consequently, for any class $\boldsymbol{K} \neq \emptyset$ of $\mathcal{L}$-structures, $\equiv_{\boldsymbol{K}} := \bigcap\{\equiv_{\mathcal{A}} |\ \mathcal{A} \in \boldsymbol{K}\}$ is always a congruence. For the class $\boldsymbol{K}$ of *all* $\mathcal{L}$-structures, $\equiv_{\boldsymbol{K}}$ is identical to the logical equivalence $\equiv$, which in this section we deal with exclusively. In the following we list its most important features; they should be committed to memory, since they will continually be applied.

(1) $\forall x(\alpha \wedge \beta) \equiv \forall x\alpha \wedge \forall x\beta$, (2) $\exists x(\alpha \vee \beta) \equiv \exists x\alpha \vee \exists x\beta$,

(3) $\forall x\forall y\alpha \equiv \forall y\forall x\alpha$, (4) $\exists x\exists y\alpha \equiv \exists y\exists x\alpha$.

If x does not occur free in the formula β, then also

(5) $\forall x(\alpha \vee \beta) \equiv \forall x\alpha \vee \beta$, (6) $\exists x(\alpha \wedge \beta) \equiv \exists x\alpha \wedge \beta$,

(7) $\forall x\beta \equiv \beta$, (8) $\exists x\beta \equiv \beta$,

(9) $\forall x(\alpha \rightarrow \beta) \equiv \exists x\alpha \rightarrow \beta$, (10) $\exists x(\alpha \rightarrow \beta) \equiv \forall x\alpha \rightarrow \beta$.

The simple proofs are left to the reader. (7) and (8) were stated in (2) in **2.3**. Only (9) and (10) look at first sight surprising. But in practice these equivalences

are very frequently used. Consider for a fixed set of formulas X the evidently true metalogical assertion 'for all α: if $X \vDash \alpha, \neg\alpha$ then $X \vDash \forall x\, x \neq x$'. The latter clearly states the same as 'If there is an α such that $X \vDash \alpha, \neg\alpha$ then $X \vDash \forall x\, x \neq x$'.

Remark. In everyday speech variables tend to remain unquantified, partly because in some cases the same meaning results from quantifying with "there exists a" or "for all." For instance, consider the following three sentences, which obviously tell us the same thing, and of which the last two correspond to the logical equivalence (9):

- If a lawyer finds a loophole in the law it must be changed.
- If there is a lawyer who finds a loophole in the law it must be changed.
- For all lawyers: if one of them finds a loophole in the law it must be changed.

Often, the type of quantification in linguistic bits of information can be made out only from the context, and this leads not all too seldom to unintentional (or intentional) misunderstandings. "Logical relations in language are almost always just alluded to, left to guesswork, and not actually expressed" (G. Frege).

Let x, y be distinct variables and $\alpha \in \mathcal{L}$. One of the most important logical equivalences is *renaming of bound variables* (in short, *bound renaming*), stated in

(11) (a) $\forall x\alpha \equiv \forall y(\alpha\,\frac{y}{x})$, (b) $\exists x\alpha \equiv \exists y(\alpha\,\frac{y}{x})$ $(y \notin \mathit{var}\,\alpha)$.

(b) follows from (a) by rearranging equivalently. Note that $y \notin \mathit{var}\,\alpha$ is equivalent to $y \notin \mathit{free}\,\alpha$ and $\alpha, \frac{y}{x}$ collision-free. Writing $\mathcal{M}_x^y$ for $\mathcal{M}_x^{y^{\mathcal{M}}}$, (a) derives as follows:

$$\begin{aligned}
\mathcal{M} \vDash \forall x\alpha &\Leftrightarrow \mathcal{M}_x^a \vDash \alpha \quad \text{for all } a \quad \text{(definition)}\\
&\Leftrightarrow (\mathcal{M}_y^a)_x^a \vDash \alpha \quad \text{for all } a \quad \text{(Theorem 3.1)}\\
&\Leftrightarrow (\mathcal{M}_y^a)_x^y \vDash \alpha \quad \text{for all } a \quad \big((\mathcal{M}_y^a)_x^y = (\mathcal{M}_y^a)_x^a\big)\\
&\Leftrightarrow \mathcal{M}_y^a \vDash \alpha\,\tfrac{y}{x} \quad \text{for all } a \quad \text{(Theorem 3.5)}\\
&\Leftrightarrow \mathcal{M} \vDash \forall y(\alpha\,\tfrac{y}{x}).
\end{aligned}$$

The equivalences (12) and (13) below are also noteworthy. According to (13), substitutions are completely described up to logical equivalence by so-called *free renamings* (substitutions of the form $\frac{y}{x}$). (13) also embraces the case $x \in \mathit{var}\,t$.

(12) $\forall x(x = t \to \alpha) \equiv \alpha\,\frac{t}{x} \equiv \exists x(x = t \wedge \alpha)$ $(\alpha, \frac{t}{x}$ collision-free, $x \notin \mathit{var}\,t)$.

(13) $\forall y(y = t \to \alpha\,\frac{y}{x}) \equiv \alpha\,\frac{t}{x} \equiv \exists y(y = t \wedge \alpha\,\frac{y}{x})$ $(\alpha, \frac{t}{x}$ collision-free, $y \notin \mathit{var}\,\alpha, t)$.

Proof of (12): $\forall x(x = t \to \alpha) \vDash (x = t \to \alpha)\,\frac{t}{x} = t = t \to \alpha\,\frac{t}{x} \vDash \alpha\,\frac{t}{x}$ by Corollary 3.6. Conversely, let $\mathcal{M} \vDash \alpha\,\frac{t}{x}$ so that $\mathcal{M}_x^{t^{\mathcal{M}}} \vDash \alpha$ and $\mathcal{M}_x^a \vDash x = t$. Then $a = t^{\mathcal{M}}$ and so $\mathcal{M}_x^a \vDash \alpha$, which shows that $\mathcal{M}_x^a \vDash x = t \to \alpha$ for any $a \in A$, hence $\mathcal{M} \vDash \forall x(x = t \to \alpha)$. This proves the left equivalence in (12). The right equivalence reduces to the left one because $\exists x(x = t \wedge \alpha) = \neg\forall x\neg(x = t \wedge \alpha) \equiv \neg\forall x(x = t \to \neg\alpha) \equiv \neg\neg\alpha\,\frac{t}{x} \equiv \alpha\,\frac{t}{x}$.

Item (13) is proved similarly, using Corollary 3.6 and Exercise 1 in **2.2**. Observe that $\forall y(y = t \to \alpha\,\frac{y}{x}) \vDash \alpha\,\frac{y}{x}\,\frac{t}{y} = \alpha\,\frac{t}{x}$ and $\alpha\,\frac{t}{x}\,\frac{t}{y} \vDash \alpha\,\frac{y}{x}$.

With the above equivalences we can now regain an equivalent formula starting with any formula in which all quantifiers are standing at the beginning. But this one requires both quantifiers $\exists$ and $\forall$, in the following denoted by $Q, Q_1, Q_2, \ldots$

A formula of the form $\alpha = Q_1x_1 \cdots Q_nx_n\beta$ with an open formula β is termed a *prenex formula* or a *prenex normal form*, in short a PNF. The open β is also called the *kernel* of α. We may assume that $x_1, \ldots, x_n$ are distinct; this can always be achieved by bound renaming. These normal forms are, for instance, highly important for classifying definable number-theoretic predicates in **6.3**. Obviously, $\forall$-formulas and $\exists$-formulas are the simplest examples of prenex normal forms.

Theorem 4.2 (on the prenex normal form). *Every formula φ is equivalent to a formula in prenex normal form that can effectively be constructed from φ.*

Proof. Without loss of generality let φ contain only the logical symbols $\neg, \wedge, \forall, \exists$ (besides $=$). For each prefix Qx in φ consider the number of symbols $\neg$ or $\wedge$ standing in front of Qx in φ. Let $s\varphi$ be the sum of these numbers, summed over all prefixes occurring in φ. Clearly, φ is a PNF if and only if $s\varphi = 0$. Let $s\varphi \neq 0$. In view of

$$\neg\forall x\alpha \equiv \exists x\neg\alpha, \quad \neg\exists x\alpha \equiv \forall x\neg\alpha, \quad \beta \wedge Qx\alpha \equiv Qy(\beta \wedge \alpha\tfrac{y}{x}) \text{ for } y \notin \operatorname{var}\alpha, \beta,$$

$s\varphi$ can obviously be reduced stepwise by means of equivalent replacements. ❑

Example 2. $\forall x\exists y(x \neq 0 \to x \cdot y = 1)$ is a PNF for $\forall x(x \neq 0 \to \exists y\, x \cdot y = 1)$. And for $\exists x\varphi \wedge \forall y\forall z(\varphi\frac{y}{x} \wedge \varphi\frac{z}{x} \to y = z)$ we get the PNF $\exists x\forall y\forall z(\varphi \wedge (\varphi\frac{y}{x} \wedge \varphi\frac{z}{x} \to y = z))$ if $y, z \notin \mathit{free}\,\varphi$; if not, a bound renaming will help. An equivalent PNF for this formula with minimal quantifier rank is $\exists x\forall y(\varphi\frac{y}{x} \leftrightarrow x = y)$, see page 57.

The first formula $\forall x(x \neq 0 \to \exists y\, x \cdot y = 1)$ from the example may be abbreviated by $(\forall x \neq 0)\exists y\, x \cdot y = 1$. More generally, we shall write $(\forall x \neq t)\alpha$ for $\forall x(x \neq t \to \alpha)$ and $(\exists x \neq t)\alpha$ for $\exists x(x \neq t \wedge \alpha)$ from now on. A similar notation is used for $\leqslant$, $<$, $\in$ and their negations. For instance, $(\forall x \leqslant t)\alpha$ and $(\exists x \leqslant t)\alpha$ are to mean $\forall x(x \leqslant t \to \alpha)$ and $\exists x(x \leqslant t \wedge \alpha)$, respectively. For any binary relation symbol $\lhd$, the "prefixes" $(\forall y \lhd x)$ and $(\exists y \lhd x)$ are related to each other as are $\forall$ and $\exists$; see Exercise 2.

Exercises

1. Suppose $\alpha \equiv \beta$. Prove $\alpha\frac{\vec{t}}{\vec{x}} \equiv \beta\frac{\vec{t}}{\vec{x}}$ whenever $\alpha, \frac{\vec{t}}{\vec{x}}$ and $\beta, \frac{\vec{t}}{\vec{x}}$ are collision-free.

2. Prove that $\neg(\forall x \lhd y)\alpha \equiv (\exists x \lhd y)\neg\alpha$ and $\neg(\exists x \lhd y)\alpha \equiv (\forall x \lhd y)\neg\alpha$. Here $\lhd$ represents any binary relation symbol.

3. Show that the conjunction or disjunction of $\forall$-formulas α, β is equivalent to a $\forall$-formula. Prove the same for $\exists$-formulas (use bounded renaming if necessary).

4. Let P be a unary predicate symbol. Prove that $\exists x(Px \to \forall y Py)$ is a tautology.

5. Call $\alpha, \beta \in \mathcal{L}$ *tautologically equivalent* if $\vDash \alpha \Leftrightarrow \vDash \beta$. Confirm that the following (in general not logically equivalent) formulas are tautologically equivalent: α, $\forall x\alpha$, and $\alpha\frac{c}{x}$, where the constant symbol c does not occur in α.

2.5 Logical Consequence and Theories

Whenever $\mathcal{L}' \supseteq \mathcal{L}$, the language $\mathcal{L}'$ is called an *expansion* or *extension* of $\mathcal{L}$ and $\mathcal{L}$ a *reduct* or *restriction* of $\mathcal{L}'$. Recall the insensitivity of the consequence relation to extensions of a language, mentioned in **2.3**. Theorem 3.1 yields that establishing $X \models \alpha$ does not depend on the language to which the set of formulas X and the formula α belong. For this reason, indices for $\models$, such as $\models_{\mathcal{L}}$, are dispensable.

Because of the unaltered satisfaction conditions for $\wedge$ and $\neg$, all properties of the propositional consequence gained in **1.3** carry over to predicate logic. These include general properties such as, for example, the reflexivity and transitivity of $\models$, and the semantical counterparts of the rules $(\wedge 1)$, $(\wedge 2)$, $(\neg 1)$, $(\neg 2)$ from **1.4**, for instance $\dfrac{X \models \alpha, \beta}{X \models \alpha \wedge \beta}$.[5] Also, Gentzen-style properties such as the deduction theorem, automatically carry over. But there are also completely new properties among the following ones. Some of these will be elevated to basic rules of a logical calculus for first-order languages in **3.1**.

Examples of properties of the predicate logical consequence relation

(a) $\dfrac{X \models \forall x\alpha}{X \models \alpha\,\frac{t}{x}}$ $(\alpha, \frac{t}{x}$ collision-free), (b) $\dfrac{X \models \alpha\,\frac{s}{x}, s = t}{X \models \alpha\,\frac{t}{x}}$ $\left(\begin{smallmatrix}\alpha, \frac{s}{x} \text{ and } \alpha, \frac{t}{x} \\ \text{collision-free}\end{smallmatrix}\right)$,

(c) $\dfrac{X, \beta \models \alpha}{X, \forall x\beta \models \alpha}$ $\left(\begin{smallmatrix}\text{anterior} \\ \text{generalization}\end{smallmatrix}\right)$, (d) $\dfrac{X \models \alpha}{X \models \forall x\alpha}$ $\left(\begin{smallmatrix}x \notin \mathit{free}\, X, \text{ poste-} \\ \text{rior generalization}\end{smallmatrix}\right)$,

(e) $\dfrac{X, \beta \models \alpha}{X, \exists x\beta \models \alpha}$ $\left(\begin{smallmatrix}x \notin \mathit{free}\, X, \alpha, \text{ anter-} \\ \text{ior particularization}\end{smallmatrix}\right)$, (f) $\dfrac{X \models \alpha\,\frac{t}{x}}{X \models \exists x\alpha}$ $\left(\begin{smallmatrix}\alpha, t \text{ collision-free,} \\ \text{posterior particul.}\end{smallmatrix}\right)$.

Since $\models$ is transitive, (a) and (b) follow from $\forall x\alpha \models \alpha\,\frac{t}{x}$ and $\alpha\,\frac{s}{x}, s = t \models \alpha\,\frac{t}{x}$. This was already stated in Corollary 3.6. Analogously (c) results from $\forall x\beta \models \beta$. To prove (d), suppose that $X \models \alpha$, $\mathcal{M} \models X$, and $x \notin \mathit{free}\, X$. Then $\mathcal{M}^a_x \models X$ for any $a \in A$ by Theorem 3.1, which just means $\mathcal{M} \models \forall x\alpha$. As regards (e), observe that $X, \beta \models \alpha \;\Rightarrow\; X, \neg\alpha \models \neg\beta \;\Rightarrow\; X, \neg\alpha \models \forall x\neg\beta$ and (d), whence $X, \neg\forall x\neg\beta \models \alpha$. (e) captures deduction from an existence claim. (f) proves an existence claim and holds since $\alpha\,\frac{t}{x} \models \exists x\alpha$ by Corollary 3.6. Both (e) and (f) are permanently applied in mathematical reasoning and will briefly be discussed in Example 1 on the next page. All of the above properties have certain variants; for example,

(g) $\dfrac{X \models \alpha\,\frac{y}{x}}{X \models \forall x\alpha}$ $(y \notin \mathit{free}\, X \cup \mathit{var}\, \alpha)$.

This results from (d) with $\alpha\,\frac{y}{x}$ for α and y for x, because $\forall y\alpha\,\frac{y}{x} \equiv \forall x\alpha$ if $y \notin \mathit{var}\, \alpha$.

[5] A suggestive way of writing "$X \models \alpha, \beta$ implies $X \models \alpha \wedge \beta$," a notation that was introduced already in Exercise 3 in **1.3**. A corresponding notation will also be used in the examples below.

From these properties complicated chains of deduction can where necessary be justified step by step. But in practice this makes sense only in particular circumstances, because formalized proofs are readable only at the expense of a lot of time, just like lengthy computer programs, even with well prepared documentation.

What is most important is that a proof, when written down, can be understood and reproduced. This is why mathematical deduction tends to proceed *informally*, i.e., both claims and their proofs are formulated in a mathematical "everyday" language with the aid of fragmentary and flexible formalization. To what degree a proof is to be formalized depends on the situation and need not be determined in advance. In this way the strict syntactic structure of formal proofs is slackened, compensating for the imperfection of our brains in regard to processing syntactic information. Further, certain informal proof methods will often be described by a more or less clear reference to so-called background knowledge, and not actually carried out. This method has proven itself to be sufficiently reliable. Indeed, apart from specific cases it has not yet been bettered by any of the existing automatic proof machines. Let us present and analyse a very simple example of an informal proof in a language $\mathcal{L}$ for natural numbers that along with $0, 1, +, \cdot$ contains the symbol $\mid$ for divisibility, defined by $m \mid n \Leftrightarrow \exists k\, m \cdot k = n$. In addition, let $\mathcal{L}$ contain a symbol f for some function from $\mathbb{N}$ to $\mathbb{N}$; we shall write here f_i for $\mathsf{f}(i)$.

Example 1. We want to prove $\forall n \exists x (\forall i \leqslant n) \mathsf{f}_i \mid x$. That is, for every n, $\mathsf{f}_0, \ldots, \mathsf{f}_n$ have a common multiple. A careful proof proceeds by induction on n. Here we focus solely on the induction step $X, \exists x (\forall i \leqslant n) \mathsf{f}_i \mid x \vDash \exists x (\forall i \leqslant n+1) \mathsf{f}_i \mid x$, where X represents our prior knowledge about familiar properties of divisibility. Informally we reason as follows: Suppose $\exists x (\forall i \leqslant n) \mathsf{f}_i \mid x$ and let x denote any common multiple of $\mathsf{f}_0, \ldots, \mathsf{f}_n$. Then $x \cdot \mathsf{f}_{n+1}$ is obviously a common multiple of $\mathsf{f}_0, \ldots, \mathsf{f}_{n+1}$, whence we infer $\exists x (\forall i \leqslant n+1) \mathsf{f}_i \mid x$. That's all. To argue formally like a proof machine, we start from the obvious $(\forall i \leqslant n) \mathsf{f}_i \mid x \vDash (\forall i \leqslant n+1) \mathsf{f}_i \mid (x \cdot \mathsf{f}_{n+1})$. Posterior particularization of x is applied to get $X, (\forall i \leqslant n) \mathsf{f}_i \mid x \vDash \exists x (\forall i \leqslant n+1) \mathsf{f}_i \mid x$. Thereafter anterior particularization is used to obtain the desired $X, \exists x (\forall i \leqslant n) \mathsf{f}_i \mid x \vDash \exists x (\forall i \leqslant n+1) \mathsf{f}_i \mid x$.

Some textbooks deal with a somewhat stricter consequence relation, which we denote here by $\overset{\mathsf{G}}{\vDash}$. The reason is that in mathematics one largely considers derivations in theories. For $X \subseteq \mathcal{L}$ and $\varphi \in \mathcal{L}$ define $X \overset{\mathsf{G}}{\vDash} \varphi$ if $\mathcal{A} \vDash \varphi$ for all $\mathcal{L}$-structures $\mathcal{A}$ such that $\mathcal{A} \vDash X$. In contrast to the *local* consequence relation $\vDash$, $\overset{\mathsf{G}}{\vDash}$ can be considered as the *global* consequence relation since it cares only about $\mathcal{A}$, not about a concrete valuation w in $\mathcal{A}$, and hence not on pairs $(\mathcal{A}, w)$.

Let us collect a few properties of $\overset{\mathsf{G}}{\vDash}$. Obviously, $X \vDash \varphi$ implies $X \overset{\mathsf{G}}{\vDash} \varphi$, but the converse does not hold in general. For example, $x = y \overset{\mathsf{G}}{\vDash} \forall xy\, x = y$, however $x = y \nvDash \forall xy\, x = y$. By (d) from the beginning of this section, $X \vDash \varphi \Rightarrow X \vDash \varphi^{\mathsf{G}}$

holds in general only if the free variables of φ do not occur free in X, while $\overset{\mathfrak{G}}{\models}$ has this property unrestrictedly; indeed, for any X, by definition, $X \overset{\mathfrak{G}}{\models} \varphi \Leftrightarrow X \overset{\mathfrak{G}}{\models} \varphi^{\mathfrak{G}}$. A reduction of $\overset{\mathfrak{G}}{\models}$ to $\models$ is provided by the following equivalence which follows from $\mathcal{M} \models X^{\mathfrak{G}} \Leftrightarrow \mathcal{A} \models X^{\mathfrak{G}}$, for each model $\mathcal{M} = (\mathcal{A}, w)$:

(1) $X \overset{\mathfrak{G}}{\models} \varphi \Leftrightarrow X^{\mathfrak{G}} \models \varphi$.

Because of $S^{\mathfrak{G}} = S$ for sets of sentences S, we clearly obtain from (1)

(2) $S \overset{\mathfrak{G}}{\models} \varphi \Leftrightarrow S \models \varphi$ (in particular, $\overset{\mathfrak{G}}{\models} \varphi \Leftrightarrow \models \varphi$).

Thus, a distinction between $\models$ and $\overset{\mathfrak{G}}{\models}$ is apparent only when premises are involved that are not sentences. In such a situation the relation $\overset{\mathfrak{G}}{\models}$ must be treated with the utmost care. In particular, neither of the rules

$$\frac{X, \alpha \overset{\mathfrak{G}}{\models} \beta \,\big|\, X, \neg\alpha \overset{\mathfrak{G}}{\models} \beta}{X \overset{\mathfrak{G}}{\models} \beta} \text{ (case distinction)}, \quad \frac{X, \alpha \overset{\mathfrak{G}}{\models} \beta}{X \overset{\mathfrak{G}}{\models} \alpha \to \beta} \text{ (deduction theorem)}$$

is unrestrictedly correct; for example $x = y \overset{\mathfrak{G}}{\models} \forall xy\, x = y$, but not $\overset{\mathfrak{G}}{\models} x = y \to \forall xy\, x = y$. Thus, the deduction theorem fails to hold for $\overset{\mathfrak{G}}{\models}$. A main reason for our preference of $\models$ over $\overset{\mathfrak{G}}{\models}$ is that $\models$ extends the propositional consequence relation conservatively, so that features such as the deduction theorem carry over unrestrictedly, while this is not the case for $\overset{\mathfrak{G}}{\models}$. It should also be said that $\overset{\mathfrak{G}}{\models}$ reflects only incompletely the actual procedures of natural deduction in that formulas with free variables are frequently used also in deductions of sentences from sentences as is seen in Example 1.

We now make more precise the notion of a formalized theory in $\mathcal{L}$, where it is useful to think of the examples in **2.3**, such as group theory.

Definition. An *elementary theory* or *first-order theory* in $\mathcal{L}$, also termed an $\mathcal{L}$-*theory*, is a set of sentences $T \subseteq \mathcal{L}^0$ deductively closed in $\mathcal{L}^0$, i.e., $T \models \alpha \Leftrightarrow \alpha \in T$, for all $\alpha \in \mathcal{L}^0$. If $\alpha \in T$ then we say that α *is valid* or *holds in* T, or α *is a theorem* of T. The extralogical symbols of $\mathcal{L}$ are also called the symbols of T. If $T \subseteq T'$ then T is called a *subtheory* of T', and T' an *extension* of T. An $\mathcal{L}$-structure $\mathcal{A}$ such that $\mathcal{A} \models T$ is also termed a *model of* T, in short a T-*model*. $\operatorname{Md} T$ denotes the class of all models of T in this sense; $\operatorname{Md} T$ consist of $\mathcal{L}$-structures only.

For instance, for any set X of sentences, $T = \{\alpha \in \mathcal{L}^0 \mid X \models \alpha\}$ is a theory, in view of the transitivity of $\models$. Clearly, $\alpha \in T$ if and only if $\mathcal{A} \models \alpha$ for all $\mathcal{A} \models T$.

According to (2), there is no difference between $\models$ and $\overset{\mathfrak{G}}{\models}$ as long as deduction from theories is considered. We always have $T \models \varphi \Leftrightarrow T \models \varphi^{\mathfrak{G}}$, for an arbitrary $\varphi \in \mathcal{L}$. This fact should be taken in and remembered, since it is constantly used.

Different authors may use different definitions for a theory. For example, it is not always demanded that theories consist only of sentences. Conventions of this type each have their advantages and disadvantages. Proofs regarding theories are always adaptable enough to accommodate small modifications of the definition. Using the definition given above we set the following convention.

Convention. In talking of the *theory* S where S is a set of sentences, we always mean the theory determined by S, that is, $\{\alpha \in \mathcal{L}^0 \mid S \vDash \alpha\}$. A set of formulas X is called an *axiom system* for T whenever $T = \{\alpha \in \mathcal{L}^0 \mid X^{\mathsf{G}} \vDash \alpha\}$. Thus, we tacitly generalize all possibly open formulas in X. Axioms of a theory are always sentences. But we conforme to standard practice of writing long axioms as formulas.

We will later consider extensive axiom systems (in particular, for arithmetic and for set theory) whose axioms are partly written as open formulas just for the reason of economy. Free variables occurring in axioms have always to be generalized.

There exists a smallest theory in $\mathcal{L}$, namely the set Taut $(= \mathsf{Taut}_{\mathcal{L}})$ of all generally valid sentences in $\mathcal{L}$, also called the "logical" theory. An axiom system for Taut is the empty set of axioms. There is also a largest theory: the set $\mathcal{L}^0$ *of all* sentences, the *inconsistent* theory which possesses no models. All remaining theories are called *satisfiable* or *consistent*.[6] Moreover, the intersection $T = \bigcap_{i\in I} T_i$ of any nonempty family of theories T_i is in turn a theory: if $T \vDash \alpha \in \mathcal{L}^0$ then clearly $T_i \vDash \alpha$ holds as well, for every $i \in I$. Hence, $T \vDash \alpha$ (equivalently, $\alpha \in T$). In this book T and T', with or without indices, exclusively denote theories.

For $T \subseteq \mathcal{L}^0$ and $\alpha \in \mathcal{L}^0$ let $T + \alpha$ denote the smallest extension of T containing α. Similarly let $T + S$ for $S \subseteq \mathcal{L}^0$ be the smallest theory $\supseteq T \cup S$. If S is finite then $T' = T + S = T + \bigwedge S$ is called a *finite extension of* T. Here $\bigwedge S$ denotes the conjunction of all sentences in S. A sentence α is termed *compatible* or *consistent* with T, if $T + \alpha$ is satisfiable, and *refutable in* T if $T + \neg\alpha$ is satisfiable. Thus, the theory of fields T_F is compatible with the sentence $1 + 1 = 0$, or $1 + 1 \neq 0$ is refutable in T_F, since the 2-element field satisfies $1 + 1 = 0$.

If both α and $\neg\alpha$ are compatible with T then the sentence α is termed *independent* of T. The classic example is the independence of the parallel axiom from the remaining axioms of Euclidean plane geometry which define *absolute* geometry. Much more difficult is the independence proof of the continuum hypothesis from the axioms for set theory. These axioms are presented and discussed in **3.4**.

At this point we introduce another important concept; $\alpha, \beta \in \mathcal{L}$ are said to be *equivalent in* or *modulo* T, $\alpha \equiv_T \beta$, if $\alpha \equiv_{\mathcal{A}} \beta$ for all $\mathcal{A} \vDash T$. Being an intersection of congruences, $\equiv_T$ is itself a congruence and hence satisfies the replacement theorem. This will henceforth be used without further mention, as will the obvious equivalence of $\alpha \equiv_T \beta$, $T \vDash \alpha \leftrightarrow \beta$, and $T \vDash (\alpha \leftrightarrow \beta)^{\mathsf{G}}$.

Example 2. In T_G (page 51) holds $x \circ x = x \equiv_{T_G} x = e \equiv_{T_G} \forall y\, y \circ x = y$. The only tricky step in the proof is $T_G \vDash x \circ x = x \to x = e$. Let $x \circ x = x$ and choose some y with $x \circ y = e$. This equation implies $x = x \circ e = x \circ x \circ y = x \circ y = e$ in T_G.

[6] *Consistent* mostly refers to a logic calculus, e.g., the calculus in **3.1**. However, it will be shown in **3.2** that consistency and satisfiability coincide, thus justifying the word's ambiguous use.

Terms s, t are called *equivalent* in T, symbolically $s \approx_T t$, if $T \vDash s = t$, that is, $\mathcal{A} \vDash s = t\,[w]$ for all $\mathcal{A} \vDash T$ and $w\colon Var \to A$. For instance, in the theory $T := T_G^=$ of groups is provable $(x \circ y)^{-1} = y^{-1} \circ x^{-1}$, equivalently, $(x \circ y)^{-1} \approx_T y^{-1} \circ x^{-1}$.

If all axioms of a theory T are $\forall$-sentences then T is called a *universal* or $\forall$-theory. For such a theory, $\mathrm{Md}\,T$ is closed with respect to substructures as follows from Corollary 3.3, that is, $\mathcal{A} \subseteq \mathcal{B} \vDash T \;\Rightarrow\; \mathcal{A} \vDash T$. Examples are partial orders, orders, lattices, Boolean algebras etc. Universal theories are further classified. The most important $\forall$-theories are equational, quasi-equational, and universal Horn theories, all of which will be considered to some extent in later chapters.

Theories are frequently given by structures or classes of structures. The elementary theory $Th\,\mathcal{A}$ and the theory $Th\,\boldsymbol{K}$ of a class $\boldsymbol{K}$ of structures are defined by

$$Th\,\mathcal{A} := \{\alpha \in \mathcal{L}^0 \mid \mathcal{A} \vDash \alpha\}, \quad Th\,\boldsymbol{K} := \bigcap\{Th\,\mathcal{A} \mid \mathcal{A} \in \boldsymbol{K}\},$$

where we tacitly assume $\boldsymbol{K} \neq \emptyset$. It is easy to verify that here theories in the precise sense are being dealt with. Instead of $\alpha \in Th\,\boldsymbol{K}$ one often writes $\boldsymbol{K} \vDash \alpha$. In general, $\mathrm{Md}\,Th\,\boldsymbol{K}$ is larger than $\boldsymbol{K}$ as we shall see.

Remark. The set of formulas breaks up modulo T (more precisely, modulo $\equiv_T$) into equivalence classes; their totality is denoted by $B_\omega T$. Based on these we can define in a natural manner operations $\wedge, \vee, \neg$. For instance, $\bar{\alpha} \wedge \bar{\beta} = \overline{\alpha \wedge \beta}$ where $\bar{\varphi}$ denotes the equivalence class to which φ belongs. One shows easily that $B_\omega T$ forms a Boolean algebra with respect to $\wedge, \vee, \neg$. For every n, also the set $B_n T$ of all $\bar{\varphi}$ in $B_\omega T$ such that the free variables of φ belong to Var_n $(= \{\boldsymbol{v}_0, \ldots, \boldsymbol{v}_{n-1}\})$ is a subalgebra of $B_\omega T$. Note that $B_0 T$ is isomorphic to the Boolean algebra of all sentences modulo $\equiv_T$. The significance of the Boolean algebras $B_n T$ is revealed only in the somewhat higher reaches of model theory, and they are therefore mentioned only incidentally.

Exercises

1. Suppose $x \notin free\,X$ and c is not in X, α. Prove the equivalence of

   $$\text{(i) } X \vDash \alpha, \qquad \text{(ii) } X \vDash \forall x \alpha, \qquad \text{(iii) } X \vDash \alpha\,\tfrac{c}{x}.$$

   This holds then in particular if X is a theory or the axiom system of a theory.

2. Let S be a set of sentences, $\alpha = \alpha(x)$ and β formulas, and c be a constant not occurring in S, α, β. Show that the following statements are equivalent:

   $$\text{(i) } S \vDash \alpha\,\tfrac{c}{x} \to \beta, \qquad \text{(ii) } S \vDash \exists x \alpha \to \beta.$$

3. Show for all $\alpha, \beta \in \mathcal{L}^0$ that $\beta \in T + \alpha \;\Leftrightarrow\; \alpha \to \beta \in T$.

4. Let $T \subseteq \mathcal{L}$ be a theory, $\mathcal{L}_0 \subseteq \mathcal{L}$, and $T_0 := T \cap \mathcal{L}_0$. Prove that T_0 is also a theory (the so-called *reduct theory* in the language $\mathcal{L}_0$).

2.6 Explicit Definitions–Expanding Languages

The deductive development of a theory, be it given by an axiom system or a single structure or classes of those, nearly always goes hand in hand with expansions of the language carried out step by step. For example, in developing elementary number theory in the language $\mathcal{L}(0, 1, +, \cdot)$, the introduction of the divisibility relation by means of the (explicit) definition $x|y \leftrightarrow \exists z\, x \cdot z = y$ has certainly some advantages. This and similar examples motivate the following

Definition I. Let r be an n-ary relation symbol not in $\mathcal{L}$. An *explicit definition of r in $\mathcal{L}$* is a formula of the following form, with distinct variables in $\vec{x}$:

$\eta_r: \quad r\vec{x} \leftrightarrow \delta(\vec{x})$

with $\delta(\vec{x}) \in \mathcal{L}$, named the *defining formula*. For a theory T, $T_r := T + \eta_r^{\mathfrak{G}}$ is then called a *definitorial* expansion (or extension) of T by r. This is a theory in $\mathcal{L}[r]$, the language resulting from $\mathcal{L}$ by adjoining the relation symbol r.

T_r is a *conservative extension* of T which, in general, is to mean a theory $T' \supseteq T$ in $\mathcal{L}' \supseteq \mathcal{L}$ such that $T' \cap \mathcal{L} = T$. Thus, no new sentences from the language of T are added to T. In this sense T_r is a harmless extension of T. Our claim constitutes part of Theorem 6.1. For $\varphi \in \mathcal{L}[r]$ define the *reduced formula* $\varphi^{rd} \in \mathcal{L}$ as follows: Starting from the left, replace every prime formula $r\vec{t}$ occurring in φ by $\delta_{\vec{x}}(\vec{t})$.

Theorem 6.1 (Elimination theorem). *Let $T_r \subseteq \mathcal{L}[r]$ be a definitorial extension of $T \subseteq \mathcal{L}^0$ by the explicit definition $r\vec{x} \leftrightarrow \delta(\vec{x})$. Then for all $\varphi \in \mathcal{L}[r]$*

$$(*) \quad T_r \vDash \varphi \;\Leftrightarrow\; T \vDash \varphi^{rd}.$$

For $\varphi \in \mathcal{L}$ we have in particular $T_r \vDash \varphi \;\Leftrightarrow\; T \vDash \varphi$ (because then $\varphi^{rd} = \varphi$). Hence, $\alpha \in T_r \;\Leftrightarrow\; \alpha \in T$, for all $\alpha \in \mathcal{L}^0$. In short, T_r is a conservative extension of T.

Proof. Each $\mathcal{A} \vDash T$ is expandable to a model $\mathcal{A}' \vDash T_r$ with the same domain, setting $r^{\mathcal{A}'}\vec{a} :\Leftrightarrow \mathcal{A} \vDash \delta[\vec{a}]$ $(\vec{a} \in A^n)$. Since $r\vec{t} \equiv_{T_r} \delta(\vec{t})$ for any term sequence $\vec{t}$, we have $\varphi \equiv_{T_r} \varphi^{rd}$ for all $\varphi \in \mathcal{L}[r]$ (replacement theorem). Thus, $(*)$ follows from

$$\begin{aligned} T_r \vDash \varphi &\Leftrightarrow \mathcal{A}' \vDash \varphi \text{ for all } \mathcal{A} \vDash T \quad (\mathrm{Md}\, T_r \text{ consist of the } \mathcal{A}' \text{ with } \mathcal{A} \vDash T) \\ &\Leftrightarrow \mathcal{A}' \vDash \varphi^{rd} \text{ for all } \mathcal{A} \vDash T \quad (\text{because } \varphi \equiv_{T_r} \varphi^{rd}) \\ &\Leftrightarrow \mathcal{A} \vDash \varphi^{rd} \text{ for all } \mathcal{A} \vDash T \quad (\text{Theorem 3.1}) \\ &\Leftrightarrow T \vDash \varphi^{rd}. \qquad \square \end{aligned}$$

Operation symbols and constants can be similarly introduced, though in that case there are certain conditions to observe. For instance, in the theory of groups T_G (page 51) the operation $^{-1}$ can be defined by $y = x^{-1} \leftrightarrow x \circ y = e$. This definition is legitimate since $T_G \vDash \forall x \exists! y\, x \circ y = e$. Only this requirement ensures that $T_G + \eta^{\mathfrak{G}}$ is a conservative extension of T_G; Exercise 3. We therefore extend Definition I as

follows, keeping in mind that to the end of this section constant symbols are to be counted among the operation symbols.

Definition II. An *explicit definition* of an n-ary operation symbol f not occurring in $\mathcal{L}$ is a formula of the form

η_f : $\quad y = f\vec{x} \leftrightarrow \delta(\vec{x}, y) \qquad (\delta \in \mathcal{L}$ and $y, x_1, \ldots, x_n$ distinct).

η_f is called *legitimate* in $T \subseteq \mathcal{L}$ if $T \vDash \forall \vec{x}\, \exists! y \delta$, and $T_f := T + \eta_f^{\mathfrak{G}}$ is then called a *definitorial extension by* f. In the case $n = 0$ we write c for f and speak of an *explicit definition of the constant symbol* c. It is of the form $y = c \leftrightarrow \delta(y)$.

Some of the free variables of δ are often not explicitly named, and thus downgraded to parameter variables. More on this will be said in the discussion of the axioms for set theory in **3.4**. The elimination theorem is proved in almost exactly the same way as above, provided η_f is legitimate in T. The reduced formula φ^{rd} is defined correspondingly. For a constant c ($n = 0$ in Definition II), let $\varphi^{rd} := \exists z(\varphi\, \frac{z}{c} \wedge \delta\, \frac{z}{y})$, where $\varphi\, \frac{z}{c}$ denotes the result of replacing c in φ by z ($\notin var\,\varphi$). Now let $n > 0$. If f does not appear in φ, set $\varphi^{rd} = \varphi$. Otherwise, looking at the first occurrence of f in φ from the left, we certainly may write $\varphi = \varphi_0\, \frac{f\vec{t}}{y}$ for appropriate φ_0, $\vec{t}$, and $y \notin var\,\varphi$. Clearly, $\varphi \equiv_{T_f} \exists y(\varphi_0 \wedge y = f\vec{t}) \equiv_{T_f} \varphi_1$, with $\varphi_1 := \exists y(\varphi_0 \wedge \delta_f(\vec{t}, y))$. If f still occurs in φ_1 then repeat this procedure, which ends in, say, m steps in a formula φ_m that no longer contains f. Then set $\varphi^{rd} := \varphi_m$.

Frequently, operation symbols f are introduced by definitions of the form

$$(*) \quad f\vec{x} := t(\vec{x})$$

where of course f does not occur in the term $t(\vec{x})$. This procedure is in fact subsumed by Definition II, because the former is nothing more than a definitorial extension of T with the explicit definition η_f: $y = f\vec{x} \leftrightarrow y = t(\vec{x})$. This definition is legitimate since $\forall \vec{x}\, \exists! y\, y = t(\vec{x})$ is a tautology. It can readily be shown that $\eta_f^{\mathfrak{G}}$ is logically equivalent to $\forall \vec{x}\, f\vec{x} = t(\vec{x})$. Hence, $(*)$ can indeed be regarded as a kind of an informative abbreviation of a legitimate explicit definition with the defining formula $y = t(\vec{x})$.

Remark. Instead of introducing new operation symbols, so-called *iota-terms* from [HB] could be used. For any formula $\varphi = \varphi(\vec{x}, y)$ in a given language, let $\iota y \varphi$ be a term in which y appears as a variable *bound by* ι. Whenever $T \vDash \forall \vec{x} \exists! y \varphi$ then T is extended by the axiom $\forall \vec{x} \forall y[y = \iota y \varphi(\vec{x}, y) \leftrightarrow \varphi(\vec{x}, y)]$ so that $\iota y \varphi(\vec{x}, y)$ so to speak stands for the function term $f\vec{x}$, which could have been introduced by an explicit definition. We mention that a definitorial language expansion is not a necessity. In principle, formulas of the expanded language can always be understood as abbreviations in the original language. This is in some presentations the actual procedure, though our imagination prefers additional notions over long sentences that would arise if we were to stick to the basic notions.

Definitions I and II can be unified in a more general declaration as follows: T' is a *definitorial extension* of T whenever $T' = T + \Delta$ for some list Δ of explicit definitions

of new symbols legitimate in T, given in terms of those of T (here *legitimate* is meant to pertain to operation symbols and constants only). Δ need not be finite, but in principle it is sufficient to restrict ourselves to this case. If $\mathcal{L}'$ is the language of T', a reduced formula $\varphi^{rd} \in \mathcal{L}$ is stepwise constructed as above, for every $\varphi \in \mathcal{L}$. In this way the somewhat long-winded proof of the following theorem is reduced each time to the case for extension by a single symbol:

Theorem 6.2 (General elimination theorem). *Let T' be a definitorial extension of T. Then $\alpha \in T' \Leftrightarrow \alpha^{rd} \in T$, and T' is a conservative extension of T.*

A relation or operation symbol ζ occurring in $T \subseteq \mathcal{L}$ is termed *explicitly definable in T* if T is a definitorial extension of $T_0 := T \cap \mathcal{L}_0$, where $\mathcal{L}_0$ denotes the language of the extralogical symbols of T without ζ. For example, in the theory T_G of groups the constant e is explicitly defined by $x = e \leftrightarrow x \circ x = x$ (Example 2 page 65). In such a case each T_0-model can be expanded in only one way to a T-model. If this special condition is fulfilled then ζ is also called *implicitly definable* in T. This could also be stated as follows: if T' is distinct from T only in that the symbol ζ is everywhere replaced by a new symbol ζ', then $T \cup T' \vDash \forall \vec{x}(\zeta\vec{x} \leftrightarrow \zeta'\vec{x})$ or $T \cup T' \vDash \forall \vec{x}(\zeta\vec{x} = \zeta'\vec{x})$, depending on whether ζ, ζ' are relation or operation symbols. It is noteworthy that the latter is already sufficient for the explicit definability of ζ in T. But we will go without the proof, preferring instead to quote the following interesting theorem:

Beth's definability theorem. *A relation or operation symbol implicitly definable in a theory T is also explicitly definable in T.*

Definitorial expansions of a language should be conscientiously distinguished from expansions of languages that arise from the introduction of so-called *Skolem functions*. These are useful for many purposes and are therefore briefly described.

Skolem normal forms. According to Theorem 4.2, every formula α can be converted into an equivalent PNF, $\alpha \equiv Q_1 x_1 \cdots Q_k x_k \alpha'$, where α' is open. Obviously then $\neg\alpha \equiv \overline{Q}_1 x_1 \cdots \overline{Q}_k x_k \neg\alpha'$, where $\overline{\forall} = \exists$ and $\overline{\exists} = \forall$. Because $\vDash \alpha$ if and only if $\neg\alpha$ is unsatisfiable, the decision problem for general validity can first of all be reduced to the satisfiability problem for formulas in PNF. Using Theorem 6.3 below, the latter—at the cost of introducing new operation symbols—is then completely reduced to the satisfiability problem for $\forall$-formulas.

Call formulas α and β *satisfiably equivalent* if both are satisfiable (not necessarily in the same model), or both are unsatisfiable. We construct for every formula, which w.l.o.g. is assumed to be given in prenex form $\alpha = Q_1 x_1 \cdots Q_k x_k \beta$, a satisfiably equivalent $\forall$-formula $\hat{\alpha}$ with additional operation symbols such that $\mathit{free}\,\hat{\alpha} = \mathit{free}\,\alpha$. The construction of $\hat{\alpha}$ will be completed after m steps, where m is the number of $\exists$-quantifiers among the $Q_1, \ldots, Q_k$. Take $\alpha = \alpha_0$ and α_i to be already constructed. If α_i is already an $\forall$-formula then let $\hat{\alpha} = \alpha_i$. Otherwise α_i has the form $\forall x_1 \cdots \forall x_n \exists y \beta_i$

for some $n \geqslant 0$. With an n-ary operation symbol f not yet used let $\alpha_{i+1} = \forall\vec{x}\beta_i \frac{f\vec{x}}{y}$. Thus, after m steps an $\forall$-formula $\hat{\alpha}$ is obtained such that $\mathit{free}\,\hat{\alpha} = \mathit{free}\,\alpha$; this formula is called a *Skolem normal form* (SNF) of α.

Example 1. If $\alpha = \forall x\exists y\, x < y$ then $\hat{\alpha} = \forall x\, x < fx$. For $\alpha = \exists x\forall y\, x \cdot y = y$ we have $\hat{\alpha} = \forall y\, c \cdot y = y$. If $\alpha = \forall x\forall y\exists z(x < z \wedge y < z)$ then $\hat{\alpha} = \forall x\forall y(x < fxy \wedge y < fxy)$.

Theorem 6.3. *Suppose that $\hat{\alpha}$ is a Skolem normal form for the formula α. Then*

(a) $\hat{\alpha} \vDash \alpha$, (b) *$\alpha$ is satisfiably equivalent to $\hat{\alpha}$.*

Proof. (a): It suffices to show that $\alpha_{i+1} \vDash \alpha_i$ for each of the described construction steps. $\beta_i \frac{f\vec{x}}{y} \vDash \exists y\beta_i$ implies $\alpha_{i+1} = \forall\vec{x}\beta_i \frac{f\vec{x}}{y} \vDash \forall\vec{x}\exists y\beta_i = \alpha_i$, by (a) and (d) in **2.5**. (b): If $\hat{\alpha}$ is satisfiable then by (a) so too is α. Conversely, suppose $\mathcal{A} \vDash \forall\vec{x}\exists y\beta_i(\vec{x}, y, \vec{z})\,[\vec{c}]$. For each $\vec{a} \in A^n$ we choose some $b \in A$ such that $\mathcal{A} \vDash \beta\,[\vec{a}, b, \vec{c}]$ and expand $\mathcal{A}$ to $\mathcal{A}'$ by setting $f^{\mathcal{A}'}\vec{a} = b$ for the new operation symbol. Then evidently $\mathcal{A}' \vDash \alpha_{i+1}\,[\vec{c}]$. Thus, we finally obtain a model for $\hat{\alpha}$ that expands the initial model. ❑

Now, for each α, a tautologically equivalent $\exists$-formula $\check{\alpha}$ (that is, $\vDash \alpha \Leftrightarrow \vDash \check{\alpha}$) is gained as well. By the above theorem, we first produce for $\beta = \neg\alpha$ a satisfiably equivalent SNF $\hat{\beta}$ and put $\check{\alpha} := \neg\hat{\beta}$. Then indeed $\vDash \alpha \;\Leftrightarrow\; \vDash \check{\alpha}$, because

$$\vDash \alpha \;\Leftrightarrow\; \beta \text{ unsatisfiable} \;\Leftrightarrow\; \hat{\beta} \text{ unsatisfiable} \;\Leftrightarrow\; \vDash \check{\alpha}.$$

Example 2. Let $\alpha := \exists x\forall y(ry \to rx)$. Clearly, $\neg\alpha \equiv \beta := \forall x\exists y(ry \wedge \neg rx)$ and $\hat{\beta} = \forall x(rfx \wedge \neg rx)$. Thus, $\check{\alpha} = \neg\hat{\beta} \equiv \exists x(rfx \to rx)$. The last formula is a tautology (in contrast to $\exists x(rx \to rfx)$). Thus, $\check{\alpha}$ and hence α are tautologies as well. This example shows how useful Skolem normal forms can be for discovering tautologies.

Exercises

1. Suppose T_f results from T by adjoining an explicit definition η for f and let α^{rd} be constructed as explained in the text. Show that T_f is a conservative extension of T if and only if η is a legitimate explicit definition.

2. Let $\mathtt{S}\colon n \mapsto n + 1$ denote the successor function in $\mathcal{N} = (\mathbb{N}, 0, \mathtt{S}, +, \cdot)$. Show that $Th\mathcal{N}$ is a definitorial extension of $Th\,(\mathbb{N}, \mathtt{S}, \cdot)$; in other words, 0 and $+$ are explicitly definable by $\mathtt{S}$ and $\cdot$ in $Th\mathcal{N}$.

3. Prove that $y = x^{-1} \leftrightarrow x \circ y = e$ is a legitimate explicit definition in T_G (which amounts to showing that $T_G \vDash x \circ y = e \wedge x \circ z = e \to y = z$). Moreover, prove that the resulting definitorial extension coincides with $T_G^{=}$.

4. Prove that the $<$-relation is not explicitly definable in $(\mathbb{Z}, 0, +)$.

5. Construct to each $\alpha \in X$ ($\subseteq \mathcal{L}$) a SNF $\hat{\alpha}$ (indexing the functions properly) such that X is satisfiably equivalent to $\hat{X} = \{\hat{\alpha} \mid \alpha \in X\}$ and $\hat{X} \vDash X$.

Chapter 3

Gödel's Completeness Theorem

Our goal is to characterize the consequence relation in a first-order language by means of a calculus similar to that of propositional logic. That this goal is attainable at all was shown for the first time by Gödel in [Go1]. The original version of Gödel's theorem refers to the axiomatization of tautologies only and does not immediately imply the compactness theorem of first-order logic; but a more general formulation of completeness in **3.2** does. The importance of the compactness theorem for mathematical applications was first revealed in 1936 by A. Malcev, see [Ma].

The characterizability of logical consequence by means of a calculus (the content of the completeness theorem) is a crucial result in mathematical logic with far-reaching applications. In spite of its metalogical origin, the completeness theorem is essentially a mathematical theorem. It satisfactorily explains the phenomenon of the well-definedness of logical deductive methods in mathematics. To seek any additional, possibly unknown methods or rules of inference would be like looking for perpetual motion in physics. Of course, this insight does not affect the development of new ideas in solving open questions. We will say somewhat more regarding the metamathematical aspect of the theorem and its applications, as well as the use of the model construction connected with its proof in a partly descriptive manner in the Sections **3.3**, **3.4**, and **3.5**.

Without beating around the bush, we deal from the outset with the case of an arbitrary, not necessarily countable first-order language. Nonetheless, the proof given, based on Henkin's idea of a constant expansion [He], is kept relatively short, mainly thanks to an astute choice of its logical basis. Although mathematical theories are countable as a rule, a successful application of methods of mathematical logic in algebra and analysis relies essentially on the unrestricted version of the completeness theorem. Only with such generality does the proof display the inherent unity that tends to distinguish the proofs of magnificent mathematical theorems.

3.1 A Calculus of Natural Deduction

As in Chapter **2**, let $\mathcal{L}$ be an arbitrary but fixed first-order language in the logical signature $\neg, \wedge, \forall, =$. We define a calculus $\vdash$ by the system of deductive rules enclosed in the box below. The calculus operates with sequents as in propositional logic. It supplements the basic rules of **1.4** with three predicate-logical rules. Note that the initial rule (IR) is subject to a minor extension.

(IR) $\dfrac{}{X \vdash \alpha}$ $(\alpha \in X \cup \{t{=}t\})$ (MR) $\dfrac{X \vdash \alpha}{X' \vdash \alpha}$ $(X \subseteq X')$

($\wedge$1) $\dfrac{X \vdash \alpha, \beta}{X \vdash \alpha \wedge \beta}$ ($\wedge$2) $\dfrac{X \vdash \alpha \wedge \beta}{X \vdash \alpha, \beta}$

($\neg$1) $\dfrac{X \vdash \beta, \neg\beta}{X \vdash \alpha}$ ($\neg$2) $\dfrac{X, \beta \vdash \alpha \mid X, \neg\beta \vdash \alpha}{X \vdash \alpha}$

($\forall$1) $\dfrac{X \vdash \forall x\alpha}{X \vdash \alpha\,\frac{t}{x}}$ ($\alpha, \frac{t}{x}$ collision-free) ($\forall$2) $\dfrac{X \vdash \alpha\,\frac{y}{x}}{X \vdash \forall x\alpha}$ ($y \notin \mathit{free}\, X \cup \mathrm{var}\,\alpha$)

(=) $\dfrac{X \vdash s{=}t, \alpha\,\frac{s}{x}}{X \vdash \alpha\,\frac{t}{x}}$ (α any prime formula)

By (IR), $X \vdash t{=}t$ for arbitrary X and t, in particular $\vdash t{=}t$. Here as everywhere, $\vdash \varphi$ stands for $\emptyset \vdash \varphi$. The remaining notation from Chapter **1** is also used here; thus, $\alpha \vdash \beta$ abbreviates $\{\alpha\} \vdash \beta$, etc. (IR) was formulated stronger than necessary only for convenience. Using (MR) it could be pared down to $\dfrac{}{\alpha \vdash \alpha}$ and $\dfrac{}{\vdash t{=}t}$.

We call $\vdash$ a *calculus of natural deduction* because it models logical inference in mathematics and other deductive sciences sufficiently well.[1] Our aim is to show that $\vDash$ is completely characterized by $\vdash$. Here the calculus is developed only insofar as the completeness proof requires. While undertaking further derivations can be instructive (see the examples and exercises), this is not the principal point of formalizing proofs unless one is after specific proof-theoretical goals. It should also be said that an acute study of formalized proofs does not really promote a human being's ability to draw correct conclusions in practice.

All basic rules are sound in the sense of **1.4**. The restrictions to the rules ($\forall$1), ($\forall$2), and (=) ensure their soundness as shown in Examples (a), (g), and (b) in **2.5**. Rule (=) could have been strengthened from the outset to allow α to be any formula such that $\alpha, \frac{s}{x}, \frac{t}{x}$ are collision-free, but we get along with the weak version. ($\forall$1) could

[1] We deal here with a version of the calculus NK from [Ge] adapted to our purpose; more involved descriptions of this and related sequent calculi are given in various textbooks on proof theory.

still be weakened; it suffices to require just $bnd\,\alpha \cap var\,t = \emptyset$. As already stated in **2.3**, we could in fact avoid any kind of restriction by means of a more involved and somewhat artificial definition for substitution. However, such measures would not simplify the matter. Weakly formulated logical calculi like the one given here often alleviate certain induction procedures, for example in proving soundness, or in verifying these rules in other logical calculi as will be done in **3.6**.

Because $\vdash$ can be understood as an extension of the corresponding calculus from **1.4**, all the examples of provable rules given there carry over automatically, the cut rule included. All further sound rules, such as the formal versions of generalization and particularization in **2.5**, are provable thanks to the completeness of the calculus. This is also true of the rule $\dfrac{X \vdash \alpha}{X \vdash \forall x\alpha}$ $(x \notin free\,X)$, which is sound by (d) in **2.5**, though it does not result directly from ($\forall$2). However, we do not want to spend too much time on the proofs of other rules; they are irrelevant for the completeness proof, which can then be used to justify these rules retrospectively.

Just as in the propositional case the following proof procedure will often be applied; it is legitimate because the proof of the corresponding principle in **1.4** depends neither on the type of language nor the concrete form of the rules.

Principle of rule induction. *Let $\mathcal{E}$ be a property of sequents (X, α) such that*

(o) $\mathcal{E}(X, \alpha)$ *provided* $\alpha \in X$ *or* α *is of the form* $t \mathrel{=\!\!=} t$,

(s) $\mathcal{E}(X, \alpha) \Rightarrow \mathcal{E}(X', \alpha)$ *for* (MR)*, and similarly for* ($\wedge$1) *through* (=).

Then $\mathcal{E}(X, \alpha)$ holds for all X, α such that $X \vdash \alpha$.

Since the basic rules are clearly sound, the *soundness of the calculus*, that is to say, $\vdash \subseteq \models$, follows immediately from the principle of rule induction. Similarly one obtains the following monotonicity property:

(*mon*) $\mathcal{L} \subseteq \mathcal{L}' \Rightarrow \vdash_{\mathcal{L}} \subseteq \vdash_{\mathcal{L}'}$.

Here the derivability relation is indexed; note that every elementary language defines its own derivability relation, and for the time being we are concerned with the comparison of these relations in various languages. Only with the completeness theorem will we see that the indices are superfluous, just as for the consequence relation $\models$. To prove (*mon*) let $\mathcal{E}(X, \alpha)$ be the property '$X \vdash_{\mathcal{L}'} \alpha$' for which the conditions (o) and (s) of rule induction are easily verified. For instance, let $X \vdash_{\mathcal{L}} \alpha, \beta$ and suppose $X \vdash_{\mathcal{L}'} \alpha, \beta$. Then ($\wedge$1), applied in $\mathcal{L}'$, yields $X \vdash_{\mathcal{L}'} \alpha \wedge \beta$ as well.

As in propositional logic we have here the easily provable

Finiteness theorem. *If $X \vdash \alpha$ then $X_0 \vdash \alpha$ for some finite $X_0 \subseteq X$.*

The only difference to the proof from **1.4** is that a few more rules have to be considered. Remember that L denotes the signature of $\mathcal{L}$, L_0 that of $\mathcal{L}_0$, etc. For the moment we require a somewhat stronger version of the theorem, namely

(*fin*) *If* $X \vdash_{\mathcal{L}} \alpha$ *then there exists a finite signature* $L_0 \subseteq L$ *and a finite subset* $X_0 \subseteq X$ *such that* $X_0 \vdash_{\mathcal{L}_0} \alpha$.

Herein the claim $X_0 \vdash_{\mathcal{L}_0} \alpha$, of course, includes $X_0 \cup \{\alpha\} \subseteq \mathcal{L}_0$. For the proof, consider the property 'there exist a finite $X_0 \subseteq X$ and $L_0 \subseteq L$ such that $X_0 \vdash_{\mathcal{L}_0} \alpha$'. It suffices to confirm the conditions (o) and (s) of the principle of rule induction. For $\alpha \in X \cup \{t \mathbin{=} t\}$ we clearly have $X_0 \vdash_{\mathcal{L}_0} \alpha$ where $X_0 = \{\alpha\}$ or $X_0 = \emptyset$. Thus, L_0 may be chosen to contain all the extralogical symbols occurring in α, and these are surely finitely many. This confirms (o). The induction step on (MR) is trivial. For $(\wedge 1)$ suppose $X_1 \vdash_{\mathcal{L}_1} \alpha$ and $X_2 \vdash_{\mathcal{L}_2} \alpha$ for some finite $X_i \subseteq X$ and $L_i \subseteq L$, $i = 1, 2$. Then (*mon*) gives $X_0 \vdash_{\mathcal{L}_0} \alpha_i$ where $X_0 = X_1 \cup X_2$ and $L_0 = L_1 \cup L_2$. Applying $(\wedge 1)$ to the language $\mathcal{L}_0$, we obtain $X_0 \vdash_{\mathcal{L}_0} \alpha_1 \wedge \alpha_2$, which is what we want. The induction steps for all remaining rules proceed similarly and are even somewhat simpler. This confirms condition (s), which in turn proves (*fin*).

In the foregoing proof, $\mathcal{L}_0$ contains at least the extralogical symbols of X_0 and α but perhaps also some others. Only with the completeness theorem can we know that the symbols occurring in X_0, α in fact suffice. This insensitivity of derivation with respect to language extensions can be derived purely proof-theoretically, albeit with considerable effort, but purely combinatorially and without recourse to the infinitistic means of semantics. A modest demonstration of such methods is the constant elimination by Lemmas 2.1 and 2.2 from the next section.

Now for some more examples of provable rules required later.

Example 1. (a) $\dfrac{X \vdash s \mathbin{=} t, s \mathbin{=} t'}{X \vdash t \mathbin{=} t'}$, (b) $\dfrac{X \vdash s \mathbin{=} t}{X \vdash t \mathbin{=} s}$, (c) $\dfrac{X \vdash t \mathbin{=} s, s \mathbin{=} t'}{X \vdash t \mathbin{=} t'}$.

To show (a) let $x \notin var\, t'$ and let α be the formula $x \mathbin{=} t'$. Then the premise of (a) is written $X \vdash s \mathbin{=} t, \alpha\frac{s}{x}$. Rule (=) yields $X \vdash \alpha\frac{t}{x}$. Now, $\alpha\frac{t}{x}$ equals $t \mathbin{=} t'$, since $x \notin var\, t'$, hence $X \vdash t \mathbin{=} t'$. (b) is obtained immediately from (a) with $t' = s$ because $X \vdash s \mathbin{=} s$. And with this follows (c), for thanks to (b), the premise of (c) now yields $X \vdash s \mathbin{=} t, s \mathbin{=} t'$ and hence, by (a), the conclusion of (c).

Example 2. In (a)–(d), n is as usual the arity of the symbols f and r. (a) and (c) are provable for $i = 1, \ldots, n$. In order to ease the writing, $X \vdash \vec{t} \mathbin{=} \vec{t'}$ abbreviates $X \vdash t_1 \mathbin{=} t'_1, \ldots, t_n \mathbin{=} t'_n$ so that, for instance, rule (b) has actually n premisses.

$$\text{(a)}\ \frac{X \vdash t_i \mathbin{=} t}{X \vdash f\vec{t} \mathbin{=} f t_1 \cdots t_{i-1} t t_{i+1} \cdots t_n}, \quad \text{(b)}\ \frac{X \vdash \vec{t} \mathbin{=} \vec{t'}}{X \vdash f\vec{t} \mathbin{=} f\vec{t'}},$$

$$\text{(c)}\ \frac{X \vdash t_i \mathbin{=} t, r\vec{t}}{X \vdash r t_1 \cdots t_{i-1} t t_{i+1} \cdots t_n}, \qquad \text{(d)}\ \frac{X \vdash \vec{t} \mathbin{=} \vec{t'}, r\vec{t}}{X \vdash r\vec{t'}}.$$

Proof of (a): Suppose $X \vdash s \mathbin{=} t$ with $s := t_i$. Let α be $f\vec{t} \mathbin{=} f t_1 \cdots t_{i-1} x t_{i+1} \cdots t_n$, where x is not to occur in any of the t_j. Since $X \vdash \alpha\frac{t_i}{x}$ $(= f\vec{t} \mathbin{=} f\vec{t})$, it follows that

$X \vdash \alpha\frac{t}{x}$ using (=). This confirms the conclusion of (a). (b) is then obtained by considering Example 1(c) and the n times iteration of (a), as can best be seen by first working through the case $n = 2$. Rule (c) is just another application of (=) by taking the formula $rt_1 \cdots t_{i-1} x t_{i+1} \cdots t_n$ for α where again, x is supposed not to occur in any of the t_j. Applying (c) n times then yields (d).

Example 3. (a) $\vdash \exists x\, t{=}x$, for all x, t with $x \notin var\, t$, (b) $\vdash \exists x\, x{=}x$.
(a) holds because (∀1) gives $\forall x\, t{\neq}x \vdash t{\neq}t$, for $t{\neq}t$ equals $(t{\neq}x)\frac{t}{x}$ (here $x \notin var\, t$ is required). Clearly, $\forall x\, t \neq x \vdash t{=}t$ as well. Thus, $\forall x\, t \neq x \vdash \exists x\, t{=}x$ by (¬1). Trivially, also $\neg\forall x\, t \neq x \vdash \exists x\, t{=}x$ $(= \neg\forall x\, t \neq x)$. Therefore, by (¬2), $\vdash \exists x\, t{=}x$. Similarly, (b) is verified, starting with $\forall x\, x \neq x \vdash x \neq x, x{=}x$. Note that the assumption $x \notin var\, t$ is essential in order to derive $\vdash \exists x\, t{=}x$ for a compound term t and hence to gain $\exists x\, t{=}x$ as a tautology. For instance, $\exists x\, fx{=}x$ with a unary operation symbol f is not a tautology, because this formula is falsified in the 2-element algebra $(\{0,1\}, f)$, with $f0 = 1$ and $f1 = 0$.

A set X $(\subseteq \mathcal{L})$ is called *inconsistent* if $X \vdash \alpha$ for all $\alpha \in \mathcal{L}$, and otherwise *consistent*, exactly as in propositional logic. A satisfiable set X is evidently consistent. By (¬1), the inconsistency of X is equivalent to $X \vdash \alpha, \neg\alpha$ for any α, hence also to $X \vdash \bot$ since $\bot = \neg\top$ and certainly $X \vdash \top$ $(= \exists \boldsymbol{v}_0\, \boldsymbol{v}_0{=}\boldsymbol{v}_0)$ by Example 3.

As in **1.4**, $\vdash$ is completely characterized by some inconsistency condition. Indeed, the proofs given there of the two properties

$$\mathtt{C}^+\!: \quad X \vdash \alpha \;\Leftrightarrow\; X, \neg\alpha \vdash \bot, \qquad \mathtt{C}^-\!: \quad X \vdash \neg\alpha \;\Leftrightarrow\; X, \alpha \vdash \bot$$

from Lemma 1.4.2 remain correct for any meaningful definition of $\bot$. $\mathtt{C}^+$ and $\mathtt{C}^-$ will permanently be used in the sequel without explicitly refering to them.

As in propositional logic, $X \subseteq \mathcal{L}$ is called *maximally consistent* if X is consistent but each proper extension of X in $\mathcal{L}$ is inconsistent. There are various characterizations of maximal consistency. For instance, the one given in Exercise 4 is easily confirmed by using one of the properties $\mathtt{C}^+$ or $\mathtt{C}^-$.

Exercises

1. Derive the rule $\dfrac{X \vdash \alpha\frac{t}{x}}{X \vdash \exists x\alpha}$ $(\alpha, \frac{t}{x}$ collision-free).
2. Prove $\forall x\alpha \vdash \forall y\alpha\frac{y}{x}$ and $\forall y\alpha\frac{y}{x} \vdash \forall x\alpha$ for $y \notin var\,\alpha$.
3. Using Exercise 2 and the cut rule prove $\dfrac{X \vdash \forall y\alpha\frac{y}{x}}{X \vdash \forall z\alpha\frac{z}{x}}$ $(y, z \notin var\,\alpha)$.
4. Show that a formula set X is maximally consistent if and only if for each $\varphi \in \mathcal{L}$ either $X \vdash \varphi$ or $X \vdash \neg\varphi$.

3.2 The Completeness Proof

Let $\mathcal{L}$ be a language and c a constant (more precisely, a constant symbol). $\mathcal{L}c$ is the result of adjoining c to $\mathcal{L}$. We have $\mathcal{L}c = \mathcal{L}$ if and only if c is already in $\mathcal{L}$. Similarly $\mathcal{L}C$ denotes the language resulting from $\mathcal{L}$ by adjoining a set C of constants, a *constant expansion* of $\mathcal{L}$. We shall also come across such expansions in Chapter **5**. Let $\alpha\frac{z}{c}$ (read "α z for c") denote the formula arising from α by replacing c with the variable z, and put $X\frac{z}{c} := \{\alpha\frac{z}{c} \mid \alpha \in X\}$. c then no longer occurs in $X\frac{z}{c}$. We actually require the following assertion only for a single variable z, but as is often the case, induction proves only a stronger version unproblematically.

Lemma 2.1 (on constant elimination). *Suppose $X \vdash_{\mathcal{L}c} \alpha$. Then $X\frac{z}{c} \vdash_{\mathcal{L}} \alpha\frac{z}{c}$ for almost all variables z.*

Proof by rule induction in $\vdash_{\mathcal{L}c}$. If $\alpha \in X$ then $\alpha\frac{z}{c} \in X\frac{z}{c}$ is clear; if α is of the form $t\mathbf{=}t$, so too is $\alpha\frac{z}{c}$. Thus, $X\frac{z}{c} \vdash_{\mathcal{L}} \alpha\frac{z}{c}$ in either case, even for all z. Only the induction steps on (∀1), (∀2) and (=) are not immediately apparent. We restrict ourselves to (∀1), because the steps for (∀2) and (=) proceed analogously. Let $X \vdash_{\mathcal{L}c} \forall x\alpha$ so that $X\frac{z}{c} \vdash_{\mathcal{L}} (\forall x\alpha)\frac{z}{c}$ for almost all z by the induction hypothesis. Suppose $\alpha, \frac{t}{x}$ are collision-free, and $z \notin var\{\forall x\alpha, t\}$. A separate induction on α readily confirms $\alpha\frac{t}{x}\frac{z}{c} = \alpha'\frac{t'}{x}$ with $\alpha' := \alpha\frac{z}{c}$ and $t' := t\frac{z}{c}$. Clearly $\alpha', \frac{t'}{x}$ are collision-free as well. Because by the induction hypothesis $X\frac{z}{c} \vdash_{\mathcal{L}} (\forall x\,\alpha)\frac{z}{c} = \forall x\alpha'$, rule (∀1) then yields $X\frac{z}{c} \vdash_{\mathcal{L}} \alpha'\frac{t'}{x} = \alpha\frac{t}{x}\frac{z}{c}$, and this holds still for almost all variables z. ❑

This lemma leads to the following derivable rule of "constant-quantification" whose semantical counterpart plays a key rule in model theory:

$$(\forall 3)\quad \frac{X \vdash \alpha\frac{c}{x}}{X \vdash \forall x\alpha} \qquad (c \text{ not in } X, \alpha).$$

Indeed, suppose $X \vdash \alpha\frac{c}{x}$. Because of the finiteness theorem we may assume that X is finite. By Lemma 2.1, where in the case at hand $\mathcal{L}c = \mathcal{L}$, some y not occurring in X, α can be found such that $X\frac{y}{c} \vdash \alpha\frac{c}{x}\frac{y}{c} = \alpha\frac{y}{x}$ (the latter holds because c does not occur in α). Since $X\frac{y}{c} = X$, we thus obtain $X \vdash \alpha\frac{y}{x}$. Hence $X \vdash \forall x\alpha$ by (∀2), which confirms (∀3). A likewise useful consequence of constant elimination is

Lemma 2.2. *Let C be any set of constants and $\mathcal{L}' = \mathcal{L}C$. Then $X \vdash_{\mathcal{L}} \alpha \Leftrightarrow X \vdash_{\mathcal{L}'} \alpha$, for all $X \subseteq \mathcal{L}$ and $\alpha \in \mathcal{L}$. Thus, $\vdash_{\mathcal{L}'}$ is a conservative expansion of $\vdash_{\mathcal{L}}$.*

Proof. (*mon*) states that $X \vdash_{\mathcal{L}} \alpha \Rightarrow X \vdash_{\mathcal{L}'} \alpha$. Suppose conversely $X \vdash_{\mathcal{L}'} \alpha$. To prove $X \vdash_{\mathcal{L}} \alpha$ we may assume, thanks to (*fin*) and (MR), that C is finite. Since the adjunction of finitely many constants can be undertaken stepwise, we may suppose for the purpose of the proof that $\mathcal{L}' = \mathcal{L}c$ for a single constant c not occurring in $\mathcal{L}$. Lemma 2.1 then yields $X\frac{z}{c} \vdash_{\mathcal{L}} \alpha\frac{z}{c}$ for at least one variable z. $X\frac{z}{c} \vdash_{\mathcal{L}} \alpha\frac{z}{c}$ means the same as $X \vdash_{\mathcal{L}} \alpha$ because c occurs neither in X nor in α. ❑

In the following, we denote the derivability relation in $\mathcal{L}$ and in every constant expansion $\mathcal{L}'$ of $\mathcal{L}$ with the same symbol $\vdash$. By Lemma 2.2 no misunderstandings can arise from this notation. Since the consistency of X is equivalent to $X \nvdash \bot$, there is also no need to distinguish between the consistency of $X \subseteq \mathcal{L}$ with respect to $\mathcal{L}$ or $\mathcal{L}'$. This is highly significant for the proofs of the next two Lemmas.

The proof of the completeness theorem essentially proceeds with a model construction from the syntactic material of a certain constant expansion of $\mathcal{L}$. We first choose for each variable x and each $\alpha \in \mathcal{L}$ a constant $c_{x,\alpha}$ not occurring in $\mathcal{L}$; more precisely, we choose exactly one such constant for each pair x, α. Define

$$(*) \quad \alpha^x := \neg\forall x\alpha \wedge \alpha\tfrac{c}{x} \quad (c := c_{x,\alpha}).$$

Here it is insignificant how many free variables α contains, and whether x occurs at all in α. We mention that the formula $\neg\alpha^x$ is logically equivalent to $\exists x\neg\alpha \to \neg\alpha\tfrac{c}{x}$. This formula states that under the hypothesis $\exists x\neg\alpha$, the constant c represents a *counterexample* for the validity of α, that is, an example for the validity of $\neg\alpha$.

Lemma 2.3. *Let $\Gamma_{\mathcal{L}} := \{\neg\alpha^x \mid \alpha \in \mathcal{L}, x \in Var\}$ where α^x is defined as in $(*)$, and let $X \subseteq \mathcal{L}$ be consistent. Then $X \cup \Gamma_{\mathcal{L}}$ is consistent as well.*

Proof. Assume that $X \cup \Gamma_{\mathcal{L}} \vdash \bot$. Since $X \nvdash \bot$, there is some $n \geqslant 0$ and formulas $\neg\alpha_0^{x_0}, \ldots, \neg\alpha_n^{x_n} \in \Gamma_{\mathcal{L}}$ such that (a): $X \cup \{\neg\alpha_i^{x_i} \mid i \leqslant n\} \vdash \bot$. Choose n to be minimal so that (b): $X' := X \cup \{\neg\alpha_i^{x_i} \mid i < n\} \nvdash \bot$, and set $x := x_n$, $\alpha := \alpha_n$, and $c := c_{x,\alpha}$. By (a), $X' \cup \{\neg\alpha^x\} \vdash \bot$. Hence, $X' \vdash \alpha^x$, and so $X' \vdash \neg\forall x\alpha, \alpha\tfrac{c}{x}$, by $(\wedge 2)$. But $X' \vdash \alpha\tfrac{c}{x}$ yields $X' \vdash \forall x\alpha$ using $(\forall 3)$, since c does not occur in X' and α. Thus, $X' \vdash \forall x\alpha, \neg\forall x\alpha$, whence $X' \vdash \bot$, contradicting (b) and hence our assumption. ❏

Call $X \subseteq \mathcal{L}$ a *Henkin set* if X satisfies the following two conditions:

(H1) $X \vdash \neg\alpha \;\Leftrightarrow\; X \nvdash \alpha$, (equivalently, $X \vdash \alpha \;\Leftrightarrow\; X \nvdash \neg\alpha$),

(H2) $X \vdash \forall x\alpha \;\Leftrightarrow\; X \vdash \alpha\tfrac{c}{x}$ for all constants c in $\mathcal{L}$.

(H1) and (H2) produce yet another useful property of a Henkin set X, namely

(H3) For each term t there exists a constant c such that $X \vdash t = c$.

Indeed, $X \vdash \exists x t = x$ $(= \neg\forall x\, t \neq x)$ for $x \notin var\, t$ by Example 3 in **3.1**. Hence, $X \nvdash \forall x\, t \neq x$ by (H1). Thus $X \nvdash t \neq c$ for some c by (H2), and so $X \vdash t = c$ by (H1).

As regards the following lemma, we mention that in the framework of the original language $\mathcal{L}$, consistent sets are not generally embeddable in Henkin sets.

Lemma 2.4. *Let $X \subseteq \mathcal{L}$ be consistent. Then there exists a Henkin set $Y \supseteq X$ in a suitable constant expansion $\mathcal{L}C$ of $\mathcal{L}$.*

Proof. Put $\mathcal{L}_0 := \mathcal{L}$, $X_0 := X$ and assume $\mathcal{L}_n$, X_n have been given. Let $\mathcal{L}_{n+1}$ result from $\mathcal{L}_n$ by adopting new constants $c_{x,\alpha,n}$ for all $x \in Var$, $\alpha \in \mathcal{L}_n$; more precisely $\mathcal{L}_{n+1} = \mathcal{L}_n C_n$, with the set C_n of constants $c_{x,\alpha,n}$. Further let $X_{n+1} = X_n \cup \Gamma_{\mathcal{L}_n}$.

Here $\Gamma_{\mathcal{L}_n}$ is defined as in Lemma 2.3, so that $X_{n+1} \subseteq \mathcal{L}_{n+1}$. Using Lemma 2.3 we have $X_n \nvdash \bot$ for each n. Let $X' := \bigcup_{n\in\mathbb{N}} X_n$, hence $X' \subseteq \mathcal{L}' := \bigcup_{n\in\mathbb{N}} \mathcal{L}_n = \mathcal{L}C$, where $C := \bigcup_{n\in\mathbb{N}} C_n$. Then $X' \nvdash \bot$ since X', as the union of a chain of consistent sets, is surely consistent (in $\mathcal{L}'$). Let $\alpha \in \mathcal{L}'$, $x \in \mathit{Var}$, and, say, $\alpha \in \mathcal{L}^n$ with minimal n, and let α^x be the formula defined as in $(*)$ but with respect to $\mathcal{L}^n$. Then $\neg\alpha^x$ belongs to X_{n+1}. Hence $\neg\alpha^x \in X'$. Now let $(H, \subseteq)$ be the partial order of all consistent extensions of X' in $\mathcal{L}'$. Every chain $K \subseteq H$ has the upper bound $\bigcup K$ in H, because if all members of K are consistent so is $\bigcup K$. Also $H \neq \emptyset$; for instance $X' \in H$. By Zorn's lemma, H therefore contains a maximal element Y. In short, Y is a maximally consistent set containing X'. Further, what is significant here, Y is at the same time a Henkin set. Here is the proof:
(H1) $\Rightarrow$: $Y \vdash \neg\alpha$ implies $Y \nvdash \alpha$ due to the consistency of Y. $\Leftarrow$: If $Y \nvdash \alpha$ then surely $\alpha \notin Y$. As a result, $Y, \alpha \vdash \bot$, for Y is maximally consistent. Thus $Y \vdash \neg\alpha$.
(H2) $\Rightarrow$: Clear by $(\forall 1)$. $\Leftarrow$: Let $Y \vdash \alpha\frac{c}{x}$ for all c in $\mathcal{L}'$, so also $Y \vdash \alpha\frac{c}{x}$ for $c := c_{x,\alpha,n}$, where n is minimal with $\alpha \in \mathcal{L}_n$. Assume that $Y \nvdash \forall x\alpha$. Then $Y \vdash \neg\forall x\alpha$ by (H1). But $Y \vdash \neg\forall x\alpha, \alpha\frac{c}{x}$ implies $Y \vdash \neg\forall x\alpha \wedge \alpha\frac{c}{x} = \alpha^x$ using $(\wedge 1)$. Now, since Y is consistent, $Y \vdash \alpha^x$ contradicts $Y \vdash \neg\alpha^x$. The latter is certainly the case because $\neg\alpha^x \in X' \subseteq Y$. Thus, our assumption was wrong and indeed $Y \vdash \forall x\alpha$. ❏

Lemma 2.5. *Every Henkin set $Y \subseteq \mathcal{L}$ possesses a model.*

Proof. The model constructed in the following is called a *term model*. Let $t \approx t'$ whenever $Y \vdash t = t'$. The relation $\approx$ is a congruence in the term algebra $\mathcal{T}$ of $\mathcal{L}$. This means (repeating the definitions on page 41),

(a) $\approx$ is an equivalence relation,

(b) $t_1 \approx t_1', \ldots, t_n \approx t_n' \;\Rightarrow\; f\vec{t} \approx f\vec{t'}$, for operation symbols f in $\mathcal{L}$.

The claim (a) follows immediately from $Y \vdash t = t$ and Example 1 in **3.1**; (b) is just another way of formulating Example 2(b). Let $A := \{\bar{t} \mid t \in \mathcal{T}\}$. Here $\bar{t}$ denotes the equivalence class of $\approx$ to which the term t belongs, so that

(c) $\bar{t} = \bar{s} \;\Leftrightarrow\; t \approx s \;\Leftrightarrow\; Y \vdash t = s$.

This set A is the domain of the sought model $\mathcal{M} = (\mathcal{A}, w)$ for Y. The factorization of $\mathcal{T}$ will ensure that $=$ means identity in the model. Let C be the set of constants in $\mathcal{L}$. By (H3) there is for each term t in $\mathcal{T}$ some $c \in C$ such that $c \approx t$. Therefore even $A = \{\bar{c} \mid c \in C\}$. Now, let $x^{\mathcal{M}} := \overline{x}$ and $c^{\mathcal{M}} := \bar{c}$ for variables and constants in $\mathcal{L}$. An operation symbol f occurring in $\mathcal{L}$ of arity n is interpreted by $f^{\mathcal{M}}$ where

$$f^{\mathcal{M}}(\bar{t}_1, \ldots, \bar{t}_n) := \overline{ft_1 \cdots t_n}.$$

This definition is sound because $\approx$ is a congruence in the term algebra $\mathcal{T}$. Finally, define $r^{\mathcal{M}}$ for an n-ary relation symbol r by

$$r^{\mathcal{M}}\bar{t}_1 \cdots \bar{t}_n \;\Leftrightarrow\; Y \vdash r\vec{t}.$$

This definition is also sound, since $Y \vdash r\vec{t} \;\Rightarrow\; Y \vdash r\vec{t'}$ whenever $t_1 \approx t'_1, \ldots, t_n \approx t'_n$. Here we use Example 2(d) in **3.1**. Induction then yields

$$\text{(d) } t^{\mathcal{M}} = \bar{t}; \quad \text{(e) } \mathcal{M} \vDash \alpha \;\Leftrightarrow\; Y \vdash \alpha,$$

of which (e) may be regarded as the goal of the constructions. (d) is evident for prime terms, and the induction hypothesis $t_i^{\mathcal{M}} = \bar{t}_i$ for $i = 1, \ldots, n$ leads to

$$(f\vec{t})^{\mathcal{M}} = f^{\mathcal{M}}(t_1^{\mathcal{M}}, \ldots, t_n^{\mathcal{M}}) = f^{\mathcal{M}}(\bar{t}_1, \ldots, \bar{t}_n) = \overline{f\vec{t}}.$$

(e) follows by induction on $\operatorname{rk}\alpha$. We begin with formulas of rank 0 (prime formulas). Induction proceeds under consideration of $\operatorname{rk}\alpha < \operatorname{rk}\neg\alpha$, $\operatorname{rk}\alpha, \operatorname{rk}\beta < \operatorname{rk}(\alpha \wedge \beta)$ and $\operatorname{rk}\alpha\,\frac{c}{x} < \operatorname{rk}\forall x\alpha$, analogously to formula induction:

$$\begin{array}{lcll}
\mathcal{M} \vDash t = s & \Leftrightarrow & t^{\mathcal{M}} = s^{\mathcal{M}} & \Leftrightarrow \;\; \bar{t} = \bar{s} \qquad \text{(by (d))} \\
 & & & \Leftrightarrow \;\; Y \vdash t = s \qquad \text{(by (c))}. \\
\mathcal{M} \vDash r\vec{t} & \Leftrightarrow & r^{\mathcal{M}} t_1^{\mathcal{M}} \cdots t_n^{\mathcal{M}} & \Leftrightarrow \;\; r^{\mathcal{M}} \bar{t}_1 \cdots \bar{t}_n \;\; \Leftrightarrow \;\; Y \vdash r\vec{t}. \\
\mathcal{M} \vDash \alpha \wedge \beta & \Leftrightarrow & \mathcal{M} \vDash \alpha, \beta & \Leftrightarrow \;\; Y \vdash \alpha, \beta \qquad \text{(induction hypothesis)} \\
 & & & \Leftrightarrow \;\; Y \vdash \alpha \wedge \beta \qquad \text{(using } (\wedge 1), (\wedge 2)\text{)}. \\
\mathcal{M} \vDash \neg\alpha & \Leftrightarrow & \mathcal{M} \nvDash \alpha & \Leftrightarrow \;\; Y \nvdash \alpha \qquad \text{(induction hypothesis)} \\
 & & & \Leftrightarrow \;\; Y \vdash \neg\alpha \qquad \text{(using (H1))}.
\end{array}$$

$$\begin{array}{lcll}
\mathcal{M} \vDash \forall x\alpha & \Leftrightarrow & \mathcal{M}_x^{\bar{c}} \vDash \alpha \text{ for all } c \in C & \text{(because } A = \{\bar{c} \mid c \in C\}\text{)} \\
 & \Leftrightarrow & \mathcal{M}_x^{c^{\mathcal{M}}} \vDash \alpha \text{ for all } c \in C & \text{(because } c^{\mathcal{M}} = \bar{c}\text{)} \\
 & \Leftrightarrow & \mathcal{M} \vDash \alpha\,\frac{c}{x} \text{ for all } c \in C & \text{(substitution theorem)} \\
 & \Leftrightarrow & Y \vdash \alpha\,\frac{c}{x} \text{ for all } c \in C & \text{(induction hypothesis)} \\
 & \Leftrightarrow & Y \vdash \forall x\alpha & \text{(using (H2))}.
\end{array}$$

Because of $Y \vdash \alpha$ for all $\alpha \in Y$, (e) immediately implies $\mathcal{M} \vDash Y$. ❏

Just as for propositional logic, the equivalence of consistency and satisfiability, and the completeness of $\vdash$, result from the above. These results, stated in the next two theorems, are what we aimed at in this section. Information about the size of the model constructed in the next theorem will be given in Theorem 4.1.

Theorem 2.6 (Model existence theorem). *Let $X \subseteq \mathcal{L}$ be consistent. Then X has a model.*

Proof. Let $Y \supseteq X$ be a Henkin extension of X, i.e., a Henkin set in a suitable constant expansion $\mathcal{L}C$ applying Lemma 2.4. According to Lemma 2.5, Y and hence also X has a model $\mathcal{M}'$ in $\mathcal{L}C$. Let $\mathcal{M}$ denote the $\mathcal{L}$-reduct of $\mathcal{M}'$. In other words, "forget" the interpretation of the constants not occurring in $\mathcal{L}$. Then, by Theorem 2.3.1, $\mathcal{M} \vDash X$ holds as well. ❏

Theorem 2.7 (Completeness theorem). *Let $\mathcal{L}$ be any first-order language. Then for all $X \subseteq \mathcal{L}$ and $\alpha \in \mathcal{L}$ holds $X \vdash \alpha \;\Leftrightarrow\; X \vDash \alpha$.*

Proof. The soundness of $\vdash$ states that $X \vdash \alpha \Rightarrow X \vDash \alpha$. The converse follows indirectly. Let $X \nvdash \alpha$, so that $X, \neg\alpha$ is consistent. Theorem 2.6 then provides model for $X \cup \{\neg\alpha\}$, whence $X \nvDash \alpha$. ❑

Thus, $\vDash$ and $\vdash$ can henceforth be freely interchanged. We will often verify $X \vdash \alpha$ by proving that $X \vDash \alpha$. In particular, for theories T, $T \vDash \alpha$ is equivalent to $T \vdash \alpha$, for which in the following we mostly write $\vdash_T \alpha$. Clearly, $\vdash_T \alpha$ means the same as $\alpha \in T$ for sentences α. More generally, let $X \vdash_T \alpha$ stand for $X \cup T \vdash \alpha$ and $\alpha \vdash_T \beta$ for $\{\alpha\} \vdash_T \beta$. We will also occasionally abbreviate $\alpha \vdash_T \beta \,\&\, \beta \vdash_T \gamma$ to $\alpha \vdash_T \beta \vdash_T \gamma$. In subsequent chapters, equivalences such as $\alpha \vdash_T \beta \Leftrightarrow \vdash_T \alpha \to \beta \Leftrightarrow \vdash_{T+\alpha} \beta$, and $\vdash_T \alpha \Leftrightarrow \vdash_T \alpha^{\mathfrak{G}}$, will be used without further mentioning and should be committed to memory. Some more useful equivalences are listed in Exercise 5.

Remark. The methods in this section easily provide also completeness of a logical calculus for *identity-free* (or $\boldsymbol{=}$-free) languages in which the symbol $\boldsymbol{=}$ does not appear. Simply discard from the calculus in **3.1** everything that refers to $\boldsymbol{=}$, including rule (=). Almost everything runs as before. The factorization in Lemma 2.5 is now dispensable and the domain A is the set of all terms of $\mathcal{L}C$. The last induction step in Lemma 2.5 has to be modified. We will not go into details since we will need in Chapter **4** only Exercise 2. The restriction to $\forall$-formulas therein is not really essential, because by Exercise 5 in **2.6** any X can be replaced by a satisfiably equivalent set of $\forall$-formulas after expanding the language by suitable Skolem functions.

Exercises

1. Show that a set $X \subseteq \mathcal{L}$ is maximally consistent iff there is a model $\mathcal{M}$ such that $X \vdash \alpha \Leftrightarrow \mathcal{M} \vDash \alpha$, for all $\alpha \in \mathcal{L}$.

2. Let $X \subseteq \mathcal{L}$ be a consistent set of identity-free $\forall$-formulas. Construct a model $\mathfrak{T} \vDash X$ on the domain $\mathcal{T}$ of all $\mathcal{L}$-terms by setting $r^{\mathfrak{T}}\vec{t} :\Leftrightarrow r^{\mathcal{M}}\vec{t}$, $c^{\mathfrak{T}} := c$, $f^{\mathfrak{T}}\vec{t} := f\vec{t}$, and $x^{\mathfrak{T}} = x$. Show in addition that if $X \subseteq \mathcal{L}^0$ and $\mathcal{L}$ contains at least one constant, then X has a model on the domain of all ground terms.

3. Let $K \neq \emptyset$ be a chain of theories in $\mathcal{L}$, i.e., $T \subseteq T'$ or $T' \subseteq T$, for all $T, T' \in K$. Show that $\bigcup K$ is a theory that is consistent iff all $T \in K$ are consistent.

4. Suppose T is consistent and $Y \subseteq \mathcal{L}$. Prove the equivalence of

   (i) $Y \vdash_T \bot$, (ii) $\vdash_T \neg\alpha$ for some conjunction α of formulas in Y.

5. Let $x \notin \mathit{var}\, t$ and $\alpha, \frac{t}{x}$ collision-free. Verify the equivalence of

   (i) $\vdash_T \alpha\,\frac{t}{x}$, (ii) $x \boldsymbol{=} t \vdash_T \alpha$, (iii) $\vdash_T \forall x(x \boldsymbol{=} t \to \alpha)$, (iv) $\vdash_T \exists x(x \boldsymbol{=} t \wedge \alpha)$.

3.3 First Applications–Nonstandard Models

In this section we draw important conclusions from the completeness theorem and the corresponding model-construction procedure. Since the finiteness theorem holds for the provability relation $\vdash$, Theorem 2.7 immediately yields

Theorem 3.1 (Finiteness theorem for the consequence relation). *$X \models \alpha$ implies $X_0 \models \alpha$ for some finite subset $X_0 \subseteq X$.*

Let us consider a first application. The elementary theory of fields of characteristic 0 is obviously axiomatized by the set X consisting of the axioms for fields and the formulas $\neg \mathtt{char}_p$ (page 39). We claim

(1) *A sentence α valid in all fields of characteristic 0 is also valid in all fields of sufficiently high prime characteristic p which, of course, depends on α.*

Indeed, since $X \models \alpha$, for some finite subset $X_0 \subseteq X$ we have $X_0 \models \alpha$. If p is a prime number larger than all prime numbers q such that $\neg \mathit{char}_q \in X_0$, then α holds in all fields of characteristic p, since these satisfy X_0. Thus (1) holds. From (1) we obtain, for instance, the information, easily formalized in $\mathcal{L}\{0, 1, +, \cdot\}$, that two given polynomials that are coprime over all fields of characteristic 0 are also coprime over fields of sufficiently high prime characteristic.

A noteworthy consequence of Theorem 3.1 is also the nonfinite axiomatizability of many elementary theories. Before presenting examples, we clarify finite axiomatizability in a somewhat broader context.

A set Z of strings of a given alphabet A is called *decidable* if there is an algorithm (a mechanical decision procedure) that after finitely many calculation steps provides us with an answer to the question whether a string ξ of symbols of A belongs to Z; otherwise Z is called *undecidable*. Thus it is certainly decidable whether ξ is a formula. While this is all intuitively plausible, it nonetheless requires more precision (undertaken in **6.2**). A theory T is called *recursively axiomatizable*, or just *axiomatizable*, if it possesses a decidable axiom system. This is the case, for instance, if T is *finitely axiomatizable*, i.e., if it has a finite axiom system.

From (1) it follows straight away that the theory of fields of characteristic 0 is not finitely axiomatizable. For were F a finite set of axioms, their conjunction $\alpha = \bigwedge F$ would, by (1), also have a field of finite characteristic as a model.

Now for another instructive example. An abelian group $\mathcal{G}$ is called *n-divisible* if $\mathcal{G} \models \vartheta_n$ with $\vartheta_n := \forall x \exists y\, x = ny$ where ny is the n-fold sum $y + \cdots + y$, and $\mathcal{G}$ is called *divisible* if $\mathcal{G} \models \vartheta_n$ for all $n \geqslant 1$. Thus, the theory of divisible abelian groups, DAG, is axiomatized by the set X consisting of the axioms for abelian groups plus all sentences ϑ_n. Also DAG is not finitely axiomatizable. This follows as above from

(2) *Every sentence* $\alpha \in \mathcal{L}\{+,0\}$ *valid in all divisible abelian groups is also valid in at least one nondivisible abelian group.*

To prove (2), let $\alpha \in \mathsf{DAG}$, or equivalently $X \vDash \alpha$. According to Theorem 3.1 we have $X_0 \vDash \alpha$ for some finite $X_0 \subseteq X$. Let $\mathbb{Z}_p$ be the cyclic group of order p, where p is a prime number $> n$ for all n with $\vartheta_n \in X_0$. The mapping $x \mapsto nx$ from $\mathbb{Z}_p$ to itself is surjective for $0 < n < p$, otherwise $\{na \mid a \in \mathbb{Z}_p\}$ would be a nontrivial subgroup of $\mathbb{Z}_p$. Hence, $\mathbb{Z}_p \vDash \vartheta_n$ for all $n < p$. Thus, $\mathbb{Z}_p \vDash X_0$ and so $\mathbb{Z}_p \vDash \alpha$. On the other hand, $\mathbb{Z}_p$ is not p-divisible because $px = 0$ for all $x \in \mathbb{Z}_p$. In exactly the same way, we can show that the theory of *torsion-free* abelian groups is not finitely axiomatizable. In these groups is $na \neq 0$ for all $n \neq 0$ and $a \neq 0$.

In a similar manner, it is possible to prove for many theories that they are not finitely axiomatizable. However, this may often demand more involved methods than the above ones. For instance, consider the theory of a.c. fields (see page 38), denoted by ACF, which results from adjoining to the theory of fields the schema of all sentences $\forall \vec{a}\, \exists x\, p(\vec{a}, x) = 0$, where $p(\vec{a}, x)$ denotes the term

$$x^{n+1} + a_n x^n + \cdots + a_1 x + a_0 \qquad (n = 0, 1, \dots),$$

called a *monic polynomial of degree* $n+1$. Here let $a_0, \dots, a_n, x$ denote distinct variables. Thus, every monic polynomial has a zero, and so *every* polynomial of positive degree. Nonfinite axiomatizability of ACF follows from the by no means trivial existence proof of fields in which all polynomials up to a certain degree do factorize but irreducible polynomials of higher degree still exist. The same holds for the theory ACF_p of a.c. fields of fixed characteristic p ($p = 0$ or a prime number).

As in propositional logic, the finiteness theorem for the consequence relation leads immediately to the corresponding compactness result:

Theorem 3.2 (Compactness theorem). *Any set* X *of first-order formulas is satisfiable provided every finite subset of* X *is satisfiable.*

Because of the greater power of expression of first-order languages, this theorem is somewhat more amenable to certain applications than its propositional counterpart. It can be proved in various ways, even quite independent of a logical calculus; for instance, by means of ultraproducts as will be carried out in **5.7**. It can also be reduced to the propositional compactness theorem, for X is satisfiably equivalent to a set of propositional formulas; see Remark 1 in **4.1**. For applications of Theorem 3.2 we concentrate on the construction of nonstandard models; to this end we introduce some more important concepts.

A theory T ($\subseteq \mathcal{L}^0$) is called *complete* if it is consistent and has no consistent proper extension in the same language. It is easily seen that this property is equivalent to either $\vdash_T \alpha$ or $\vdash_T \neg\alpha$ but not both, for each $\alpha \in \mathcal{L}^0$ (for other equivalences, see Theorem 5.2.1). Hence, for an arbitrary $\mathcal{A}$, the theory $Th\,\mathcal{A}$ is always complete.

We will frequently come across the theory $Th\mathcal{N}$ where $\mathcal{N} = (\mathbb{N}, 0, \mathtt{S}, +, \cdot)$ with the *successor function* $\mathtt{S}\colon n \mapsto n+1$. The choice of signature is a matter of convenience; for instance, one could replace $\mathtt{S}$ by the constant 1. Of the relations and functions definable in $\mathcal{N}$, we name just $\leqslant$, defined by $x \leqslant y \leftrightarrow \exists z\, z+x = y$, and the *predecessor function* $\mathtt{Pd}\colon \mathbb{N} \to \mathbb{N}$, defined by $y = \mathtt{Pd}\, x \leftrightarrow y = 0 \vee x = \mathtt{S}y$, so that $\mathtt{Pd}\, 0 = 0$.

Certain axiomatic subtheories of $Th\mathcal{N}$ are even more frequently dealt with, in particular *Peano arithmetic* PA in the arithmetical language $\mathcal{L}_{ar} := \mathcal{L}\{0, \mathtt{S}, +, \cdot\}$. This theory is important for many investigations in mathematical foundations and theoretical computer science (see e.g. [Kr]). The axioms of PA run as follows:

$$\begin{array}{lll} \forall x\, \mathtt{S}x \neq 0, & \forall x\, x + 0 = x, & \forall x\, x \cdot 0 = 0, \\ \forall x \forall y (\mathtt{S}x = \mathtt{S}y \to x = y), & \forall x \forall y\, x + \mathtt{S}y = \mathtt{S}(x+y), & \forall x \forall y\, x \cdot \mathtt{S}y = x \cdot y + x, \end{array}$$

$$\text{IS:} \quad \varphi\tfrac{0}{x} \wedge \forall x(\varphi \to \varphi\tfrac{\mathtt{S}x}{x}) \to \forall x \varphi.$$

IS is called the *induction schema* and should not be mixed up with the induction axiom IA discussed on the next page. In IS, φ is any formula in $\mathcal{L}_{ar}$ with $x \in \mathit{free}\,\varphi$. IS reads more precisely $[\varphi\tfrac{0}{x} \wedge \forall x(\varphi \to \varphi\tfrac{\mathtt{S}x}{x}) \to \forall x \varphi]^{\mathtt{G}}$, see our convention in **2.5**. Thus, to prove $\vdash_{\mathsf{PA}} \forall x \varphi$, one has to confirm $\vdash_{\mathsf{PA}} \varphi\tfrac{0}{x}$ (*induction initiation*), and $\vdash_{\mathsf{PA}} \forall x(\varphi \to \varphi\tfrac{\mathtt{S}x}{x})$ or equivalently, $\varphi \vdash_{\mathsf{PA}} \varphi\tfrac{\mathtt{S}x}{x}$ (*induction step*, the derivation of the *induction claim* $\varphi\tfrac{\mathtt{S}x}{x}$ from the *induction hypothesis* φ).

Example. Let $\varphi = \varphi(x) := x \neq 0 \to \exists v\, \mathtt{S}v = x$. We want to prove $\vdash_{\mathsf{PA}} \forall x \varphi(x)$. In words, each $x \neq 0$ has a predecessor, not something seen at once from the axioms. Trivially, $\vdash_{\mathsf{PA}} \varphi\tfrac{0}{x}$. Since $\mathtt{S}v = x \vdash_{\mathsf{PA}} \mathtt{SS}v = \mathtt{S}x$, we get $\exists v\, \mathtt{S}v = x \vdash_{\mathsf{PA}} \exists v \mathtt{S}v = \mathtt{S}x$ by particularization. Therefore $x \neq 0 \to \exists v \mathtt{S}v = x \vdash_{\mathsf{PA}} x \neq 0 \to \exists v \mathtt{S}v = \mathtt{S}x$ (cf. Exercise 2 in **1.3**), that is, $\varphi \vdash_{\mathsf{PA}} \varphi\tfrac{\mathtt{S}x}{x}$ (the induction step), and so $\vdash_{\mathsf{PA}} \forall x \varphi$ by IS. This proof is easily supplemented by an inductive proof of $\vdash_{\mathsf{PA}} \forall x\, \mathtt{S}x \neq x$.

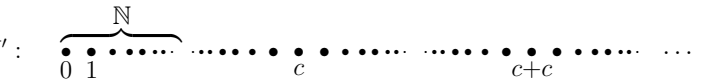
Remark 1. Only few arithmetical facts (like $x \leqslant y \leftrightarrow \mathtt{S}x \leqslant \mathtt{S}y$) are derivable in PA without IS. Already the derivation of $x \leqslant x$ needs IS when $x \leqslant y$ is defined as above by $\exists z\, z + x = y$. More in the exercises; these are exclusively devoted to PA, in order to get familiar in time with this important theory. In **7.1** it will then become clear that PA fully embraces elementary number theory and practically the whole of discrete mathematics. It is not of any import that subtraction is only partially defined in models of PA. A theory of integers formulated similarly to PA may be more convenient for number theory, but is actually not stronger than PA; it is interpretable in PA in the sense of **6.6**. We mention that PA is not finitely axiomatizable, shown for the first time in [Ry].

We will now prove that not only PA but also the complete theory $Th\mathcal{N}$ has alongside the standard model $\mathcal{N}$ other models not isomorphic to $\mathcal{N}$, called *nonstandard models*. In these models, exactly the same theorems hold as in $\mathcal{N}$. The existence proof of a nonstandard model $\mathcal{N}'$ of $Th\mathcal{N}$ is strikingly simple. Let $x \in \mathit{Var}$ and $X := Th\mathcal{N} \cup \{\underline{n} < x \mid n \in \mathbb{N}\}$. Here and elsewhere we use $\underline{n}$ to denote the term $\mathtt{S}^n 0 := \underbrace{\mathtt{S} \cdots \mathtt{S}}_{n} 0$. Thus $\underline{1} = \mathtt{S}0$, $\underline{2} = \mathtt{S}\underline{1}, \ldots$ Instead of $\underline{0}$ $(= \mathtt{S}^0 0)$ one writes just 0.

$\underline{n} < x$ is the formula $\underline{n} \leqslant x \wedge \underline{n} \neq x$. One may x replace here by a constant symbol c, thus expanding the language. But both approaches lead to the same result.

Every finite subset $X_0 \subseteq X$ possesses a model. Indeed, there is evidently some m such that $X_0 \subseteq X_1 := Th\mathcal{N} \cup \{\underline{n} < x \mid n < m\}$, and X_1 certainly has a model: one need only appoint to x in $\mathcal{N}$ the number m. Thus, by Theorem 3.2, X has a model $(\mathcal{N}', c)$ with the domain $\mathbb{N}'$, where $c \in \mathbb{N}'$ denotes the interpretation of x. Because $\mathcal{N}'$ satisfies all sentences valid in $\mathcal{N}$, including in particular the sentences $\mathsf{S}\underline{n} = \underline{\mathsf{S}n}$, $\underline{n+m} = \underline{n} + \underline{m}$ and $\underline{n \cdot m} = \underline{n} \cdot \underline{m}$, it is easily seen that $n \mapsto \underline{n}^{\mathcal{N}'}$ constitutes an embedding from $\mathcal{N}$ into $\mathcal{N}'$ whose image can be thought of as coinciding with $\mathcal{N}$.[2] Thus, it is legitimate to presume that $\underline{n}^{\mathcal{N}'} = n$ and hence $\mathcal{N} \subseteq \mathcal{N}'$.

Because $\mathcal{N}' \vDash X$, on the one hand $\mathcal{N}'$ is elementarily equivalent to $\mathcal{N}$, and on the other $n < a$ for all n and any $a \in \mathbb{N}' \backslash \mathbb{N}$, since in $\mathcal{N}$ and hence in $\mathcal{N}'$ holds $(\forall x \leqslant \underline{n}) \bigvee_{i \leqslant n} x = \underline{i}$. In short, $\mathbb{N}$ is a (proper) initial segment of $\mathbb{N}'$, or $\mathcal{N}'$ is an *end extension* of $\mathcal{N}$. The elements of $\mathbb{N}' \backslash \mathbb{N}$ are called *nonstandard numbers*. Alongside c, other examples are $c + c$ and $c + \underline{n}$ for $n \in \mathbb{N}$. Clearly, c has both an immediate successor and an immediate predecessor in the order, because $\mathcal{N}' \vDash (\forall x \neq 0) \exists y\, x = \mathsf{S}y$. The figur gives a rough picture of a nonstandard model $\mathcal{N}'$:



$\mathcal{N}'$ has the same number-theoretical features as $\mathcal{N}$, at least all those that can be formulated in $\mathcal{L}_{ar}$. These include nearly all the interesting ones, as will turn out to be the case in **7.1**. For example, $\forall x \exists y (x = \underline{2}y \vee x = \underline{2}y + \underline{1})$ holds in every model of $Th\mathcal{N}$, that is, every nonstandard number is either even or odd. Clearly, $\mathbb{N}'$ contains gaps in the sense of **2.1**, $(\mathbb{N}, \mathbb{N}' \backslash \mathbb{N})$ being an example.

Remark 2. Theorem 4.1 will show that $Th\mathcal{N}$ has countable nonstandard models. The order of such a model $\mathcal{N}'$ is easy to make intuitive: it arises from the half-open interval $[0, 1)$ of rational numbers by replacing 0 with $\mathbb{N}$ and every other $r \in [0, 1)$ by a specimen from $\mathbb{Z}$. On the other hand, neither $+^{\mathcal{N}'}$ nor $\cdot^{\mathcal{N}'}$ is effectively describable; see e.g. [HP].

Replacing IS in the axiom system for PA by the so-called *induction axiom*

$$\text{IA:} \quad \forall P(P0 \wedge \forall x(Px \to P\mathsf{S}x) \to \forall x Px) \qquad (P \text{ a predicate variable})$$

results in a *categorical* axiom system that, up to isomorphism, has just a single model (see e.g. [Ra2]). How is it possible that $\mathcal{N}$ is uniquely determined up to isomorphism by a few axioms, but at the same time nonstandard models exist for $Th\mathcal{N}$? The answer: IA cannot be adequately formulated in $\mathcal{L}_{ar}$. That is, IA is not an axiom or perhaps an axiom scheme of the first-order language of $\mathcal{N}$. It

[2] Whenever $\mathcal{A}$ is embeddable into $\mathcal{B}$ there is a structure $\mathcal{B}'$ isomorphic to $\mathcal{B}$ such that $\mathcal{A} \subseteq \mathcal{B}'$. The domain B' arises from B by interchanging the images of the elements of $\mathcal{A}$ with their originals.

is a sentence of a second-order language, about which we shall say more in **3.7**. However, this intimated limitation regarding the possibilities of formulation in first-order languages is merely an apparent one, as the undertakings of the rest of the book will show, especially those concerning axiomatic set theory in **3.4**.

In no nonstandard model $\mathcal{N}'$ is the initial segment $\mathbb{N}$ definable, indeed not even *parameter definable*, i.e., there exist no $\alpha = \alpha(x, \vec{y})$ and no $b_1, \ldots, b_n \in \mathbb{N}'$ such that $\mathbb{N} = \{a \in \mathbb{N}' \mid \mathcal{N}' \vDash \alpha\,[a, \vec{b}]\}$. Otherwise we would have $\mathcal{N}' \vDash \alpha\,\frac{0}{x} \wedge \forall x(\alpha \to \alpha\,\frac{\mathrm{S}x}{x})\,[\vec{b}]$. This statement yields $\mathcal{N}' \vDash \forall x \alpha\,[\vec{b}]$ by IS, in contradiction to $\mathbb{N}' \backslash \mathbb{N} \neq \emptyset$. The same reasoning shows that no proper initial segment $A \subset \mathbb{N}'$ without a largest element is definable in $\mathbb{N}'$, because such an A would clearly define a gap in the order of $\mathbb{N}'$. The situation can also be described as *gaps in* $\mathbb{N}'$ *are not recognizable from within.*

Introductory courses in real analysis tend to give the impression that a meaningful study of the subject requires the axiom of continuity: *Every nonempty bounded set of real numbers has a supremum.* On this basis, Cauchy and Weierstrass reformed analysis, thus banishing from mathematics the somewhat mysterious infinitesimal arguments of Leibniz, Newton, and Euler. But mathematical logic has developed methods that, to a large extent, justify the original arguments. This is undertaken in the framework of *nonstandard analysis*, developed above all by A. Robinson around 1950. In the following, we provide an indication of its basic idea.

The same construction as for $\mathcal{N}$ also provides a nonstandard model for the theory of $\mathcal{R} = (\mathbb{R}, +, \cdot, <, \{\boldsymbol{a} \mid a \in \mathbb{R}\})$, where for each real number a, a name $\boldsymbol{a}$ was added to the signature. Consider $X = Th\mathcal{R} \cup \{\boldsymbol{a} < x \mid a \in \mathbb{R}\}$. Every finite subset of X has a model on the domain $\mathbb{R}$. Thus, X is consistent and as above, a model of X represents a proper extension $\mathcal{R}^*$ of $\mathcal{R}$, a so-called *nonstandard model of analysis.* In each such model the same theorems hold as in $\mathcal{R}$. For instance, in $\mathcal{R}^*$ every polynomial of positive degree can be decomposed into linear and quadratic factors. In Chapter **5** it will be shown that the nonstandard models of $Th\mathcal{R}$ are precisely the real closed extensions of $\mathcal{R}$. All these are elementarily equivalent to $\mathcal{R}$.

For analysis, it is now decisive that the language can be enriched from the very beginning, say by the adoption of the symbols $\exp, \ln, \sin, \cos$ for the exponential, logarithmic and trigonometric functions, and further symbols for further functions. We denote a thus expanded standard model once again by $\mathcal{R}$ and a corresponding nonstandard model by $\mathcal{R}^*$. The mentioned real functions available in $\mathcal{R}$ carry over to $\mathcal{R}^*$ and maintain all properties that can be elementarily formulated. That means in fact almost all properties with interesting applications, for example

$$\forall xy\, \exp(x+y) = \exp x \cdot \exp y, \quad (\forall x{>}0)\, \exp \ln x = x, \quad \forall x\, \sin^2 x + \cos^2 x = 1,$$

as well as the addition theorems for the trigonometric functions and so on. All these functions remain continuous and repeatedly differentiable. However, the Bolzano–

Weierstrass theorem and other topological properties cannot be salvaged in full generality. They are replaced by the aforementioned infinitesimal arguments.

In a nonstandard model $\mathcal{R}^*$ of $Th\,\mathcal{R}$ with $\mathcal{R} \subseteq \mathcal{R}^*$ there not only exist infinitely large numbers c (i.e., $r < c$ for all $r \in \mathbb{R}$), but also infinitely many small positive numbers. Let c be infinite. Then $\frac{1}{r} < c \Leftrightarrow \frac{1}{c} < r$, i.e., $\frac{1}{c}$ is smaller than each positive real r, and yet is positive. That is, $\frac{1}{c}$ is fairly precisely what Leibniz once named an *infinitesimal*. Taking a somewhat closer look reveals the following picture: every real number a is sitting in a nest of nonstandard numbers $a^* \in \mathcal{R}^*$ that are only infinitesimally distinct from a. In other words, $|a^* - a|$ is an infinitesimal. Hence, quantities such as dx, dy *exist* in mathematical reality, and may once again be considered as infinitesimals in the sense of their inventor Leibniz. These quantities are precisely the elements of $\mathcal{R}^*$ infinitesimally distinct from 0.

From the existence of nonstandard models for $Th\,\mathcal{R}$, it can be concluded that the continuity axiom, just like IA, cannot be elementarily formulated. For by adjoining this axiom to those for ordered fields, $\mathcal{R}$ is characterized, up to isomorphism, as the only continuously ordered field; see e.g. [Ta4]. Hence, the order of a nonstandard model $\mathcal{R}^*$ of $Th\,\mathcal{R}$ possesses gaps. Here, too, the gaps are "not recognizable from within," since every nonempty, bounded parameter-definable subset of $\mathbb{R}^*$ has a supremum in $\mathbb{R}^*$. That is the case because in $\mathcal{R}$ and thus also in $\mathcal{R}^*$, the following *continuity schema* holds, which ensures the existence of a supremum for those sets; here $\varphi = \varphi(x, \vec{y})$ runs over all formulas such that $y, z \notin free\,\varphi$:

CS: $\exists x\varphi \wedge \exists y \forall x(\varphi \to x \leqslant y) \to \exists z \forall x[(\varphi \to x \leqslant z) \wedge \forall y((\varphi \to x \leqslant y) \to z \leqslant y)]$.

Analogous remarks can be made with respect to the complex numbers. $\mathcal{R}^*$ has an algebraically closed field extension $\mathcal{R}^*[i]$ in which familiar facts such as Euler's formula $e^{ix} = \cos x + i \cdot \sin x$ continue to hold, in particular $e^{i\pi} = -1$.

Exercises

1. Prove in PA the associativity and commutativity of $+, \cdot$, along with the law of distributivity. Before proving that $+$ is commutative derive $\mathsf{S}x + y = x + \mathsf{S}y$ in PA by induction on y. The basic arithmetical laws, including the ones about $\leqslant$ and $<$, are collected in the axiom system N on page 182.

2. Define $\leqslant$ in PA as in the text. Reflexivity and transitivity of $\leqslant$ are obvious. Derive in PA the important $x < y \leftrightarrow \mathsf{S}x \leqslant y$ (or equivalently, $y < \mathsf{S}x \leftrightarrow y \leqslant x$). Use this to prove $\vdash_{\mathsf{PA}} x \leqslant y \vee y \leqslant x$ inductively on x.

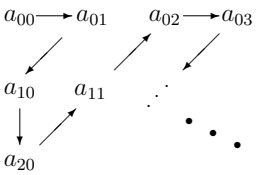
3. Verify (a) $\vdash_{\mathsf{PA}} \forall x((\forall y{<}x)\alpha\,\frac{y}{x} \to \alpha) \to \forall x\alpha$, the *schema of* $<$*-induction*,
(b) $\vdash_{\mathsf{PA}} \exists x\beta \to \exists x(\beta \wedge (\forall y{<}x)\neg\beta\,\frac{y}{x})$, the *well-ordering* (or *minimum*) *schema*,
(c) $\vdash_{\mathsf{PA}} (\forall x{<}v)\exists y\gamma \to \exists z(\forall x{<}v)(\exists y{<}z)\gamma$, the *schema of bounds*.
Here α, β, γ are any formulas in $\mathcal{L}_{ar}$ with $y \notin var\,\{\alpha, \beta\}$ and $z \notin var\,\gamma$.

3.4 ZFC and Skolem's Paradox

Before turning to further consequences of the results from **3.2**, we collect a few basic facts about countable sets. The proofs are simple and can be found in any textbook on basic set theory. A set M is called *countable* if $M = \emptyset$ or there is a surjective mapping $f\colon \mathbb{N} \to M$ (i.e., $M = \{a_n \mid n \in \mathbb{N}\}$ provided $fn = a_n$), and otherwise *uncountable*. Every subset of a countable set is itself countable. If $f\colon M \to N$ is surjective and M is countable then clearly so too is N. Sets M, N are termed *equipotent*, briefly $M \sim N$, if a bijection from M to N exists. If $M \sim \mathbb{N}$, then M is said to be *countably infinite*. A countable set can only be countably infinite or *finite*, which is to mean equipotent to $\{1, \ldots, n\}$ for some $n \in \mathbb{N}$.

The best-known uncountable set is $\mathbb{R}$, which is equipotent to $\mathfrak{P}\mathbb{N}$. The uncountability of $\mathfrak{P}\mathbb{N}$ is a particular case of an important theorem from Cantor: *The power set $\mathfrak{P}M$ of any set M has a higher cardinality than M*, i.e., no injection from M to $\mathfrak{P}M$ is surjective. The cardinality of sets will be explained to some extend in **5.1**. Here it suffices to know that two sets M, N are of the same cardinality iff $M \sim N$, and that there are countable and uncountable infinite sets.

If M, N are countable so too are $M \cup N$ and $M \times N$, as is easy to see. Moreover, a countable union $U = \bigcup_{i\in\mathbb{N}} M_i$ of countable sets M_i is again countable. A familiar proof consists in writing down U as an infinite matrix where the nth line is an enumeration of $M_n = \{a_{nm} \mid m \in \mathbb{N}\}$. Then enumerate the matrix in the zigzag manner indicated by the figure on the right, beginning with a_{00}. Accordingly, for countable M, in particular $\bigcup_{n\in\mathbb{N}} M^n$, the set of all finite sequences of elements in M is again countable, because every M^n is countable. Hence, every elementary language with a countable signature is itself countable, more precisely countably infinite.



By a *countable theory* we always mean a theory formalized in a countable language $\mathcal{L}$. We now formulate a theorem significant for many reasons.

Theorem 4.1 (Löwenheim–Skolem). *A countable consistent theory T always has a countable model.*

Proof. By Theorem 2.6, T ($\subseteq \mathcal{L}$) has a model $\mathcal{M}$ with domain A, consisting of the equivalence classes $\bar{c}$ for $c \in C$ in the set of all terms of $\mathcal{L}' = \mathcal{L}C$, where $C = \bigcup_{n\in\mathbb{N}} C_n$ is a set of new constants. By construction, C_0 is equipotent to $Var \times \mathcal{L}$ and thus countable. The same holds for every C_n, and so C is also countable. The map $c \mapsto \bar{c}$ from C to A is trivially surjective, so that $\mathcal{M}$ has a countable (possibly finite) domain, and this was the very claim. ❏

In **5.1** we will significantly generalize the theorem, but even in the above formulation it leads to noteworthy consequences. For example, there exist also countable ordered fields $\mathcal{R} = (\mathbb{R}, 0, 1, +, <, \cdot, \exp, \sin, \dots)$ as nonstandard models of $Th\,\mathcal{R}$ in which the usual theorems about real functions retain their validity. Thus, one need not really overstep the countable to obtain a rich theory of analysis.

Especially surprising is the existence, ensured by Theorem 4.1, of countable models of formalized set theory. Although set theory can be regarded as the basis for the whole of presently existing mathematics, it embraces only a few set-building principles. The most important system of formalized set theory is ZFC.

Remark. Z stands for E. Zermelo, F for A. Fraenkel, and C for AC, the axiom of choice. ZF denotes the theory resulting from the removal of AC. ZFC sets out from the principle that every element of a set is again a set, so that a distinction between sets and families of sets vanishes. Thus, ZFC speaks exclusively about sets, unlike B. Russell's type-theoretical system, in which, along with sets, so-called *urelements* (objects that are members of sets but are themselves not sets) are considered. Set theory without urelements is fully sufficient as a foundation of mathematics and for nearly all practical purposes. Even from the epistemological point of view there is no evidence that urelements occur in reality: each object can be identified with the set of all properties that distinguish it from other objects. Nonetheless, urelements are still in use as a technical tool in certain set-theoretical investigations. We mention in passing that neither ZF nor ZFC are finitely axiomatizable. This seems plausible if looking at the axioms given below, but the proof is not easy.

To make clear that ZFC is a countable first-order theory and hence belongs to the scope of applications of Theorem 4.1, we present in the following its axioms. Each of the axioms will be briefly discussed. This will be at the same time an excellent exercise in advanced formalization technics. The set-theoretical language already denoted in **2.2** by $\mathcal{L}_\in$ is one of the most conceivably simple languages and is certainly countable. Alongside $=$ it contains only the membership symbol $\in$. This symbol should be distinguished from the somewhat larger $\in$ that is used throughout in our metatheory. The variables are now called *set variables*. These will as a rule be denoted by lowercase letters as in other elementary languages. In order to make the axioms more legible, we use the abbreviations $(\forall y{\in}x)\varphi := \forall y(y \in x \to \varphi)$, $(\exists y{\in}x)\varphi := \exists y(y \in x \wedge \varphi)$. In addition, we introduce the relation of inclusion by the explicit definition $x \subseteq y \leftrightarrow \forall z(z \in x \to z \in y)$. Note also that all free variables occurring in the axioms below have to be thought of as being generalized according to our convention in **2.5**. The axioms of the theory ZFC are then the following:

AE : $\forall z(z \in x \leftrightarrow z \in y) \to x = y$ (axiom of extensionality).

AS : $\exists y \forall z(z \in y \leftrightarrow \varphi \wedge z \in x)$ (schema of separation).

Here φ runs over all $\mathcal{L}_\in$-formulas with $y \notin \mathit{free}\,\varphi$. Let $\varphi = \varphi(x, z, \vec{a})$. From AS and AE is derivable $\forall x\, \exists! y \forall z(z \in y \leftrightarrow \varphi \wedge z \in x)$. Indeed, observe the obvious derivability of

$(z\in y \leftrightarrow \varphi\wedge z\in x)\wedge(z\in y' \leftrightarrow \varphi\wedge z\in x)\to(z\in y \leftrightarrow z\in y')$ $(y,y'\notin \mathit{free}\,\varphi)$. This implies $\forall z(z\in y \leftrightarrow \varphi\wedge z\in x)\wedge\forall z(z\in y' \leftrightarrow \varphi\wedge z\in x)\to y=y'$ and hence the claim. Therefore,

$$y=\{z\in x \mid \varphi\} \leftrightarrow \forall z(z\in y \leftrightarrow \varphi\wedge z\in x)$$

is a legitimate definition in the sense of **2.6**. $\{z\in x \mid \varphi\}$ is called a *set term* and is just a suggestive writing of a function term $f_{\vec{a}}x$. This term still depends on the "parameters" $a_1,\dots,a_n$, which are the variables from $\mathit{free}\,\varphi\setminus\{x,z\}$.

The empty set can explicitly be defined by $y=\emptyset \leftrightarrow \forall z\, z\notin y$. Indeed, thanks to AS, $\exists y\forall z(z\in y \leftrightarrow z\notin x\wedge z\in x)$ is provable. This formula is clearly equivalent to $\exists y\forall z\, z\notin y$. Now, using AE, $\forall z\, z\notin y\wedge\forall z\, z\notin y'\to y=y'$ is provable, hence also $\exists! y\forall z\, z\notin y$, which legitimates the explicit definition $y=\emptyset \leftrightarrow \forall z\, z\notin y$. The next axiom is

AU : $\forall x\exists y\forall z(z\in y \leftrightarrow (\exists u\in x)\, z\in u)$ (axiom of union).

Here again, because of AE, $\exists y$ can be replaced by $\exists! y$. As in **2.6**, we may therefore define an operator on the universe,[3] denoted by $x\mapsto \bigcup x$. AU is equivalent to $\forall x\exists y\forall z((\exists u\in x)z\in u\to z\in y)$, because $\bigcup x$ can be separated from such a set y by means of AS. The following axiom could be analogously weakened.

AP : $\forall x\exists y\forall z(z\in y \leftrightarrow z\subseteq x)$ (power set axiom).

Let $\mathfrak{P}x$ denote the y that in view of AE is uniquely determined by x in AP. What first can be proved is $\forall x(x\in\mathfrak{P}\emptyset \leftrightarrow x=\emptyset)$ and $\forall x(x\in\mathfrak{P}\mathfrak{P}\emptyset \leftrightarrow x=\emptyset\vee x=\mathfrak{P}\emptyset)$. Thus, $\mathfrak{P}\mathfrak{P}\emptyset$ contains exactly two members. This is decisive for defining the pair set below.

The next axiom (again a schema) was added to those of Zermelo by Fraenkel.

AR : $\forall x\exists! y\varphi\to\forall u\exists v\forall y(y\in v \leftrightarrow (\exists x\in u)\,\varphi)$ (axiom of replacement).

Here $\varphi=\varphi(x,y,\vec{a})$ and $u,v\notin \mathit{free}\,\varphi$. If $\forall x\exists! y\varphi$ is provable, then we know from **2.6** that an operator $x\mapsto Fx$ can be introduced. By AR, the image of a set u under F is again a set v, as a rule denoted by $\{Fx \mid x\in u\}$. F may depend on further parameters $a_1,\dots,a_n$, so we better write $F_{\vec{a}}$ for F. AR is very strong; it can even be shown that AS is derivable from it. An instructive example of an application of AR, for which $\forall x\exists! y\varphi$ is certainly provable, is provided by the formula

$$\varphi(x,y,a,b) \;:=\; x=\emptyset\wedge y=a\vee x\neq\emptyset\wedge y=b.$$

For the operator $F=F_{a,b}$ defined by φ, clearly holds $F\emptyset=a$ and $Fx=b$ if $x\neq\emptyset$. Accordingly, the image of the 2-element set $\mathfrak{P}\mathfrak{P}\emptyset$ under $F_{a,b}$ contains precisely the two members a,b. We therefore define $\{a,b\}:=\{F_{a,b}(x) \mid x\in\mathfrak{P}\mathfrak{P}\emptyset\}$ and call this the *pair set* of a,b. We then put $a\cup b:=\bigcup\{a,b\}$ (while $a\cap b:=\{z\in a \mid z\in b\}$ already exists from AS). Further, let $\{a\}:=\{a,a\}$ and $\{a_1,\dots,a_{n+1}\}=\{a_1,\dots,a_n\}\cup\{a_{n+1}\}$ for $n\geqslant 2$. Now we can write and moreover prove that $\mathfrak{P}\emptyset=\{\emptyset\}$, $\mathfrak{P}\mathfrak{P}\emptyset=\{\emptyset,\{\emptyset\}\},\dots$ The *ordered pair* of a,b is defined after Kuratowski as $(a,b):=\{\{a\},\{a,b\}\}$.

[3] A frequently used synonym for the domain of a ZFC-model. The word "function" is avoided here because functions are specific objects of a universe, namely sets of ordered pairs.

We now have at our disposal the implements necessary to develop elementary set theory. Beginning with sets of ordered pairs it is possible to model relations and functions and all concepts building upon them, even though the existence of an infinite set remains unprovable. Mathematical requirements demand their existence, though then the borders of our experience with finite sets are transgressed. The easiest way to get infinite sets is using the set operator $x \mapsto \mathtt{S}x$, where $\mathtt{S}x := x \cup \{x\}$.

$\mathsf{AI}: \quad \exists u[\emptyset \in u \wedge \forall x(x \in u \to \mathtt{S}x \in u)]$ (axiom of infinity).

Such a set u contains $\emptyset$, $\mathtt{S}\emptyset = \emptyset \cup \{\emptyset\} = \{\emptyset\}$, $\mathtt{SS}\emptyset = \{\emptyset, \{\emptyset\}\}, \ldots$ and is therefore infinite in the naive sense. This holds in particular for the smallest set u of this type, denoted by ω. In formalized set theory ω plays the role of the set of natural numbers. ω contains $0 := \emptyset$, $1 := \mathtt{S}0 = \{0\}$, $2 := \mathtt{S}1 = \{\emptyset, \{\emptyset\}\}$, etc. Generally, $n + 1 := \mathtt{S}n = n \cup \{n\}$ which easily computes to $n + 1 = \{0, \ldots, n\}$. Thus, natural numbers are represented by certain variable-free set terms, called *ω-terms*.

In everyday mathematics the following axiom is basically dispensable:

$\mathsf{AF}: \quad (\forall x \neq \emptyset)(\exists y \in x)(\forall z \in x)\, z \notin y$ (axiom of foundation or regularity).

Put intuitively: Every $x \neq \emptyset$ contains an $\in$-minimal element y. AF precludes the possibility of "$\in$-circularity" $x_0 \in \cdots \in x_n \in x_0$. In particular, there are no sets x with $x \in x$. Other consequences of AF will not be discussed here.

From the theory denoted by ZF with the axioms so far, ZFC results by adjoining the *axiom of choice*, which has various equivalent formulations.

$\mathsf{AC}: \quad \forall u[\emptyset \notin u \wedge (\forall x \in u)(\forall y \in u)(x \neq y \to x \cap y = \emptyset) \to \exists z(\forall x \in u)\exists! y(y \in x \wedge y \in z)]$.

It states that for every set (or family thereof) u of disjunct nonempty sets x there exist a set z, a *choice set*, that picks up precisely one element from each x in u.

The above expositions clearly show that ZFC can be understood as a first-order theory. In some sense, ZFC is even the purest such theory, because all sophisticated proof methods that occur in mathematics, for instance transfinite induction and recursion and every other type of induction and recursion, can be made explit and derived purely predicate logically in ZFC without particular difficulty.

Whereas mathematicians regularly transgress the framework of a theory, even one that is unambiguously defined by first-order axioms, in that they make use of combinatorial, number- or set-theoretical tools wherever it suits them, set theory, as it stands now, imposes upon itself an upper limit. Within ZFC, all sophisticated proof and definition techniques gain an elementary character, so to speak.

As a matter of fact, there are no pertinent arguments against the claim that the whole of mathematics can be treated within the frame of ZFC as a single first-order theory, a claim based on general mathematical experience that is highly interesting for the philosophy of mathematics. However, one should not make a religion out of this insight, because for mathematical practice it is of limited significance.

If ZFC is consistent—and no one really doubts this assumption although there is no way of proving it—then by Theorem 4.1, ZFC also has a countable model. The existence of such a ZFC-model $\mathcal{V} = (V, \in^{\mathcal{V}})$ is at first glance paradoxical because the existence of uncountable sets is easily provable *within* ZFC. An example is $\mathfrak{P}\omega$. On the other hand, because $(\mathfrak{P}\omega)^{\mathcal{V}} \subseteq V$, it must be true (from outside) that $(\mathfrak{P}\omega)^{\mathcal{V}}$ contains only countably many elements. Thus, the notion *countable* has a different meaning "inside and outside the world $\mathcal{V}$," which comes completely unexpectedly. This is the so-called *paradox of Skolem*.

The explanation of Skolem's paradox is that the countable model $\mathcal{V}$, to put it figuratively, is "thinned out" and contains fewer sets and functions than expected. Indeed, roughly put, it contains just enough to satisfy the axioms, yet not, for instance, some bijection from $\omega^{\mathcal{V}}$ to $(\mathfrak{P}\omega)^{\mathcal{V}}$, which, seen from the outside, certainly exists. Therefore, the countable set $(\mathfrak{P}\omega)^{\mathcal{V}}$ is uncountable from the perspective of the world $\mathcal{V}$. In other words, uncountability is not an absolute concept.

Moreover, the universe V of a ZFC-model is by definition a set, whereas it is easy to prove $\vdash_{\mathsf{ZFC}} \neg\exists v\forall z\, z \in v$, i.e., there is no "universal set." Thus, seen from within, V is too big to be a set. $\neg\exists v\forall z\, z \in v$ is verified as follows: the hypothesis $\exists v\forall z\, z \in v$ entails with AE and AS the existence of the "Russellian set" $u = \{x \in v \mid x \notin x\}$. That is, $\exists v\forall z\, z \in v \vdash_{\mathsf{ZFC}} \exists u\forall x(x \in u \leftrightarrow x \notin x)$. On the other hand, by Example 1 on page 58, $\vdash_{\mathsf{ZFC}} \neg\exists u\forall x(x \in u \leftrightarrow x \notin x)$, whence $\vdash_{\mathsf{ZFC}} \neg\exists v\forall z\, z \in v$. Accordingly, even the notion of a set depends on the model. There is no absolute definition of a set.

None of the above has anything to do with ZFC's being incomplete.[4] Mathematics has no problem with the fact that its basic theory is incomplete and, in principle, cannot be rendered complete. More of a problem is the lack of undisputed criteria for extending ZFC in a way coinciding with truth or at least with our intuition.

Exercises

1. Let T be an elementary theory with arbitrarily large finite models. Prove using the compactness theorem that T also has an infinite model.

2. Suppose $\mathcal{A} = (A, <)$ is an infinite, well-ordered set (see **2.1**). Show that there exists a non-well-ordered set elementarily equivalent to $\mathcal{A}$.

3. Using the ZFC axioms, confirm the well-definedness of ω in the text. For this assertion it suffices to prove $\vdash_{\mathsf{ZFC}} \exists u[\emptyset \in u \wedge \forall x(x \in u \to x \cup \{x\} \in u)]$.

4. Let $\mathcal{V} \vDash \mathsf{ZFC}$. Show that there exists a model $\mathcal{V}' \vDash \mathsf{ZFC}$ such that $\mathcal{V}' \supseteq \mathcal{V}$ and a $U \in V'$ with $a \in^{\mathcal{V}'} U$ for all $a \in V$. Then necessarily $V' \supset V$.

[4] In **6.5** the incompleteness of ZFC and all its axiomatic extensions is proved. The most prominent example of a sentence independent of ZFC is the continuum hypothesis stated on page 135.

3.5 Enumerability and Decidability

Of all the far-reaching consequences of the completeness theorem, perhaps the most significant is the effective enumerability of all tautologies of a countable first-order language. Once Gödel had proved this theorem, the hope grew that the decidability problem for tautologies might soon be resolved. Indeed, the wait was not long, and a few years after Gödel's result Church proved the problem to be unsolvable for sufficiently expressive languages. This section is intended to provide only a brief glimpse of enumeration and decision problems as they appear in logic. We consider them more rigorously in the Chapters **5** and **6**.

The term *effectively enumerable* will be made more precise in **6.1** by the notion of *recursive enumerability*. At this stage, our explanation of this notion must be somewhat superficial, though like that for a decidable set it is highly visualizable. Put roughly, a set M of natural numbers, say, or syntactic objects, finite structures, or similar objects is called *effectively* (or *recursively*) *enumerable* if there exist an algorithm that delivers stepwise the elements of M. Thus, in the case of an infinite set M, the algorithm does not stop its execution by itself.

The calculus of natural deduction enables first of all an effective enumeration of all provable finite sequences of a first-order language with at most countably many logical symbols, i.e., all pairs (X, α) such that $X \vdash \alpha$ and X is finite, at least in principle. First of all, we imagine all initial sequents as enumerated in an ongoing, explicitly producible sequence $S_0, S_1, \ldots$ Then it is systematically checked whether one of the sequent rules is applicable; the resulting sequents are then enumerated in a second sequence and so on. Leaving aside problems concerning the storage capacity of such a deduction machine, as well as the difficulties involved in evaluating the flood of information that would pour from it, it is simply a question of organization to create a program that enumerates all provable finite sequents.

Moreover, it can be seen without difficulty that the tautologies of a countable language $\mathcal{L}$ are effectively enumerable; one need only pick out from an enumeration procedure of provable sequents (X, α) those such that $X = \emptyset$. In short, the aforementioned deduction machine delivers stepwise a sequence $\alpha_0, \alpha_1, \ldots$ (without repetitions if so desired) that consists of exactly the tautologies of $\mathcal{L}$. This would be somewhat easier with the calculus in **3.6**. However, we cannot in this way obtain a decision procedure as to whether or not any given formula $\alpha \in \mathcal{L}$ is a tautology, for we do not know whether α ever appears in the produced sequence. We prove rigorously in **6.5** that in fact such an algorithm does not exist provided $\mathcal{L}$ contains at least a binary predicate or operation symbol. Decision procedures exist only for $\mathcal{L}_=$ (cf. **5.2**) or when the signature contains only unary predicate and constant symbols, and at most one unary operation symbol; see also [BGG].

The deduction machine can also be applied to enumerate the theorems of a given axiomatizable theory T, in that parallel to the enumeration process for all provable sequents of the language, a process is also set going that enumerates all axioms of T. It must then continually be checked for the enumerated sequents whether all their premises occur as already-enumerated assertions; if so, then the conclusion of the sequent in question is provable in T. The preceding considerations constitute an informal proof of the following theorem. A rigorous proof free of merely intuitive arguments is provided by Theorem 6.2.4.

Theorem 5.1. *The theorems of an axiomatizable theory are effectively enumerable.*

Almost all theories considered in mathematics are axiomatizable, including formalized set theory ZFC and Peano arithmetic PA. While the axiom systems of these two theories are infinite and cannot be replaced by finite ones, these sets of axioms are evidently decidable. Our experience hitherto shows us that all those theorems of mathematics held to be proved are also provable in ZFC, and therefore, according to Theorem 5.1, all mathematical theorems can in principle be stepwise generated by a computer. This fact is theoretically important, even if it has little far-reaching practical significance at present.

Recall the notion of a complete theory. Among the most important examples is the theory of the real closed fields (Theorem 5.5.5). A noteworthy feature of complete and axiomatizable theories is their *decidability*. We call a theory *decidable* if the set of its theorems is a decidable set of formulas, and otherwise *undecidable*. We prove the next theorem intuitively; it is generalized by Exercise 3. A strict proof, based on the rigorous definition of decidability in **6.1**, will later be provided by Theorem 6.4.4 on page 191.

Theorem 5.2. *A complete axiomatizable theory T is decidable.*

Proof. By Theorem 5.1 let $\alpha_0, \alpha_1, \ldots$ be an effective enumeration of all sentences provable in T. A decision procedure consists simply in comparing for given $\alpha \in \mathcal{L}^0$ the sentences α and $\neg\alpha$ in the nth construction step of $\alpha_0, \alpha_1, \ldots$ with α_n. If $\alpha = \alpha_n$ then $\vdash_T \alpha$; if $\alpha = \neg\alpha_n$ then $\nvdash_T \alpha$. This process certainly terminates, because due to the completeness of T, either α or $\neg\alpha$ will appear in the enumeration sequence $\alpha_0, \alpha_1, \ldots$ of the theorems of T. ❑

Conversely, a complete decidable theory is trivially axiomatizable (by T itself). Thus, for complete theories, "decidable" and "axiomatizable" mean one and the same thing. A consistent theory has a model and hence at least one *completion*, i.e., a complete extension in the same language. The only completion of a complete theory T is T itself. An incomplete theory has at least two distinct completions. A decidable incomplete theory even possesses a decidable completion (Exercise 4).

Hence, a theory all completions of which are undecidable is itself undecidable. We will meet such theories, even finitely axiomatizable ones, in **6.5**. On the other hand, if T has finitely many completions only, $T_0, \ldots, T_n$, all of which are decidable, then so is T.[5] Indeed, according to Exercise 2, $\alpha \in T \Leftrightarrow \alpha \in T_i$ for all $i \leqslant n$.

In the early stages in the development of fast computing machines, high hopes were held concerning the practical carrying out of mechanized decision procedures. For various reasons, this optimism has since been muted, though skillfully employed computers can be helpful not only in verifying proofs but also in finding them. This area of applied logic is called *automated theorem proving* (ATP). Convincing examples include computer-supported proofs of the four-colour conjecture, the Robbins problem about a particular axiomatization of Boolean algebras, and Bieberbach's conjecture in function theory. ATP is used today both in hardware and software verification, for instance, in integrated circuit (chip) design and verification. A quick source of information about automated theorem proving is the Internet.

Despite of these applications, even a highly developed artificial-intelligence system has presently no chance of simulating the heuristic approach in mathematics, where a precise proof from certain hypotheses is frequently only the culmination of a series of considerations flowing from the imagination. However, that is not to say that such a system may not be creative in a new way, for it is not necessarily the case that the human procedural method, influenced by all kinds of pictorial thoughts, is the sole means to gaining mathematical knowledge.

Exercises

1. Let $T' = T + \alpha$ ($\alpha \in \mathcal{L}^0$) be a finite extension of T. Show that if T is decidable so too is T' (cf. Lemma 6.5.3).

2. Prove that a consistent theory T coincides with the intersection of all its completions, in short $T = \bigcap\{T' \supseteq T \mid T' \text{ complete}\}$.

3. Show that the following are equivalent for a consistent theory T:

   (i) T has finitely many extensions, (ii) T has finitely many completions.

   Moreover, show that a consistent theory T with n completions has $2^n - 1$ consistent extensions, T included ($n = 1$ iff T itself is complete).

4. Using the Lindenbaum construction of **1.4**, show that an incomplete decidable and countable theory T has a decidable completion ([TMR, p. 15]).

[5] The elementary absolute (plane) geometry T has precisely two completions, Euclidean and non-Euclidean (or hyperbolic) geometry. Both are axiomatizable, hence decidable. Completeness follows in both cases from that of the elementary theory of real numbers, Theorem 5.5.5. Thus, absolute geometry is decidable as well. Further applications can be found in **5.2**.

3.6 Complete Hilbert Calculi

The sequent calculus of **3.1** models natural deduction sufficiently well. But it is nonetheless advantageous to use a Hilbert calculus for some purposes, for instance the arithmetization of formal proofs. Such calculi are based on logical axioms and rules of inference like modus ponens MP: $\alpha, \alpha \to \beta / \beta$, also called *Hilbert-style rules*. These rules can be understood as premiseless sequent rules. In a Hilbert calculus, deductions are drawn from a fixed set of formulas X, for instance, the axioms a theory, with the inclusion of the logical axioms, as in **1.6**. In the case $X = \emptyset$ one deduces from the logical axioms alone, and only tautologies are established.

In the following we prove the completeness of a Hilbert calculus in the logical symbols $\neg, \wedge, \forall, =$. It will be denoted here by $\vdash$. MP is its only rule of inference. The calculus refers to an arbitrary elementary language $\mathcal{L}$ and is essentially an extension of the corresponding propositional Hilbert calculus treated in **1.6**. Once again, implication, defined by $\alpha \to \beta := \neg(\alpha \wedge \neg\beta)$, will play a useful part.

The *logical axiom system* Λ of our calculus is taken to consist of all formulas $\forall x_1 \cdots \forall x_n \varphi$, where φ is a formula of the form $\Lambda 1$–$\Lambda 10$ below, and $n \geqslant 0$. For example, due to $\Lambda 9$, $x = x$, $\forall x\, x = x, \forall y\, x = x$, $\forall x \forall y\, x = x$ are logical axioms, even though $\forall y$ is meaningless in the last two formulas. One may also say that Λ is the set of all formulas that can be derived from $\Lambda 1$–$\Lambda 10$ by means of the rule MQ: $\alpha / \forall x \alpha$. However, MQ is not a rule of inference of the calculus, nor is it provable. We will later take a closer look at this rule.

$\Lambda 1$: $(\alpha \to \beta \to \gamma) \to (\alpha \to \beta) \to \alpha \to \gamma$, $\Lambda 2$: $\alpha \to \beta \to \alpha \wedge \beta$,

$\Lambda 3$: $\alpha \wedge \beta \to \alpha$, $\alpha \wedge \beta \to \beta$, $\Lambda 4$: $(\alpha \to \neg\beta) \to \beta \to \neg\alpha$,

$\Lambda 5$: $\forall x \alpha \to \alpha\, \frac{t}{x}$ ($\alpha, \frac{t}{x}$ collision-free), $\Lambda 6$: $\alpha \to \forall x \alpha$ ($x \notin \mathit{free}\, \alpha$)

$\Lambda 7$: $\forall x(\alpha \to \beta) \to \forall x \alpha \to \forall x \beta$, $\Lambda 8$: $\forall y \alpha\, \frac{y}{x} \to \forall x \alpha$ ($y \notin \mathit{var}\, \alpha$),

$\Lambda 9$: $t = t$, $\Lambda 10$: $x = y \to \alpha \to \alpha\, \frac{y}{x}$ (α prime).

It is easy to recognize $\Lambda 1$–$\Lambda 10$ as tautologies. For $\Lambda 1$–$\Lambda 4$ this is clear by **1.6**. For $\Lambda 5$–$\Lambda 8$ the reasoning proceeds straightforwardly by accounting for the corollary on page 56 and the logical equivalences in **2.4**. For $\Lambda 9$ and $\Lambda 10$ this is obvious.

Axiom $\Lambda 5$ corresponds to the rule $(\forall 1)$ of the calculus in **3.1**, while $\Lambda 6$ serves to deal with superfluous prefixes. The role of $\Lambda 7$ will become clear in the completeness proof for $\vdash$, and $\Lambda 8$ is part of bound renaming. $\Lambda 9$ and $\Lambda 10$ control the treatment of identity. If φ is a tautology then, for any prefix block $\forall \vec{x}$, so too is $\forall \vec{x} \varphi$. Thus, Λ consists solely of tautologies. The same holds for all formulas derivable from Λ using MP, for $\vDash \alpha, \alpha \to \beta$ obviously implies $\vDash \beta$.

Let $X \vdash \alpha$ if there exists a *proof* $\Phi = (\varphi_0, \ldots, \varphi_n)$ *of* α *from* X, that is, $\alpha = \varphi_n$, and for all $k \leqslant n$ either $\varphi_k \in X \cup \Lambda$ or there exists some φ such that φ and $\varphi \to \varphi_k$

appear as members of Φ before φ_k. This definition and its consequences are the same as in **1.6**. As is the case there, it holds that $X \mathrel{|\!\sim} \alpha, \alpha \to \beta \Rightarrow X \mathrel{|\!\sim} \beta$. Moreover, the induction theorem 1.6.1 also carries over unaltered, and its application will often be announced by the heading "proof by induction on $X \mathrel{|\!\sim} \alpha$." For instance, the soundness of $\mathrel{|\!\sim}$ is proved by induction on $X \mathrel{|\!\sim} \alpha$, where soundness is defined as usual, that is to mean $X \mathrel{|\!\sim} \alpha \;\Rightarrow\; X \models \alpha$, for all X and α. In short, $\mathrel{|\!\sim} \subseteq \models$.

The completeness of $\mathrel{|\!\sim}$ can now be relatively easily be traced back to that of the rule calculus $\vdash$ of **3.1**. Indeed, much of the work was already undertaken in **1.6**, and we can immediately formulate the completeness of $\mathrel{|\!\sim}$.

Theorem 6.1 (Completeness theorem for $\mathrel{|\!\sim}$). $\mathrel{|\!\sim} = \models$.

Proof. $\mathrel{|\!\sim} \subseteq \models$ has already been verified. $\models \subseteq \mathrel{|\!\sim}$ follows from the claim that $\mathrel{|\!\sim}$ satisfies all nine basic rules of $\vdash$. This implies $\vdash \subseteq \mathrel{|\!\sim}$, and since $\vdash = \models$ we then have $\models \subseteq \mathrel{|\!\sim}$. For the propositional rules $(\wedge 1)$ through $(\neg 2)$ the claim holds according to their proof for the Hilbert calculus in **1.6**. The Lemmas 1.6.2 through 1.6.5 carry over word for word, because we have kept the four axioms on which the proofs are based and have taken no new rules into account. $(\forall 1)$ follows immediately from $\Lambda 5$ using MP, and (IR) is dealt with by $\Lambda 9$. Only $(\forall 2)$ and $(=)$ provide us with a little work which, by the way, will clear up the role of axioms $\Lambda 6$, $\Lambda 7$, and $\Lambda 8$.

$(\forall 2)$: Suppose $x \notin \mathit{free}\, X$. We first prove $X \mathrel{|\!\sim} \alpha \Rightarrow X \mathrel{|\!\sim} \forall x\alpha$ by induction on $X \mathrel{|\!\sim} \alpha$. *Initial step*: If $\alpha \in X$ then x is not free in α. So $X \mathrel{|\!\sim} \alpha \to \forall x\alpha$ using $\Lambda 6$, and MP yields $X \mathrel{|\!\sim} \forall x\alpha$. If $\alpha \in \Lambda$ then also $\forall x\alpha \in \Lambda$, and hence likewise $X \mathrel{|\!\sim} \forall x\alpha$. *Induction step*: Let $X \mathrel{|\!\sim} \alpha, \alpha \to \beta$ and $X \mathrel{|\!\sim} \forall x\alpha, \forall x(\alpha \to \beta)$ according to the induction hypothesis. This yields $X \mathrel{|\!\sim} \forall x\alpha, \forall x\alpha \to \forall x\beta$ by Axiom $\Lambda 7$ and MP and hence the induction claim $X \mathrel{|\!\sim} \forall x\beta$. Now, to verify $(\forall 2)$, let $X \mathrel{|\!\sim} \alpha\, \frac{y}{x}$ and $y \notin \mathit{free}\, X \cup \mathit{var}\, \alpha$. By what we have just proved, we get $X \mathrel{|\!\sim} \forall y \alpha\, \frac{y}{x}$. This, MP, and $X \mathrel{|\!\sim} \forall y \alpha\, \frac{y}{x} \to \forall x\alpha$ (Axiom $\Lambda 8$) yield the conclusion $X \mathrel{|\!\sim} \forall x\alpha$ of $(\forall 2)$. Thus, $\mathrel{|\!\sim}$ satifies rule $(\forall 2)$.

$(=)$: Let α be a prime formula and $X \mathrel{|\!\sim} s = t, \alpha\, \frac{s}{x}$. Further, let y be a variable $\neq x$ not appearing in s and α. Then certainly $X \mathrel{|\!\sim} \forall x \forall y (x = y \to \alpha \to \alpha\, \frac{y}{x})$, because the latter is a logical axiom in view of $\Lambda 10$. By the choice of y, rule $(\forall 1)$ then yields

$$X \mathrel{|\!\sim} [\forall y (x = y \to \alpha \to \alpha\, \tfrac{y}{x})]\, \tfrac{s}{x} \;=\; \forall y (s = y \to \alpha\, \tfrac{s}{x} \to \alpha\, \tfrac{y}{x}).$$

Because of $y \notin \mathit{var}\, \alpha, s$ and $\alpha\, \frac{y}{x}\, \frac{t}{y} = \alpha\, \frac{t}{x}$, a repeated application of $(\forall 1)$ gives

$$X \mathrel{|\!\sim} [s = y \to \alpha\, \tfrac{s}{x} \to \alpha\, \tfrac{y}{x}]\, \tfrac{t}{y} \;=\; s = t \to \alpha\, \tfrac{s}{x} \to \alpha\, \tfrac{y}{x}\, \tfrac{t}{y} \;=\; s = t \to \alpha\, \tfrac{s}{x} \to \alpha\, \tfrac{t}{x}\ .$$

Since $X \mathrel{|\!\sim} s = t, \alpha\, \frac{s}{x}$ by assumption, two applications of MP then leads to the desired conclusion $X \mathrel{|\!\sim} \alpha\, \frac{t}{x}$. ❑

A special case of the completeness theorem 6.1 is the following

Corollary 6.2. *For any $\alpha \in \mathcal{L}$, the following properties are equivalent:*

(i) $\vdash\alpha$, *that is,* α *is derivable from* Λ *by means of* MP *only,*

(ii) α *is derivable from* $\Lambda 1$–$\Lambda 10$ *by means of* MP *and* MQ,

(iii) $\vDash \alpha$, *i.e.,* α *is a tautology.*

The equivalence of (i) and (iii) renders especially intuitive the possibility to construct a "deduction machine" that effectively enumerates the set of all tautologies of $\mathcal{L}$. Here, we are dealing with just one rule of inference, modus ponens, so we need just the help of a machine to list the logical axioms, a "deducer" to check whether MP is applicable and, if so, to apply it, and an output unit that emits the results and feeds them back into the deducer for further processing. However, similar to the case of a sequent calculus, such a procedure is not actually practicable; the distinction between significant and insignificant derivations is too difficult to be taken into account. Who would be interested to find in the listing such a weird looking tautology as for instance $\exists x(rx \to \forall y\, ry)$?

Next we want to show that the global consequence relation $\overset{\mathsf{G}}{\models}$ defined in **2.5** can also be completely characterized by a Hilbert calculus. It is necessary only to adjoin the generalization rule MQ to the calculus $\vdash$. Thus, the resulting Hilbert calculus, defined by $\overset{\mathsf{G}}{\vdash}$, then has two rules of inference, MP and MQ. Like every Hilbert calculus, $\overset{\mathsf{G}}{\vdash}$ is transitive, that is, $X \overset{\mathsf{G}}{\vdash} Y \;\&\; Y \overset{\mathsf{G}}{\vdash} \alpha \Rightarrow X \overset{\mathsf{G}}{\vdash} \alpha$. To see this, let $X \overset{\mathsf{G}}{\vdash} Y$, $Y \overset{\mathsf{G}}{\vdash} \alpha$ and let Φ be a proof of α from Y. By replacing every formula $\varphi \in Y$ appearing in Φ by a proof of φ from X, the resulting sequence is clearly a proof of α from X. The completeness of $\overset{\mathsf{G}}{\vdash}$ follows essentially from that of $\vdash$:

Theorem 6.3 (Completeness theorem for $\overset{\mathsf{G}}{\vdash}$). $\overset{\mathsf{G}}{\vdash} = \overset{\mathsf{G}}{\models}$.

Proof. Certainly $\overset{\mathsf{G}}{\vdash} \subseteq \overset{\mathsf{G}}{\models}$, since both MP and MQ are sound for $\overset{\mathsf{G}}{\models}$. Now let $X \overset{\mathsf{G}}{\models} \alpha$, so that $X^{\mathsf{G}} \vDash \alpha$ by (1) of **2.5**. This yields $X^{\mathsf{G}} \vdash \alpha$ by Theorem 6.1, and thus a fortiori $X^{\mathsf{G}} \overset{\mathsf{G}}{\vdash} \alpha$. But since $X \overset{\mathsf{G}}{\vdash} X^{\mathsf{G}}$, transitivity provides $X \overset{\mathsf{G}}{\vdash} \alpha$. ❑

We now discuss a notion of equal interest for both logic and computer science. $\alpha \in \mathcal{L}^0$ is called *generally valid in the finite* if $\mathcal{A} \vDash \alpha$ for all finite structures $\mathcal{A}$. Examples of such sentences α not being tautologies can be constructed in every signature that contains at least a unary function symbol or a binary relation symbol. For instance, consider $\forall x\forall y(fx = fy \to x = y) \to \forall y\exists x\, y = fx$. This states in (A, f) that if $f^{\mathcal{A}}$ is injective, it is also surjective, which is true iff A is finite. Thus, *Taut* is properly extended by the set of sentences generally valid in the finite, *Tautfin*.

In each signature, *Tautfin* is an example of a theory T with the *finite model property*, i.e., every sentence α compatible with T has a finite T-model. More generally, the theory $T = Th\boldsymbol{K}$ of any class $\boldsymbol{K}$ of finite $\mathcal{L}$-structures has the finite model property. Indeed, if $T + \alpha$ is consistent, i.e., $\neg\alpha \notin T$, then $\mathcal{A} \nvDash \neg\alpha$ for some $\mathcal{A} \in \boldsymbol{K}$, hence $\mathcal{A} \vDash \alpha$. This is the case, for example, for the theories FSG and FG of all finite

semigroups and finite groups, respectively, in $\mathcal{L}\{\circ\}$. Both theories are undecidable. As regards FSG, the proof is not particularly difficult; see **6.6**.

Unlike *Taut*, as a rule, *Tautfin* is not axiomatizable. This is the claim of

Theorem 6.4 (Trachtenbrot). *$\mathit{Tautfin}_L$ is not (recursively) axiomatizable for any signature L containing at least one binary operation or relation symbol.*

Proof. We restrict ourselves to the first case; for a binary relation symbol, the same follows easily by means of interpretation (Theorem 6.6.3). If $\mathit{Tautfin}_L$ were axiomatizable it would also be decidable because of the finite model property, Exercise 2. This also clearly holds for $\mathit{Tautfin}_{\mathcal{L}\{\circ\}}$, and by Exercise 1 in **3.5**, so too for FSG, because FSG is $\mathit{Tautfin}_{\mathcal{L}\{\circ\}}$ extended by a single sentence, the law of associativity. But as already mentioned, FSG is undecidable. ❑

The theorem is in fact a corollary of much stronger results that have been established in the meantime. For the newer literature on decision problems of this type consult [Id]. Unlike FG, the theory of finite abelian groups, as well as of all abelian groups, is decidable ([Sz]). The former is a proper extension of the latter; for instance, $\forall x \exists y\, y + y \equiv x \to \forall x(x + x \equiv 0 \to x \equiv 0)$ does not hold in all abelian groups, though it does in all finite ones. Verifying this is a highly informative exercise.

As early as 1922 Behmann discovered by quantifier elimination that *Taut* possesses the finite model property provided the signature contains only unary predicate symbols; one can also prove this without difficulty by the Ehrenfeucht game of **5.3**. In this case then, $\mathit{Tautfin} = \mathit{Taut}$, because $\alpha \notin \mathit{Taut}$ implies $\neg\alpha$ is satisfiable and therefore has a finite model. Thus, $\alpha \notin \mathit{Tautfin}$. This proves $\mathit{Tautfin} \subseteq \mathit{Taut}$ and hence $\mathit{Tautfin} = \mathit{Taut}$. With the Ehrenfeucht game also a quite natural axiomatization of the theory FO of all finite ordered sets is obtained. This is an exercise in **5.3**.

Exercises

1. Show that MQ is unprovable in $\vdash$ ($X \vdash \alpha \Rightarrow X \vdash \forall x\alpha$ does not hold in general).

2. Suppose (i) a theory T has the finite model property, (ii) the finite T-models are effectively enumerable (more precisely, a system of representatives thereof up to isomorphism). Show that (a) the sentences α refutable in T are effectively enumerable, (b) if T is axiomatizable then it is also decidable.

3. Let T be a *finitely* axiomatizable theory with the finite model property. Show by working back to Exercise 2 that T is decidable.

4. Show that $\forall x \exists y\, y + y = x \to \forall x(x + x = 0 \to x = 0)$ holds in all finite abelian groups. Moreover, provide an example of an infinite abelian group for which the above proposition fails.

3.7 First-Order Fragments and Extensions

Subsequent to Gödel's completeness theorem it makes sense to investigate some fragments and extensions of first-order languages aiming at a formal characterization of deduction *inside* the fragment or extension. In this section we shall present some results in this regard. First-order fragments are formalisms that come along without the full means of expression in an elementary language, for instance by the omission of some or all logical connectives, or restricted quantification. These formalisms are interesting for various reasons, partly because of the growing interest in automatic information processing with its more or less restricted user interface. The poorer a linguistic fragment, the more modest the possibilities for the formulation of sound rules. Therefore, the completeness problem for fragments is in general nontrivial.

A useful example dealt with more closely is the *language of equations*, whose only formulas are equations of a fixed algebraic signature. We think tacitly of the variables in the equations as being generalized and call them *identities*, though we often speak of equations. Theories with axiom systems of identities are called *equational theories* and their model classes *equational-defined classes* or *varieties*.

Let Γ denote a set of equations defining an equational theory, γ a single equation, and assume $\Gamma^{\mathtt{G}} \models \gamma$. By Theorem 2.7 there is a formal proof for γ from Γ. But because of the special form of the equations, it can be expected that one need not the whole formalism to verify $\Gamma^{\mathtt{G}} \models \gamma$. Indeed, Theorem 7.2 states that the *Birkhoff rules* (B0)–(B4) below, taken from [Bi], suffice. This result is so pleasing because operating with (B0)–(B4) remains completely inside the language of equations. The rules define a Hilbert-style calculus denoted by $\vdash^B$ and look as follows:

(B0) $/t = t$, (B1) $s = t / t = s$, (B2) $t = s, s = t' / t = t'$,

(B3) $t_1 = t'_1, \ldots, t_n = t'_n / f t_1 \cdots t_n = f t'_1 \cdots t'_n$, (B4) $s = t / s^\sigma = t^\sigma$.

Here σ is a global substitution, though as explained in **2.2** it would suffice to consider just simple σ. (B0) has no premise which means that $t = t$ is derivable from any set of identities (or $t = t$ is added as an axiom to Γ). These rules are formally stated with respect to unquantified equations. However, we think of all variables as being generalized in a formal derivation sequence. We are forced to do this by the soundness requirement of (B4), because $(s = t)^{\mathtt{G}} \models s^\sigma = t^\sigma$ but not $s = t \models s^\sigma = t^\sigma$, in general. To verify $\Gamma \vdash^B \gamma \Rightarrow \Gamma^{\mathtt{G}} \models \gamma$, we need only to show that the property $\Gamma^{\mathtt{G}} \models \gamma$ is closed under (B0)–(B4), i.e., $\mathcal{A} \models t = t$ (which is trivial), $\mathcal{A} \models s = t \Rightarrow \mathcal{A} \models t = s$, etc. We have already come across the rules of $\vdash^B$ in **3.1**, stated there as Gentzen-style rules; they ensure that by $s \approx t :\Leftrightarrow \Gamma \vdash^B s = t$ a congruence in the term algebra $\mathcal{T}$ is defined, similar as in Lemma 2.5. (B4) states the *substitution invariance* of $\approx$, which is to mean $s \approx t \Rightarrow s^\sigma \approx t^\sigma$. Let $\mathcal{F}$ be the factor structure of $\mathcal{T}$ by $\approx$, and let $\bar{t}$ denote the congruence class modulo $\approx$ determined by the term t, so that

(1) $\overline{t_1} = \overline{t_2} \Leftrightarrow \Gamma \vdash^B t_1 = t_2$.

Further let $w\colon Var \to \mathcal{F}$, say $x^w = \overline{t_x}$, with arbitrarily chosen $t_x \in x^w$. Any such choice determines a global substitution $\sigma_w\colon x \mapsto t_x$. Induction on t easily yields

(2) $t^{\mathcal{F},w} = \overline{t^\sigma}$ $(\sigma := \sigma_w)$.

Lemma 7.1. $\Gamma \vdash^B t_1 \equiv t_2 \;\Leftrightarrow\; \mathcal{F} \vDash t_1 \equiv t_2$.

Proof. Let $\Gamma \vdash^B t_1 = t_2$. By (B4) also $\Gamma \vdash^B t_1^\sigma = t_2^\sigma$, so that $\overline{t_1^\sigma} = \overline{t_2^\sigma}$. Therefore, $t_1^{\mathcal{F},w} = t_2^{\mathcal{F},w}$ using (2). Since w was arbitrary, it follows that $\mathcal{F} \vDash t_1 = t_2$. Now suppose the latter and let $\varkappa$ be the so-called *canonical valuation* $x \mapsto \overline{x}$. Here we choose $\sigma_\varkappa = \iota$ (the identical substitution), hence $t_i^{\mathcal{F},\varkappa} = \overline{t_i}$ by (2). $\mathcal{F} \vDash t_1 = t_2$ implies $\mathcal{F} \vDash t_1^{\mathcal{F},\varkappa} = t_2^{\mathcal{F},\varkappa}$, and since $t_i^{\mathcal{F},\varkappa} = \overline{t_i}$ we get $\overline{t_1} = \overline{t_2}$ and so $\Gamma \vdash^B t_1 = t_2$ by (1). ❑

Theorem 7.2 (Birkhoff's completeness theorem). $\Gamma \vdash^B t_1 = t_2 \Leftrightarrow \Gamma^{\mathtt{G}} \vDash t_1 = t_2$.

Proof. The direction $\Rightarrow$ is the soundness of $\vdash^B$. Now let $\Gamma^{\mathtt{G}} \vDash t_1 = t_2$. According to Lemma 7.1, certainly $\mathcal{F} \vDash \Gamma$, or equivalently $\mathcal{F} \vDash \Gamma^{\mathtt{G}}$. Therefore $\mathcal{F} \vDash t_1 = t_2$. Applying Lemma 7.1 once again then yields $\Gamma \vdash^B t_1 = t_2$. ❑

This proof is distinguished on the one hand by its simplicity and on the other by its highly abstract character. It has manifold variations and is valid in a corresponding sense, for example, for sentences of the form $\forall\vec{x}\pi$ with *arbitrary* prime formulas π of any given first-order language. It is rather obvious how to strengthen the Birkhoff rules to cover this more general case: Keep (B0), (B1), and (B3) and replace the conclusions of (B3) and (B4) by arbitrary prime formulas of the language.

There is also a special calculus for sentences of the form

(3) $\forall\vec{x}\,(\gamma_1 \wedge \cdots \wedge \gamma_n \to \gamma_0)$ $\qquad (n \geqslant 0$, all γ_i equations),

called *quasi-identities*. The classes of models of axioms of the form (3) are called *quasi-varieties*. The latter are highly important both in algebra and logic. (B0) is retained and (B1)–(B3) are replaced by the following premiseless rules:

$$/x = y \to y = x, \quad /x = y \wedge y = z \to x = z, \quad /\bigwedge\nolimits_{i=1}^n x_i = y_i \to f\vec{x} = f\vec{y}.$$

Besides an adaptation of (B4), some rules are required for the formal handling of the premises $\gamma_1, \ldots, \gamma_n$ in (3), for instance their permutability (for details see e.g., [Se]). A highly important additional rule is here a variant of the cut rule, namely

$$\alpha \wedge \delta \to \gamma, \alpha \to \delta / \alpha \to \gamma \qquad (\alpha \text{ a conjunction of equations}).$$

The most interesting case for automated information processing, where Hilbert rules remaining inside the fragment still provide completeness, is that of universal Horn theories. Here, roughly speaking, the equations γ_i in (3) may be arbitrary prime formulas. Horn theories are treated in Chapter **4**. But for enabling a real machine implementation, the calculus considered there (the resolution calculus) is different from a Hilbert- or a Gentzen-style calculus.

Now we consider a few of the numerous possibilities for extending first-order languages to increase the power of expression: We say a language $\mathcal{L}' \supseteq \mathcal{L}$ *of the same* signature as $\mathcal{L}$ is *more expressive* than $\mathcal{L}$ if for at least one sentence $\alpha \in \mathcal{L}'$, $\operatorname{Md}\alpha$ is distinct from all $\operatorname{Md}\beta$ for $\beta \in \mathcal{L}$. In $\mathcal{L}'$, some of the properties of first-order languages are lost. Indeed, the claim of the next theorem is that first-order languages are optimal in regard to the richness of their applications.

Lindström's Theorem (see [EFT] or [CK]). *There is no language of a given signature that is more expressive than the first-order language and for which both the compactness theorem and the Löwenheim–Skolem theorem hold.*

Many-sorted languages. In describing geometric facts it is convenient to use several variables, for points, lines, and, depending on dimension, also for geometrical objects of higher dimension. For every argument of a predicate or operation symbol of such a language, it is useful to fix its *sort*. For instance, the incidence relation of plane geometry has arguments for points and lines. For function symbols, the sort of their values must additionally be given. If $\mathcal{L}$ is of sort k and $\boldsymbol{v}_0^s, \boldsymbol{v}_1^s, \ldots$ are variables of sort s $(1 \leqslant s \leqslant k)$ then every relation symbol r is assigned a sequence $(s_1, \ldots, s_n)$; in languages not containing function symbols, prime formulas beginning with r have the form $rx_1^{s_1} \cdots x_n^{s_n}$, where $x_i^{s_i}$ denotes a variable of sort s_i.

Many-sorted languages represent only an inessential extension of the concept hitherto expounded, provided the sorts are given equal rights. Instead of a language $\mathcal{L}$ with k sorts of variables, we can consider a one-sorted language $\mathcal{L}'$ with additional unary predicate symbols $P_1, \ldots, P_k$ and the adoption of certain new axioms: $\exists x P_i x$ for $i = 1, \ldots, k$ (no sort is empty, otherwise it could be omitted) and $\neg\exists x(P_i x \wedge P_j x)$ for $i \neq j$ (sort disjunction). For example, plane geometry could also be described in a one-sorted language with the additional predicates *pt* (to be a point) and *li* (to be a line). Apart from a few differences in dealing with term insertion, many-sorted languages behave almost exactly like one-sorted languages.

Second-order languages. Some frequently quoted axioms, e.g., the induction axiom IA, may be looked upon as second-order sentences. The simplest extension of an elementary language to one of higher order is the *monadic second-order language*, a two-sorted language that has a special interpretation for the second sort. Let us consider such a language $\mathcal{L}$ with variables $x, y, z, \ldots$ for individuals, variables $X, Y, Z, \ldots$ for sets of these individuals, along with at least one binary relation symbol $\in$ but no function symbols. Prime formulas are $x = y$, $X = Y$, and $x \in X$. An $\mathcal{L}$-structure is generally of the form $(A, B, \in)$ where $\in \subseteq A \times B$. The goal is that by formulating additional axioms such as $\forall XY[\forall x(x \in X \leftrightarrow x \in Y) \rightarrow X = Y]$ (which corresponds to the axiom AE in **3.4**), $\in$ should be interpretable as the membership relation $\in$, hence B should consist of the subsets of A. This goal is not fully attain-

able, but nearly so: axioms can be found such that B can be regarded only as a subset of $\mathfrak{P}A$, with $\in$ interpreted as $\in$. This also works by adding sort variables for members of $\mathfrak{P}\mathfrak{P}A$, $\mathfrak{P}\mathfrak{P}\mathfrak{P}A$, etc. This "completeness of the theory of types" plays a basic role in the higher nonstandard analysis.

A more enveloping second-order language, $\mathcal{L}_{II}$, is won by adopting quantifiable variables for any relations and operations on the domains of individuals. But even for $\mathcal{L} = \mathcal{L}_{=}$, $\mathcal{L}_{II}$ fails to satisfy both the finiteness theorem and the Löwenheim–Skolem theorem (Theorem 4.1). The former does not hold because a theorem α_{fin} can be given in $\mathcal{L}_{II}$ such that $\mathcal{A} \vDash \alpha_{\text{fin}}$ if and only if A is finite. For it is not difficult to prove that A is finite iff A can be ordered such that every nonempty subset of A possesses both a smallest and largest element. This property can effortlessly be formulated by means of a binary and a unary predicate variable.

The Löwenheim–Skolem theorem is also easily refutable for $\mathcal{L}_{II}$; one need only write down in $\mathcal{L}_{II}$ the sentence 'there exists a continuous order on A without smallest or largest element'. This sentence has no countable model. For if there were such a model, it would be isomorphic to the ordered set of rationals according to a theorem of Cantor (Example 2 in **5.2**) and therefore has gaps, contradicting our assumptions.

There is still a more serious problem as regards $\mathcal{L}_{II}$: The ZFC-axioms, seen as axioms of the underlying set theory, do not suffice to establish what a tautology in $\mathcal{L}_{II}$ should actually be. For instance, the continuum hypothesis CH (see page 135) can be easily formulated as an $\mathcal{L}_{II}$-sentence, α_{CH}. But CH is independent of ZFC. Thus, if CH is true, α_{CH} is an $\mathcal{L}_{II}$ tautology, otherwise not. It does not look as though mathematical intuition suffices to decide this question unambiguously.

New quantifiers. A simple syntactic extension $\mathcal{L}_{\mathfrak{O}}$ of a first-order language $\mathcal{L}$ is obtained by taking on a new quantifier denoted by $\mathfrak{O}$, which formally is to be handled as the $\forall$-quantifier. However, in a model $\mathcal{M} = (\mathcal{A}, w)$, a new interpretation of $\mathfrak{O}$ is provided by means of the satisfaction clause

(0) $\quad \mathcal{M} \vDash \mathfrak{O}x\alpha \;\Leftrightarrow\; \{a \in A \mid \mathcal{M}^a_x \vDash \alpha\}$ is infinite.

With this interpretation, we write $\mathcal{L}^0_{\mathfrak{O}}$ instead of $\mathcal{L}_{\mathfrak{O}}$, since yet another interpretation of $\mathfrak{O}$ will be discussed. $\mathcal{L}^0_{\mathfrak{O}}$ is more expressive than $\mathcal{L}$, as seen by the fact, for example, that the finiteness theorem for $\mathcal{L}^0_{\mathfrak{O}}$ no longer holds: Let X be the collection of all sentences $\exists_n$ (there exist at least n elements) plus $\alpha_{\text{fin}} := \neg\mathfrak{O}x\, x = x$ (there exist only finitely many elements). Every finite subset of X has a model, but X itself does not. All the same, $\mathcal{L}^0_{\mathfrak{O}}$ still satisfies the Löwenheim–Skolem theorem. This can be proved straightforwardly with the methods of **5.1**. Once again, because of the missing finiteness theorem there cannot be a complete rule calculus for $\mathcal{L}^0_{\mathfrak{O}}$. Otherwise, just as in **3.1**, one could prove the finiteness theorem after all. However, there are several nontrivial, correct Hilbert-style rules for $\mathcal{L}^0_{\mathfrak{O}}$, for instance

(Q1) $/\neg\mathfrak{O}x(x = y \vee x = z)$ $(x \neq y, z)$, (Q2) $\mathfrak{O}x\alpha/\mathfrak{O}y\alpha\,{}^{y}_{x}$ $(y \notin free\,\alpha)$,

(Q3) $\forall x(\alpha \to \beta)/\mathfrak{O}x\alpha \to \mathfrak{O}x\beta$, (Q4) $\mathfrak{O}x\exists y\alpha, \neg\mathfrak{O}y\exists x\alpha/\exists y\mathfrak{O}x\,\alpha$.

Intuitively, rule (Q1) (which has no premises) says that the pair $\{y, z\}$ is finite. (Q2) is bound renaming. (Q3) says that a set containing an infinite subset is itself infinite. (Q4) is rendered intuitive for $\mathcal{M} = (\mathcal{A}, w) \vDash \mathfrak{O}x\exists y\alpha, \neg\mathfrak{O}y\exists x\alpha$ and for $\alpha = \alpha(x, y)$ as follows: Let $A_b = \{a \in A \mid \mathcal{A} \vDash \alpha(a, b)\}$. Then $\mathcal{M} \vDash \mathfrak{O}x\exists y\,\alpha$ states '$\bigcup_{b\in A} A_b$ is infinite'. $\mathcal{M} \vDash \neg\mathfrak{O}y\exists x\,\alpha$ says 'there exist only finitely many indices b such that $A_b \neq \emptyset$', the conclusion $\exists y\mathfrak{O}x\alpha$ therefore 'A_b is infinite for at least one index b'. Hence (Q4) expresses altogether the fact that the union of a finite system of finite sets is itself finite. Now replace the satisfaction clause (0) by

(1) $\mathcal{M} \vDash \mathfrak{O}x\alpha \;\Leftrightarrow\; \{a \in A \mid \mathcal{M}^a_x \vDash \alpha\}$ is uncountable.

Also with this interpretation, (Q1)–(Q4) are sound for $\mathcal{L}^1_{\mathfrak{O}}$ $(= \mathcal{L}_{\mathfrak{O}}$ with the interpretation (1)$)$. Rule (Q4) now evidently expresses that a countable union of countable sets is again countable. Moreover, the logical calculus $\vdash^1$ resulting from the basic rules of **3.1** by adjoining (Q1)–(Q4) is, surprisingly, complete for these semantics when restricted to countable sets X. Thus, $X \vdash^1 \alpha \;\Leftrightarrow\; X \vDash \alpha$, for any countable $X \subseteq \mathcal{L}^1_{\mathfrak{O}}$ ([CK]). This fact implies the following compactness theorem for $\mathcal{L}^1_{\mathfrak{O}}$: *If every finite subset of a countable set of formulas $X \subseteq \mathcal{L}^1_{\mathfrak{O}}$ has a model then so too does X.* For uncountable sets of formulas this is false in general.

Programming languages. All languages hitherto discussed are of static character inasmuch as there are spatially and temporally independent truth values for given valuations w in a structure $\mathcal{A}$. But one can also connect a first-order language $\mathcal{L}$ in various ways with a *programming language* having *dynamic* character.

We describe here a simple example of such language, $\mathcal{PL}$. The elements of $\mathcal{PL}$ are called *programs*, denoted by $\mathcal{P}, \mathcal{Q}, \ldots$ and are defined below. The dynamic character arises by modifying traditional semantics as follows: A program $\mathcal{P}$ starts with a valuation $w\colon Var \to A$ (the domain of a given $\mathcal{L}$-structure $\mathcal{A}$) and alters stepwise the values of the variables as a run of the program $\mathcal{P}$ proceeds in time. If $\mathcal{P}$ terminates upon feeding in w, i.e., the calculation ends, the result is a new valuation $w^{\mathcal{P}}$. Otherwise we take $w^{\mathcal{P}}$ to be undefined. The description of this in general only partially defined operation $w \mapsto w^{\mathcal{P}}$ is called the *procedural semantics* of $\mathcal{PL}$.

It is possible to meaningfully consider issues of completeness, say, for this type of semantics, too. The syntax of $\mathcal{PL}$ is specified as follows: The logical signature of $\mathcal{L}$ is extended by the symbols `WHILE`, `DO`, `END`, `:=`, and `;` (the semicolon serves only as a separator for concatenated programs and could be omitted if programs are arranged 2-dimensionally, which we will not do for the sake of brevity). *Programs* on $\mathcal{L}$ are defined inductively as strings of symbols in the following manner:

1. For any variable $x \in Var$ and term $t \in \mathcal{T}_{\mathcal{L}}$, the string $x := t$ is a program.

2. If α is an open formula in $\mathcal{L}$ and $\mathcal{P}, \mathcal{Q}$ are programs, so too are the strings $\mathcal{P}\,;\mathcal{Q}$ and $\mathtt{WHILE}\,\alpha\,\mathtt{DO}\,\mathcal{P}\,\mathtt{END}$.

No other strings are programs in this context. $\mathcal{P}\,;\mathcal{Q}$ is to mean that first $\mathcal{P}$ and then $\mathcal{Q}$ are executed. Let $\mathcal{P}^n$ be the n-times repeated execution of $\mathcal{P}$, more precisely $\mathcal{P}^0$ is the *empty* program ($w^{\mathcal{P}^0} = w$) and $\mathcal{P}^{n+1} = \mathcal{P}^n\,;\mathcal{P}$. The procedural semantics for $\mathcal{PL}$ are made more precise by the following stipulations:

(a) $w^{x:=t} = w\,\frac{t^w}{x}$ (i.e., w alters at most the value of the variable x).

(b) If $w^{\mathcal{P}}$ and $(w^{\mathcal{P}})^{\mathcal{Q}}$ are defined, so too is $w^{\mathcal{P};\mathcal{Q}}$, and $w^{\mathcal{P};\mathcal{Q}} = (w^{\mathcal{P}})^{\mathcal{Q}}$.

(c) For $\mathcal{Q} := \mathtt{WHILE}\,\alpha\,\mathtt{DO}\,\mathcal{P}\,\mathtt{END}$ let $w^{\mathcal{Q}} = w^{\mathcal{P}^k}$ with k specified below.

According to our intuition regarding the "WHILE loop," k is the smallest number such that $\mathcal{A} \vDash \alpha\,[w^{\mathcal{P}^i}]$ for all $i < k$ and $\mathcal{A} \nvDash \alpha\,[w^{\mathcal{P}^k}]$, provided such a k exists and all $w^{\mathcal{P}^i}$ for $i \leqslant k$ are well defined. Otherwise $w^{\mathcal{Q}}$ is considered to be undefined. If $k = 0$, that is, $\mathcal{A} \nvDash \alpha\,[w]$, then $w^{\mathcal{Q}} = w$, which amounts to saying that $\mathcal{P}$ is not executed at all, in accordance with the meaning of WHILE in all programming languages.

Example. Let $\mathcal{L} = \mathcal{L}\{0, \mathtt{S}, \mathtt{Pd}\}$ and let $\mathcal{A} = (\mathbb{N}, 0, \mathtt{S}, \mathtt{Pd})$, where S and Pd respectively denote the successor and predecessor functions, and let $\mathcal{P}$ be the program

$$z := x\,;v := y\,;\mathtt{WHILE}\,v \neq 0\,\mathtt{DO}\,z := \mathtt{S}z\,;v := \mathtt{Pd}\,v\,\mathtt{END}.$$

If x and y initially have the values $x^w = m$ and $y^w = n$, the program ends with $z^{w^{\mathcal{P}}} = m+n$. In other words, $\mathcal{P}$ terminates for every input m, n for x, y and computes the output $m + n$ in the variable z while x, y keep their initial values.

In $\mathcal{PL}$, the well-known program schema $\mathtt{IF}\,\alpha\,\mathtt{THEN}\,\mathcal{P}\,\mathtt{ELSE}\,\mathcal{Q}\,\mathtt{END}$ is definable by $x := 0\,;\mathtt{WHILE}\,\alpha \wedge x = 0\,\mathtt{DO}\,\mathcal{P}\,;x := \mathtt{S}0\,\mathtt{END}\,;\mathtt{WHILE}\,x = 0\,\mathtt{DO}\,\mathcal{Q}\,;x := \mathtt{S}0\,\mathtt{END}$, where x is a variable not appearing in $\mathcal{P}$, $\mathcal{Q}$, and α.

Exercises

1. Show that a variety $\boldsymbol{K}$ is closed with respect to homomorphism, subalgebra, and direct product.[6]

2. Show that $\mathcal{L}^1_{\mathfrak{O}}$ and $\mathcal{L}_{II}$ do not satisfy the Löwenheim–Skolem theorem, and that $\mathcal{L}^1_{\mathfrak{O}}$ violates the finiteness theorem for uncountable sets of formulas.

3. Express the continuum hypothesis as a theorem of $\mathcal{L}_{II}$.

4. Verify the correctness of the definition of the program $\mathtt{IF}\,\alpha\,\mathtt{THEN}\,\mathcal{P}\,\mathtt{ELSE}\,\mathcal{Q}\,\mathtt{END}$ given in the text.

5. Define the loop $\mathtt{DO}\,\mathcal{P}\,\mathtt{UNTIL}\,\alpha\,\mathtt{END}$ by means of the $\mathtt{WHILE}\,\alpha\,\mathtt{DO}\,\mathcal{P}\,\mathtt{END}$-loop.

[6] If, conversely, a class $\boldsymbol{K}$ has these three properties and is closed under isomorphisms then $\boldsymbol{K}$ is a variety. This is Birkhoff's **HSP** theorem, a theorem of Universal algebra; see e.g. [Mo].

Chapter 4

The Foundations of Logic Programming

Logic programming aims not so much at solving numerical problems in science and technology, rather at treating information processing in general, in particular at the creation of expert systems of artificial intelligence. A distinction has to be made between logic programming as theoretical subject matter and the widely used programming language for practical tasks of this kind, PROLOG. In regards to the latter, we confine ourselves to a presentation of a somewhat simplified version, FF nonetheless preserves the typical features.

The notions dealt with in **4.1** are of fairly general nature. Their origin lies in certain theoretical questions posed by mathematical logic, and they took shape before the invention of the computer. For certain sets of formulas, in particular for sets of universal Horn formulas, which are very important for logic programming, term models are obtained canonically. For a full understanding of **4.1**, Chapters **1** and **2** should have been read, and to some extent also Chapter **3**. The newcomer need not understand all details of **4.1** at once, but should learn at least what a Horn formula is and after a glance at the theorems may then continue with **4.2**.

The resolution method and its combination with unification proposed in [Rob] and applied in PROLOG were directly inspired by mechanical information processing. This method is also of significance for tasks of automated theorem proving which extends beyond logic programming. We treat resolution first in the framework of propositional logic in **4.2**. Its highlight, the resolution theorem, is proved constructively, without recourse to the propositional compactness theorem. In **4.3** unification is dealt with in an understandable way. **4.4** presents the combination of resolution with unification and its application to logic programming. An elementary introduction to this area is also offered by [Ll], while [Do] is more challenging. For practical PROLOG programming, [CM] may be a good reference.

4.1 Term Models and Horn Formulas

In the proof of Lemma 3.2.5 as well as in Lemma 3.7.1 we have come across models whose domains are equivalence classes of terms of a first-order language $\mathcal{L}$. In general, a *term model* is to mean an $\mathcal{L}$-model $\mathcal{F}$ whose domain F is the set of congruence classes $\overline{t}$ of a congruence $\approx_{\mathcal{F}}$ on the algebra $\mathcal{T}$ of all $\mathcal{L}$-terms t. If $\approx_{\mathcal{F}}$ is the identity in $\mathcal{T}$, one identifies F with $\mathcal{T}$ so that then $\overline{t} = t$. Function symbols and constants are interpreted canonically: $f^{\mathcal{F}}(\overline{t_1}, \ldots, \overline{t_n}) := \overline{ft_1 \cdots t_n}$ and $c^{\mathcal{F}} := \overline{c}$. No particular condition is posed on realizing the relation symbols of $\mathcal{L}$. Further let $\kappa : x \mapsto \overline{x}$ ($x \in Var$). This is called the *canonical* valuation. In the terminology of **2.3**, $\mathcal{F} = (\mathfrak{F}, \kappa)$, where $\mathfrak{F}$ denotes the underlying $\mathcal{L}$-structure with the domain $F = \{\overline{t} \mid t \in \mathcal{T}\}$. We claim that independent of a specification of the $r^{\mathcal{F}}$,

(1) $t^{\mathcal{F}} = \overline{t}$ for all $t \in \mathcal{T}$,

(2) $\mathcal{F} \vDash \forall \vec{x} \alpha \;\Leftrightarrow\; \mathcal{F} \vDash \alpha \frac{\vec{t}}{\vec{x}}$ for all $\vec{t} \in \mathcal{T}^n$ (α open).

(1) is proved by an easy term induction (cf. (d) page 79). (2) follows from left to right by Corollary 2.3.6. The converse runs as follows: $\mathcal{F} \vDash \alpha \frac{\vec{t}}{\vec{x}}$ for all $\vec{t} \in \mathcal{T}^n$ implies $\mathcal{F}^{\overline{t_1} \cdots \overline{t_n}}_{x_1 \cdots x_n} \vDash \alpha$ for all $t_1, \ldots, t_n \in \mathcal{T}$ in view of Theorem 2.3.5 and (1). But this means that $\mathcal{F} \vDash \forall \vec{x} \alpha$, because the $\overline{t}$ for $t \in \mathcal{T}$ exhaust the domain of $\mathcal{F}$.

Essential for both theoretical logic and automated theorem proving is the question for which consistent formula sets $X \subseteq \mathcal{L}$ can a term model be constructed *inside* $\mathcal{L}$. For certain sets X a positive answer is given by Theorems 1.1 and 1.3 below.

Definition. The *term model* $\mathcal{F} = \mathcal{F}X$ *associated* with a given set of formulas X is that term model for which $\approx_{\mathcal{F}X}$ and $r^{\mathcal{F}X}$ are defined by

$$s \approx_{\mathcal{F}X} t \;\Leftrightarrow\; X \vdash s = t; \quad r^{\mathcal{F}X}\overline{t_1} \cdots \overline{t_n} \;\Leftrightarrow\; X \vdash rt_1 \cdots t_n\,.$$

By (1), $\mathcal{F}X \vDash s = t \Leftrightarrow \overline{s} = \overline{t} \Leftrightarrow X \vdash s = t$. Similarly $\mathcal{F}X \vDash r\vec{t} \Leftrightarrow X \vdash r\vec{t}$. In general, $\mathcal{F}X$ is not a model for X. What follows from our definition is only

(3) $\mathcal{F}X \vDash \pi \;\Leftrightarrow\; X \vdash \pi$ (π prime).

Most of the time X will be the axiom system of some theory T. We then also write $\mathcal{F}T$ for $\mathcal{F}X$ and $s \approx_T t$ for $s \approx_{\mathcal{F}X} t$ (s, t are equivalent in T, see page 66). An example in which $\mathcal{F}T$ is indeed a model for T (a special case of Theorem 1.3) is

Example 1. Let T be the theory of semigroups in $\mathcal{L}\{\circ\}$. Every term t is equivalent in T to a term in left association, denoted by $x_1 \cdots x_n$ ($\circ$ is not written); here $x_1, \ldots, x_n$ is an enumeration of the variables of t in the order of appearance from left to right, possibly with repetitions. In other words, $t \approx_T x_1 \cdots x_n$; for instance, $\boldsymbol{v}_0((\boldsymbol{v}_1\boldsymbol{v}_0)\boldsymbol{v}_1) \approx_T \boldsymbol{v}_0\boldsymbol{v}_1\boldsymbol{v}_0\boldsymbol{v}_1$. Further, $(x_1 \cdots x_n) \circ (y_1 \cdots y_m) \approx_T x_1 \cdots x_n y_1 \cdots y_m$, as is easily seen inductively on m. Moreover, $x_1 \cdots x_n \approx_T y_1 \cdots y_m \Leftrightarrow m = n \;\&\; x_i = y_i$. Therefore, one can identify the term classes modulo $\approx_T$ with the words over the

alphabet *Var*. More precisely, the algebra $\mathfrak{F}$ underlying $\mathcal{F}T$ is isomorphic to the word-semigroup over the alphabet *Var* and is thereby also a model of T.

As already announced earlier, we slightly extend the concept of a model. Let $\mathcal{L}^k$ and Var_k be defined as in **2.2**. Pairs $(\mathcal{A}, w)$ with $dom\, w \supseteq Var_k$ are called $\mathcal{L}^k$*-models*. Here w need not be defined for $\boldsymbol{v}_k, \boldsymbol{v}_{k+1}, \ldots$, or an allocation to these variables may have been deliberately "forgotten." In this sense $\mathcal{L}$-structures are also $\mathcal{L}^0$-models; simply choose the empty valuation whenever $k = 0$, hence $Var_k = \emptyset$. Note that an $\mathcal{L}^n$-model can be understood as an $\mathcal{L}^k$-model whenever $k \leqslant n$.

Let $\mathcal{T}_k := \{t \in \mathcal{T} \mid var\, t \subseteq Var_k\}$. To ensure that the set of ground terms $\mathcal{T}_0$ is nonempty, we tacitly assume in the following that $\mathcal{L}$ contains at least one constant when considering $\mathcal{T}_0$. Clearly $\mathcal{T}_k$ is a subalgebra of $\mathcal{T}$, for $t_1, \ldots, t_n \in \mathcal{T}_k \Rightarrow f\vec{t} \in \mathcal{T}_k$. The concept of a term model can equally be related to $\mathcal{L}^k$ as follows:

Let $\approx$ be a congruence in $\mathcal{T}_k$ and $\mathfrak{F}_k$ the factor structure $\mathcal{T}_k/\!\approx$ whose domain is $F_k = \{\bar{t} \mid t \in \mathcal{T}_k\}$. $\mathfrak{F}_k$ is extended canonically to an $\mathcal{L}^k$-model by the valuation $x \mapsto \overline{x}$ for $x \in Var_k$. This $\mathcal{L}^k$-model is subsequently denoted by $\mathcal{F}_k$. For each k, the following conditions are verified as with (1), (2), (3). $\mathcal{F}_k X$ in (3_k) is defined analogously to $\mathcal{F}X$ but with respect to sets of formulas $X \subseteq \mathcal{L}^k$.

(1_k) $\quad t^{\mathcal{F}_k} = \bar{t}$ for all $t \in \mathcal{T}_k$,

(2_k) $\quad \mathcal{F}_k \vDash \forall \vec{x}\alpha \;\Leftrightarrow\; \mathcal{F}_k \vDash \alpha\, \frac{\vec{t}}{\vec{x}}$ for all $\vec{t} \in \mathcal{T}_k^n \quad (\alpha$ open),

(3_k) $\quad \mathcal{F}_k X \vDash \pi \;\Leftrightarrow\; X \vdash \pi \qquad (\pi$ a prim formula from $\mathcal{L}^k)$.

Let $\varphi = \forall \vec{x}\alpha$ with an open formula α. Then $\alpha\, \frac{\vec{t}}{\vec{x}}$ is called an *instance* of φ. And if $t_i \in \mathcal{T}_k$ for $i = 1, \ldots, n$ then $\alpha\, \frac{\vec{t}}{\vec{x}}$ is called a $\mathcal{T}_k$*-instance*, for $k = 0$ also a *ground instance of* φ. Let $\mathrm{GI}(X)$ denote the set of ground instances of all $\varphi \in X$. Note that $\mathrm{GI}(X) \neq \emptyset$ whenever $X \neq \emptyset$ since $\mathcal{L}$ contains constants if $k = 0$ is considered.

Theorem 1.1. *Let U $(\subseteq \mathcal{L})$ be a set of universal formulas and $\tilde{U}$ the set of all instances of the formulas in U. Then the following are equivalent:*

(i) *U is consistent,* (ii) *$\tilde{U}$ is consistent,* (iii) *U has a term model in $\mathcal{L}$.*

The same holds if $U \subseteq \mathcal{L}^k$ and $\tilde{U}$ denotes the set of all $\mathcal{T}_k$-instances of the formulas in U. In particular, a set $U \subseteq \mathcal{L}^0$ of $\forall$-sentences is consistent iff $\mathrm{GI}(U)$ *is consistent, provided $\mathcal{L}$ contains constants.*

Proof. (i) $\Rightarrow$ (ii) is clear, because $U \vdash \tilde{U}$. (ii) $\Rightarrow$ (iii): Let $\mathcal{M} \vDash \tilde{U}$ and $\mathcal{F} := \mathcal{F}X$ for $X := \{\varphi \in \mathcal{L} \mid \mathcal{M} \vDash \varphi\}$. By (3), $\mathcal{F} \vDash \pi \Leftrightarrow \mathcal{M} \vDash \pi$ $(\Leftrightarrow X \vdash \pi)$ for prime formulas π. Induction on $\wedge, \neg$ yields $\mathcal{F} \vDash \varphi \Leftrightarrow \mathcal{M} \vDash \varphi$, for all open φ. Since $\mathcal{M} \vDash \tilde{U}$ we thus have $\mathcal{F} \vDash \tilde{U}$. But this yields $\mathcal{F} \vDash U$ according to (2). (iii) $\Rightarrow$ (i): Trivial. For the case $U \subseteq \mathcal{L}^k$ the proof runs similarly using (3_k), (2_k), and $\mathcal{F}_k = \mathcal{F}_k X$. ❑

By this theorem, a consistent set U of universal *sentences* has a term model $\mathcal{F}_0$. For logic programming the important case is that in which U is $=$-free. Then U

has a model on the set of all terms (Exercise 2 in **3.2**). By choosing such $\mathcal{M}$ in the proof of Theorem 1.1, we can do without a factorization in the construction of $\mathcal{F}_0 X$ ($X = \{\alpha \in \mathcal{L}^0 \mid \mathcal{M} \vDash \alpha\}$). Such a model is called a *Herbrand model* for U. Its domain $\mathcal{T}_0$ consists of all ground terms and is named the *Herbrand universe* of $\mathcal{L}$. In general, U has many Herbrand models $\mathcal{A}$ on the same domain $\mathcal{T}_0$ with the same constants and functions: $c^{\mathcal{A}} = c$ and $f^{\mathcal{A}}(t_1, \ldots, t_n) = f\vec{t}$ for all $\vec{t} \in \mathcal{T}_0^n$. Only the relations may vary. If U is a universal Horn theory (to be explained below), then U has a distinguished Herbrand model, the *minimal* Herbrand model; see page 110.

Example 2. Let $U \subseteq \mathcal{L}\{0, \mathtt{S}, <\}$ consist of the two $=$-free universal sentences

$$\text{(a) } \forall x\, x < \mathtt{S}x; \quad \text{(b) } \forall x \forall y \forall z (x < y \wedge y < z \to x < z).$$

Here the Herbrand universe $\mathcal{T}_0$ consists of all ground terms $\underline{n}$ ($= \mathtt{S}^n 0$). Obviously, $\mathcal{N} := (\mathbb{N}, 0, \mathtt{S}, <) \vDash U$. Hence, we may choose $\mathcal{N}$ in the construction of a Herbrand model $\mathcal{F}_0$ for U in the proof of Theorem 1.1. One may even say that $\mathcal{N}$ itself is a Herbrand model for U and indeed the minimal one; see Example 5.

Remark 1. With Theorem 1.1 the problem of satisfiability for $X \subseteq \mathcal{L}$ can basically be reduced to a propositional satisfiability problem. By Exercise 5 in **2.6**, X is after adding new operation symbols satisfiably equivalent to a set U of $\forall$-formulas which, by Theorem 1.1, is in turn satisfiably equivalent to the set of open formulas $\tilde{U}$. Now replace the prime formulas π occurring in the formulas of $\tilde{U}$ with propositional variables p_π, distinct variables for distinct prime formulas, as in the examples of **1.5**. One then obtains a satisfiably equivalent set of propositional formulas. This works straight on $=$-free sets of $\forall$-formulas. By dealing with the congruence conditions for $=$ (page 109), this method can be generalized for sets of $\forall$-formulas with identity but is then more involved.

Although we will focus on a certain variant of the next theorem, its basic concern (the construction of explicit solutions of existential assertions) is the same in logic programming and other areas of automated information processing. Herbrand's theorem was originally a purely proof-theoretic statement.

Theorem 1.2 (Herbrand's theorem). *Let $U \subseteq \mathcal{L}$ be a set of universal formulas and $\exists \vec{x} \alpha \in \mathcal{L}$ for an open formula α. Finally, let $\tilde{U}$ be the set of all $\mathcal{T}$-instances of formulas in U. Then the following properties are equivalent:*

(i) $U \vdash \exists \vec{x} \alpha$,

(ii) $U \vdash \bigvee_{i \leqslant m} \alpha \frac{\vec{t}_i}{\vec{x}}$ for some m and some $\vec{t_0}, \ldots, \vec{t_m} \in \mathcal{T}^n$,

(iii) $\tilde{U} \vdash \bigvee_{i \leqslant m} \alpha \frac{\vec{t}_i}{\vec{x}}$ for some m and some $\vec{t_0}, \ldots, \vec{t_m} \in \mathcal{T}^n$.

The same holds if $\mathcal{L}$ is replaced here by $\mathcal{L}^k$, $\mathcal{T}$ by $\mathcal{T}_k$, and $\mathcal{T}^n$ by $\mathcal{T}_k^n$, for each $k \geqslant 0$.

Proof. Because $U \vdash \tilde{U}$, certainly (iii)⇒(ii)⇒(i). It therefore remains to be shown (i)⇒(iii): by (i), $X = U, \forall \vec{x} \neg \alpha$ is inconsistent, hence also $\tilde{U} \cup \{\neg \alpha \frac{\vec{t}}{\vec{x}} \mid \vec{t} \in \mathcal{T}_k^n\}$ by Theorem 1.1. With this, (iii) follows already propositionally (Exercise 1 in **1.4**). ❏

The theorem's assumption that α is open is essential, as can be seen from the example $\vdash \exists x\alpha$ with $\alpha = \forall y(ry \to rx)$ (Example 2 in **2.6**). There are no terms $t_0, \dots, t_m$ (variables in this case) such that $\vdash \bigvee_{i \leqslant m} \alpha \frac{t_i}{x}$, as is readily confirmed.

We now define *Horn formulas* for a given language $\mathcal{L}$ inductively. The definition covers also the propositional case; omit everything that refers to quantification.

Definition. (a) Literals are basic Horn formulas. If α is a prime formula and β a basic Horn formula, then $\alpha \to \beta$ is a basic Horn formula. (b) Basic Horn formulas are Horn formulas. If α, β are Horn formulas then so too is $\alpha \wedge \beta$, along with $\forall x\alpha$ and $\exists x\alpha$. Horn formulas without free variables will be called *Horn sentences*.

For instance, $\forall y(ry \to rx)$ and $\forall x(y \in x \to x \notin y)$ are Horn formulas. By definition, $\alpha_1 \to \dots \to \alpha_n \to \beta$ $(n \geqslant 0)$ is the general form of a basic Horn formula of $\mathcal{L}$, where the α_i are prime formulas and β is a literal. Note that in the propositional case the α_i are propositional variables and β is a propositional literal.

We also call any formula α a (basic) Horn formula if it is equivalent to an original (basic) Horn formula. Thus, since $\alpha_1 \to \dots \to \alpha_n \to \beta \equiv \beta \vee \neg\alpha_1 \vee \dots \vee \neg\alpha_n$ and by writing α_0 for β in case β is prime, and $\beta = \neg\alpha_0$ otherwise, basic Horn formulas are up to logical equivalence of the type

$$\text{I: } \alpha_0 \vee \neg\alpha_1 \vee \dots \vee \neg\alpha_n \quad \text{or} \quad \text{II: } \neg\alpha_0 \vee \neg\alpha_1 \vee \dots \vee \neg\alpha_n$$

for prime formulas $\alpha_0, \dots, \alpha_n$. I and II are disjunctions of literals of which at most one is a prime formula. Basic Horn formulas are often defined in this way; but our definition above has pleasant advantages in inductive proofs as we shall see. Basic Horn formulas of type I are called *positive* and those of type II *negative*.

Each Horn formula is equivalent to a prenex Horn formula. If its prefix contains only $\forall$-quantifiers, then the formula is called a *universal Horn formula*. If the kernel of a Horn formula φ in prenex form is a conjunction of positive basic Horn formulas, φ is termed a *positive* Horn formula. A *propositional Horn formula*, i.e., a conjunction of propositional basic Horn formulas, can always be conceived of as a CNF whose disjunctions contain at most one nonnegated element. It is possible to think of an open Horn formula of $\mathcal{L}$ as resulting from replacing the propositional variables of some suitable propositional Horn formula by prime formulas of $\mathcal{L}$.

Example 3. (a) Identities and quasi-identities are universal Horn sentences, as are transitivity $(x \leqslant y \wedge y \leqslant z \to x \leqslant z)^{\mathfrak{G}}$, reflexivity $(x \leqslant x)^{\mathfrak{G}}$, and irreflexivity $(x \not\leqslant x)^{\mathfrak{G}}$, but not connexity $(x \leqslant y \vee y \leqslant x)^{\mathfrak{G}}$. The following congruence conditions for $=$ (where $\vec{x} = \vec{y}$ abbreviates $\bigwedge_{i=1}^{n} x_i = y_i$) are once again Horn sentences:

(4) $(x = x)^{\mathfrak{G}}$, $(x = y \wedge x = z \to y = z)^{\mathfrak{G}}$, $(\vec{x} = \vec{y} \to r\vec{x} \to r\vec{y})^{\mathfrak{G}}$, $(\vec{x} = \vec{y} \to f\vec{x} = f\vec{y})^{\mathfrak{G}}$.

(b) $\forall x \exists y\, x \circ y = e$ is a Horn sentence. Therefore, e.g., the theory of divisible abelian groups in $\mathcal{L}\{\circ, e\}$ is a *Horn theory*, which in the general case is to mean a theory

possessing an axiom system of Horn sentences. $\alpha := \forall x \exists y (x \neq 0 \to x \cdot y = 1)$, on the other hand, is not a Horn sentence and even not equivalent to a Horn sentence in the theory of fields, T_F. Otherwise $\mathrm{Md}\, T_F$ would be closed under direct products, Exercise 1. This is not the case: $\mathbb{Q} \times \mathbb{Q}$ has zero divisors, for example $(1,0)\cdot(0,1) = 0$. Hence, $\mathbb{Q} \times \mathbb{Q}$ is not a field.

Theorem 1.3. *Let U be a consistent set of universal Horn formulas. Then $\mathcal{F} := \mathcal{F}U$ is a model for U. In the case $U \subseteq \mathcal{L}^k$, $\mathcal{F}_k := \mathcal{F}_k U$ is a model for U as well.*

Proof. $\mathcal{F} \vDash U$ follows from $(*)\colon\ U \vdash \alpha \ \Rightarrow\ \mathcal{F} \vDash \alpha$, for all Horn formulas α. This is proved inductively on α. For prime formulas π, $(*)$ is clear, for then (3) reads as $\binom{*}{*}\colon\ U \vdash \pi \Leftrightarrow \mathcal{F} \vDash \pi$. Let $U \vdash \neg\pi$. Then $U \nvdash \pi$, for U is consistent. Hence $\mathcal{F} \nvDash \pi$ by $\binom{*}{*}$, and so $\mathcal{F} \vDash \neg\pi$. This confirms $(*)$ for all literals. Now let α be prime, β a basic Horn formula, $U \vdash \alpha \to \beta$, and assume $\mathcal{F} \vDash \alpha$. Then $U \vdash \alpha$, hence $U \vdash \beta$ and so $\mathcal{F} \vDash \beta$ by the induction hypothesis. This proves $\mathcal{F} \vDash \alpha \to \beta$. Induction on $\wedge$ is clear. Finally suppose $U \vdash \forall \vec{x} \alpha$ for some open Horn formula α, and let $\vec{t} \in \mathcal{T}_n$. Since then certainly $U \vdash \alpha \frac{\vec{t}}{\vec{x}}$, we get $\mathcal{F} \vDash \alpha \frac{\vec{t}}{\vec{x}}$ by the induction hypothesis. $\vec{t}$ was arbitrary, hence $\mathcal{F} \vDash \forall \vec{x} \alpha$ by (2). Thus $(*)$ is proved. The case $U \subseteq \mathcal{L}^k$ runs analogously by considering (2_k), (3_k) and taking $\mathcal{F}_k$ for $\mathcal{F}$. ❑

Incidentally, U's consistency in the theorem is always secured if U consists of positive Horn formulas; Exercise 2. Let U in Theorem 1.3 now be the axiom system of a *universal Horn theory* T, that is, U consists of universal Horn sentences. The theorem then yields $\mathcal{F}U \vDash T$. Since clearly $U \subseteq \mathcal{L}^k$ for each k, we likewise get $\mathcal{F}_k U \vDash T$. For more information on these particular models see Remark 2 below.

Example 4. The theory T in Example 1 is a particularly simple universal Horn theory. $\mathcal{F}T$ (more precisely, its underlaying algebra $\mathfrak{F} = \mathcal{T}/{\approx_T}$) was shown to be isomorphic to the word-semigroup on the alphabet *Var*. The semigroup $\mathfrak{F}$ is also called *the free semigroup with the free generators* $\boldsymbol{v}_0, \boldsymbol{v}_0 \ldots$ Now $\mathfrak{F} \vDash T$ follows from Theorem 1.3 without calculation. Similarly, the free semigroup with a finite number $k > 0$ of free generators is constructed by considering $\mathcal{F}_k T$. Its underlaying algebra is isomorphic to the word-semigroup on a k-element alphabet.

Remark 2. A universal Horn theory T like the one in Example 4 is said to be *nontrivial* if $\nvdash_T \forall xy\, x = y$. The generators $\overline{\boldsymbol{v}}_0, \overline{\boldsymbol{v}}_1, \ldots$ of $\mathcal{F}U$ are then distinct and $\mathcal{F}U$ is called the *free model of T with the free generators* $\overline{\boldsymbol{v}}_i$. Similarly, $\mathcal{F}_k U$ is the *free model of T with the free generators* $\overline{\boldsymbol{v}}_i$ for $i < k$. The word "free" comes from the fact that to generate a homomorphism, one can make "free use" of the values of the free generators. Free models in this sense exist only for nontrivial universal Horn theories.

Let U be as in Theorem 1.3 but $=$-free and let T be axiomatized by U. Clearly, $\mathcal{F}_0 U$ is defined only if $\mathcal{L}$ contains constant symbols; then $\mathcal{F}_0 U$ is a Herbrand model for T, called the *free* or *minimal Herbrand model for T*. It will henceforth be denoted

by $\mathcal{C}_U$ or $\mathcal{C}_T$. The domain of $\mathcal{C}_U$ is the set of ground terms. A not too simple an example for the not always easy task of identifying the minimal Herbrand model for a set U of $=$-free universal Horn sentences is the following one:

Example 5. Let U and $\mathcal{N}$ be as in Example 2. Both (a) and (b) are universal Horn sentences. We determine precisely the minimal Herbrand model $\mathcal{C}_U$ (whose domain consists of the terms $\underline{n}$) by proving $\mathcal{N} \simeq \mathcal{C}_U$, with the isomorphism $n \mapsto \underline{n}$. Since $\mathcal{C}_U \vDash \underline{m} < \underline{k} \Leftrightarrow U \vdash \underline{m} < \underline{k}$ by the definition page 106, it suffices to prove $(*)$: $m < k \Leftrightarrow U \vdash \underline{m} < \underline{k}$. The direction $\Rightarrow$ is shown by induction on k, beginning with $k = \mathtt{S}m$. The induction initiation is clear since $U \vdash \underline{m} < \underline{\mathtt{S}m}$ by (a). Let $m < \mathtt{S}k$, so that $m < k$ or $m = k$ and so $U \vdash \underline{m} < \underline{k}$ by the induction hypothesis, or $m = k$. In both cases, $U \vdash \underline{m} < \underline{\mathtt{S}k}$ by (a) and (b). The direction $\Leftarrow$ is obvious since $\mathcal{N} \vDash U$. This proves $(*)$. Note that U has many models on its Herbrand universe. $<$ may be realized by any transitive relation on $\mathbb{N}$ that extends $<$, for instance by $\leqslant^{\mathbb{N}}$. This interpretation is excluded by adding the Horn sentence $\forall x\, x \not< x$ to U, but the minimal Herbrand model remains the same for this expansion of U.

Most useful for logic programming is the following variant of Herbrand's theorem. The main difference is that in case $U \vdash \exists \vec{x} \gamma$ we get a single solution $\gamma \frac{\vec{t}}{\vec{x}}$. Theorem 1.4 does also hold with the same proof if the k is dropped throughout.

Theorem 1.4. *Let $U \subseteq \mathcal{L}^k$ $(k \geqslant 0)$ be a consistent set of universal Horn formulas, $\gamma = \gamma_0 \wedge \cdots \wedge \gamma_m$ where all γ_i are prime, and $\exists \vec{x} \gamma \in \mathcal{L}^k$. Then are equivalent*

(i) $\mathcal{F}_k U \vDash \exists \vec{x} \gamma$, (ii) $U \vdash \gamma \frac{\vec{t}}{\vec{x}}$ *for some* $\vec{t} \in \mathcal{T}_k^n$, (iii) $U \vdash \exists \vec{x} \gamma$.

In particular, for a consistent universal Horn theory T of any $=$-free language with constants, $\mathcal{C}_T \vDash \exists \vec{x} \gamma$ is always equivalent to $\vdash_T \exists \vec{x} \gamma$.

Proof. (i)$\Rightarrow$(ii): Let $\mathcal{F}_k U \vDash \exists \vec{x} \gamma$. Then $\mathcal{F}_k U \vDash \gamma \frac{\vec{t}}{\vec{x}}$ for some $\vec{t}$, because $\mathcal{F}_k U \vDash \neg \gamma \frac{\vec{t}}{\vec{x}}$ for all $\vec{t}$ implies $\mathcal{F}_k U \vDash \forall \vec{x} \neg \gamma$ by (2_k), contradicting (i). Thus, $\mathcal{F}_k U \vDash \gamma_i \frac{\vec{t}}{\vec{x}}$ for all $i \leqslant m$. Therefore $U \vdash \gamma_i \frac{\vec{t}}{\vec{x}}$ by (3_k), and so $U \vdash \gamma \frac{\vec{t}}{\vec{x}}$. (ii)$\Rightarrow$(iii): Trivial. (iii)$\Rightarrow$(i): Theorem 1.3 states that $\mathcal{F}_k U \vDash U$. Hence (iii) implies $\mathcal{F}_k U \vDash \exists \vec{x} \gamma$. The particular case follows from (i)$\Leftrightarrow$(iii) when choosing $k = 0$. Observe that $\mathcal{C}_T = \mathcal{F}_0 T$. ❑

Exercises

1. Show that $\mathrm{Md}\, T$ for a Horn theory T is closed under direct products and, if T is a universal Horn theory, then also under substructures. The former means that $(\forall i \in I) \mathcal{A}_i \vDash T \;\Rightarrow\; \mathcal{B} := \prod_{i \in I} \mathcal{A}_i \vDash T$, the latter $\mathcal{A}' \subseteq \mathcal{A} \vDash T \;\Rightarrow\; \mathcal{A}' \vDash T$.

2. Prove that a set of positive Horn formulas is always consistent.

3. Prove $\mathcal{C}_U \simeq (\mathbb{N}, 0, \mathtt{S}, \leqslant)$ for the set of $=$-free universal Horn sentences
$$U = \{\forall x\, x \leqslant x,\ \forall x\, x \leqslant \mathtt{S}x,\ \forall x \forall y \forall z (x \leqslant y \wedge y \leqslant z \to x \leqslant z)\}.$$

4.2 Propositional Resolution

We recall the problem of quickly deciding the satisfiability of propositional formulas. This problem is of eminent practical importance, since many nonnumerical (sometimes called "logical") problems can be reduced to this. The truth table method, practical for formulas with few variables, grows in terms of calculation effort exponentially with the number of variables; even the most powerful computers of the forseeable future will not be able to carry out the method for formulas with just 100 variables. Unfortunately, no better procedure is known, unless one is dealing with formulas of a particular form, for instance with certain normal forms. The general case represents an unsolved problem of theoretical computer science, not discussed here, the so-called P=NP problem; see for instance [GJ].

For conjunctive normal forms, the best procedure for contemporary computers is the *resolution procedure* introduced in the following. For the sake of a sparing presentation one switches from a disjunction $\lambda_1 \vee \cdots \vee \lambda_n$ of literals λ_i to the set $\{\lambda_1, \ldots, \lambda_n\}$. In so doing, the order of the disjuncts and their possible repetition, unessential factors for questions of satisfiability, are eliminated.

A finite, possibly empty set of literals is called a (propositional) *clause*. By a clause *in* $p_1, \ldots, p_n$ is meant a clause K with $var\, K \subseteq \{p_1, \ldots, p_n\}$. In the following K, H, G, L, P, N denote clauses, $\mathcal{K}, \mathcal{H}, \mathcal{P}, \mathcal{N}$ sets of clauses. $K = \{\lambda_1, \ldots, \lambda_n\}$ corresponds to the formula $\lambda_1 \vee \cdots \vee \lambda_n$. The *empty clause* (i.e., $n = 0$) is denoted by $\square$. It corresponds to the empty disjunction, which is identified with the falsum $\bot$. For $m > 0$, $K = \{q_1, \ldots, q_m, \neg r_1, \ldots, \neg r_k\}$ where $q_i, r_j \in PV$ is called a *positive* clause, for $m = 1$ also *definite*, and for $m = 0$ a *negative* clause. These conventions will also be adopted when the λ_i later denote literals of a first-order language.

Write $w \vDash K$ (a propositional valuation w *satisfies the clause* K) if K contains some λ with $w \vDash \lambda$. K is termed *satisfiable* if there is some w with $w \vDash K$. Note that the empty clause $\square$, as the definition's wording suggests, is not satisfiable.

w is a *model* for a set $\mathcal{K}$ of clauses, if $w \vDash K$ for all $K \in \mathcal{K}$. If $\mathcal{K}$ has a model then $\mathcal{K}$ is called *satisfiable*. In contrast to the empty clause $\square$, the empty set of clauses is satisfied by every valuation, again by the definition's wording.

w satisfies a CNF α iff w satisfies all its conjuncts, and hence all of the clauses corresponding to these conjuncts. Since every propositional formula can be transformed into a CNF, α is satisfiably equivalent to a corresponding finite set of clauses. For instance, the CNF $(p \vee q) \wedge (\neg p \vee q \vee r) \wedge (q \vee \neg r) \wedge (\neg q \vee s) \wedge \neg s$ is satisfiably equivalent to the corresponding set $\{\{p, q\}, \{\neg p, q, r\}, \{q, \neg r\}, \{\neg q, s\}, \{\neg s\}\}$ of clauses. It will turn out later that this set is not satisfiable.

We write $\mathcal{K} \vDash H$ if every model of $\mathcal{K}$ also satisfies the clause H. A set of clauses $\mathcal{K}$ is accordingly unsatisfiable if and only if $\mathcal{K} \vDash \square$.

For $\lambda \notin K$ we will frequently denote the clause $K \cup \{\lambda\}$ by K, λ. Moreover, let $\bar{\lambda} = \neg p$ for $\lambda = p$, $\bar{\lambda} = p$ for $\lambda = \neg p$ (so that always $\bar{\bar{\lambda}} = \lambda$), and $\bar{K} = \{\bar{\lambda} \mid \lambda \in K\}$.

The *resolution calculus* operates with sets of clauses and individual clauses, and has a single rule working with these objects, the so-called *resolution rule*

$$\text{RR:} \quad \frac{K, \lambda \,\big|\, L, \bar{\lambda}}{K \cup L} \quad (\lambda, \bar{\lambda} \notin K \cup L).$$

The clause $K \cup L$ is also called a *resolvent* of the clauses K, λ and $L, \bar{\lambda}$. The restriction $(\lambda, \bar{\lambda} \notin K \cup L)$ is actually not important, and can be neglected.

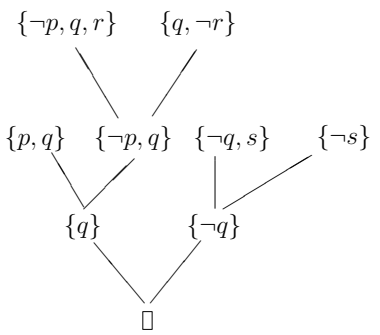
A clause H is called *derivable* from a set of clauses $\mathcal{K}$, in symbols $\mathcal{K} \vdash^{RR} H$, if H can be obtained from $\mathcal{K}$ by the stepwise application of RR; equivalently, if H belongs to the *resolution closure* $Rc\mathcal{K}$ of $\mathcal{K}$, which is the smallest set of clauses $\mathcal{H} \supseteq \mathcal{K}$ closed with respect to applications of RR. This definition corresponds completely to that of an MP-closed set of formulas in **1.6**.

Example. Let $\mathcal{K} = \{\{p, \neg q\}, \{q, \neg p\}\}$. Application of RR leads to the two resolvents $\{p, \neg p\}$ and $\{q, \neg q\}$, from which we see that a clause pair in general has several resolvents. Every subsequent application of RR yields already available clauses, so that $Rc\mathcal{K}$ contains only the clauses $\{p, \neg q\}, \{q, \neg p\}, \{p, \neg p\}, \{q, \neg q\}$.

Applying RR to $\{p\}, \{\neg p\}$ gives the empty clause $\square$. Hence $\mathcal{K} \vdash^{RR} \square$, with the unsatifiable set of clauses $\mathcal{K} = \{\{p\}, \{\neg p\}\}$. By the resolution theorem below, the derivability of the empty clause from a set of clauses $\mathcal{K}$ is characteristic of the nonsatisfiability of $\mathcal{K}$. To test this one needs only to check whether $\mathcal{K} \vdash^{RR} \square$, or $\square \in Rc\mathcal{K}$. This is effectively decidable for finite sets $\mathcal{K}$ because $Rc\mathcal{K}$ is finite. Indeed, a resolvent that results from applying RR to clauses in $p_1, \ldots, p_n$ contains at most these very same variables. Further, it is clear that there exist only finitely many clauses in $p_1, \ldots, p_n$, namely exactly 2^{2n}. But that is still an exponential increase as n increases. And aside from this the mechanical implementation of the resolution calculus mostly involves potentially infinite sets of predicate-logical clauses. We consider this problem more closely at the end of **4.4**.

The derivation of a clause H from a set of clauses $\mathcal{K}$, especially the derivation of the empty clause, can best be graphically represented by a so-called *resolution tree*. This is a tree which branches "from above" with an endpoint H without edge exits, also called the *root* of the tree. Points without entering edges are called *leaves*. A point that is not a leaf has two entrances, and the points leading to them are called their *predecessors*. The points of a tree bear sets of clauses in the sense that a point which is not a leaf is a resolvent of the two clauses above it. The following figure shows one of the many resolution trees for the already-mentioned set of clauses

$$\mathcal{K}_0 = \{\{p, q\}, \{\neg p, q, r\}, \{q, \neg r\}, \{\neg q, s\}, \{\neg s\}\}.$$



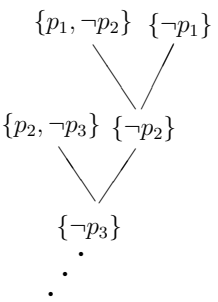
The leaves of this tree are all occupied by clauses in $\mathcal{K}_0$. It should be clear that an arbitrary clause H belongs to the resolution closure of a set of clauses $\mathcal{K}$ just when there exists a resolution tree with leaves in $\mathcal{K}$ and root H. A resolution tree with leaves in $\mathcal{K}$ and the root $\square$ as in the figure on the left for $\mathcal{K} = \mathcal{K}_0$ is called a *resolution for* $\mathcal{K}$, or more exactly a *successful resolution for* $\mathcal{K}$. By the aforementioned, $\mathcal{K}_0$ is unsatisfiable, and hence so is the conjunctive normal form that corresponds to the set of clauses $\mathcal{K}_0$, namely $(p \vee q) \wedge (\neg p \vee q \vee r) \wedge (q \vee \neg r) \wedge (\neg q \vee s) \wedge \neg s$.

Remark 1. If a resolution tree ends with a point $\neq \square$, to which RR either cannot be applied or where upon application the points are simply reproduced, then one talks of an *unsuccessful* resolution. In this case, most interpreters of the resolution calculus will "backtrack," which means the program searches backwards along the tree for the first point where one of several resolution alternatives was chosen, and picks up another alternative. Some kind of selection strategy must in any case be implemented, since just as with any logical calculus, the resolution calculus is nondeterministic, that is, no natural preferences exist regarding the order of the derivations leading to a successful resolution, even if the existence of such a resolution is known for other reasons.

We remark that despite the derivability of the empty clause, for infinite unsatisfiable sets of clauses $\mathcal{K}$ there also exist infinite resolution trees with nonrepeating points where $\square$ never appears. Such trees do not have a root. For example, the set of clauses

$$\mathcal{K} = \{\{p_1\}, \{\neg p_1\}, \{p_1, \neg p_2\}, \{p_2, \neg p_3\}, \dots\}$$

is not satisfiable. Here we obtain the infinite resolution tree in the figure on the right, occupied by leaves from $\mathcal{K}$, which has no root and does not reflect the fact that $\square$ can be derived just by a single application of RR to the first two clauses of $\mathcal{K}$. In this example the resolution calculus runs on $\mathcal{K}$ with a completely stupid strategy.



This and similar examples indicate that the resolution calculus is incapable in general of deciding the satisfiability of infinite sets $\mathcal{K}$ of clauses. Indeed, this will be confirmed in **4.4**. Nonetheless, by Theorem 2.2 below there does exist—if $\mathcal{K}$ is in actual fact unsatisfiable—a successful resolution for $\mathcal{K}$ that can in principle be found in finitely many steps.

We commence the more detailed study of the resolution calculus with

Lemma 2.1 (Soundness lemma). $\mathcal{K} \vdash^{RR} H \;\Rightarrow\; \mathcal{K} \models H$.

Proof. As in the case of a Hilbert calculus, it suffices to confirm the soundness of the rule RR, that is, to prove that a model for K, λ and $L, \bar{\lambda}$ is also one for $K \cup L$. Thus let $w \models K, \lambda$ and $w \models L, \bar{\lambda}$. **Case 1:** $w \not\models \lambda$. Then there must be a literal $\lambda' \in K$ with $w \models \lambda'$. Hence $w \models K$ and therefore $w \models K \cup L$. **Case 2:** $w \models \lambda$. Then $w \not\models \bar{\lambda}$. Similar to the above we get $w \models L$. Hence $w \models K \cup L$ as well. ❏

For the case $\mathcal{K} \vdash^{RR} \square$ the lemma shows $\mathcal{K} \models \square$, that is, the unsatisfiability of $\mathcal{K}$. The converse of Lemma 2.1 is in general not valid; for instance $\{\{p\}\} \models \{p, q\}$, but $\{\{p\}\} \nvdash^{RR} \{p, q\}$. It does hold, though, for $H = \square$. This follows from Theorem 2.2, also often stated as "$\mathcal{K}$ *is unsatisfiable iff* $\mathcal{K} \vdash^{RR} \square$." In its proof we construct a global valuation w from partial valuations, defined only for $p_1, \ldots, p_n$.

Theorem 2.2 (Resolution theorem). $\mathcal{K}$ *is satisfiable if and only if* $\mathcal{K} \nvdash^{RR} \square$.

Proof. For satisfiable $\mathcal{K}$ we have $\mathcal{K} \not\models \square$, so $\mathcal{K} \nvdash^{RR} \square$ by Lemma 2.1. Now let $\mathcal{K} \nvdash^{RR} \square$, or equivalently, $\square \notin \mathcal{H}$ where $\mathcal{H} := Rc\mathcal{K}$. We will construct a model w for $\mathcal{H}$ and hence for $\mathcal{K}$ stepwise, i.e., the values $v_n = wp_n$ will be defined inductively on n. Let $\Lambda^{(n)}$ be the set of all literals in $p_1, \ldots, p_n$, and let $\mathcal{H}^{(n)}$ be the set of all $K \in \mathcal{H}$ with $K \subseteq \Lambda^{(n)}$ such that p_n or $\neg p_n$ or both belong to K. Clearly, $\Lambda^{(0)} = \mathcal{H}^{(0)} = \emptyset$, because variable enumeration starts with p_1. Note that $var\,\mathcal{H}^{(n)} \subseteq \{p_1, \ldots, p_n\}$ and $\mathcal{H} = \bigcup_{n \in \mathbb{N}} \mathcal{H}^{(n)}$. Let $v_1, \ldots, v_n$ already be defined so that $w_n := (v_1, \ldots, v_n) \models \mathcal{H}^{(i)}$ for all $i \leqslant n$. This assumption holds trivially for $n = 0$ if we agree to say that the "empty valuation" satisfies $\mathcal{H}^{(0)} = \emptyset$. Now $v_{n+1} = wp_{n+1}$ will be defined such that

$(*)$ $\quad w_{n+1} := (v_1, \ldots, v_{n+1}) \models \mathcal{H}^{(n+1)}$ $\quad$ (induction claim).

We need to care only about those $K \in \mathcal{H}^{(n+1)}$ containing not both p_{n+1} and $\neg p_{n+1}$, and no literal $\lambda \in \Lambda^{(n)}$ with $w_n \models \lambda$, called *sensitive* clauses during this proof, since all other (insensitive) $H \in \mathcal{H}^{(n+1)}$ are satisfied by *any* expansion of w_n to w_{n+1}.[1]

Claim: either $p_{n+1} \in K$ for *all* sensitive K—then put $v_{n+1} = 1$—or else $\neg p_{n+1} \in K$ for all sensitive K, in which case put $v_{n+1} = 0$, so that $(*)$ holds in either case. To prove the claim assume that there are sensitive K, H with $p_{n+1} \in K$ and $\neg p_{n+1} \in H$ (hence $\neg p_{n+1} \notin K$, $p_{n+1} \notin H$). Then, applying RR to H, K, we obtain either $\square$ (contradicting $\square \notin \mathcal{H}$), or else a clause from $\mathcal{H}^{(i)}$ for some $i \leqslant n$ whose literals are not satisfied by w_n, a contradiction to $w_n \models \mathcal{H}^{(i)}$. This confirms the claim. Thus, $w_n \models \mathcal{H}^{(n)}$ for all n, so that $w = (v_1, v_2, \ldots)$ is a model for the whole of $\mathcal{H}$. ❏

Remark 2. The foregoing proof is constructive, that is, if $\mathcal{K} \nvdash^{RR} \square$ and the $\mathcal{H}^{(n)}$ in the proof above are computable, then a valuation satisfying $\mathcal{K}$ is computable as well. Moreover, we incidentally proved the propositional compactness theorem for countable sets of formulas

[1] The newcomer should write down all eight candidates for $\mathcal{H}^{(1)} \subseteq \{\{p_1\}, \{\neg p_1\}, \{p_1, \neg p_1\}\}$. Only $\{p_1\}$ and $\{\neg p_1\}$ are sensitive to $v_1 = wp_1$. At most one of these two clauses can belong to $\mathcal{H}^{(1)}$.

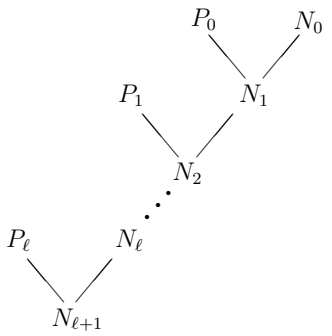
X once again. Here is the argument: every formula is equivalent to some KNF, and hence X is satisfiably equivalent to a set of clauses $\mathcal{K}_X$. So if X is not satisfiable, the same is true of $\mathcal{K}_X$. Consequently, $\mathcal{K}_X \vdash^{RR} \square$ by Theorem 2.2. Therefore $\mathcal{K}_0 \vdash^{RR} \square$ for some finite subset $\mathcal{K}_0 \subseteq \mathcal{K}_X$, for there must be some successful resolution tree whose leaves are collected in $\mathcal{K}_0$. Having this it is obvious that just a finite subset of X is not satisfiable, namely the one that corresponds to the set of clauses $\mathcal{K}_0$.

A clause belonging to a propositional basic Horn formula is called a (propositional) *Horn clause*. It is called positive or negative if the corresponding Horn formula is positive or negative. Positive Horn clauses are of the form $\{\neg q_1, \ldots, \neg q_n, p\}$ where $n \geqslant 0$, negative of the form $\{\neg q_1, \ldots, \neg q_k\}$. The empty clause ($k = 0$) is also counted among the negative ones. The affix *propositional* is omitted as long as we remain within propositional logic.

It is important in practice that the resolution calculus can be formulated even more specifically for Horn clauses. The empty clause, if it can be obtained from a set of Horn clauses at all, can also be obtained using a restricted resolution rule, which is applied only to pairs of Horn clauses where one component is positive and the other negative. This is the *rule of Horn resolution*

$$\text{HR}: \qquad \frac{K, p \mid L, \neg p}{K \cup L} \qquad (K, L \text{ negative},\ p, \neg p \notin K \cup L).$$

A positive Horn clause is clearly definite. Hence, the resolvent of an application of HR is uniquely determined and always negative. An H-Resolution tree is therefore of the simple form illustrated by the figure on the left. There $P_0, \ldots, P_\ell$ denote positive and $N_0, \ldots, N_{\ell+1}$ negative Horn clauses. Such a tree is called an *H-resolution for $\mathcal{P}, N$* (where $\mathcal{P}$ here and everywhere is taken to mean a set of positive Horn clauses and N a negative clause $\neq \square$) if it satisfies the conditions (1) $P_i \in \mathcal{P}$ for all $i \leqslant \ell$, and (2) $N_0 = N$ & $N_{\ell+1} = \square$. It is evidently also possible to regard an H-resolution for $\mathcal{P}, N$ as a sequence $(P_i, N_i)_{i \leqslant \ell}$ with the properties (0) $N_{i+1} = HR(P_i, N_i)$ for all $i \leqslant \ell$, (1), and (2). Here $HR(P, N)$ denotes the uniquely determined resolvent resulting from applying HR to the positive clause P and the negative clause N.



The calculus operating with Horn clauses and rule HR is denoted by $\vdash^{HR}$. Before proving its completeness we require a little preparation. Let $\mathcal{P}$ be a set of positive Horn clauses. In order to gain an overview of all models w of $\mathcal{P}$, consider the natural correspondence $w \longleftrightarrow V_w := \{p \in PV \mid w \vDash p\}$ between valuations w and subsets of PV. Let $w \leqslant w' :\Leftrightarrow\ V_w \subseteq V_{w'}$. Clearly, $\mathcal{P}$ is always satisfied by the "maximal" valuation w with $V_w = PV$ (i.e., $wp = 1$ for all $p \in PV$). It is obvious that $w \vDash \mathcal{P}$ if

and only if $V = V_w$ satisfies the following two conditions:

(a) $p \in V$ provided $\{p\} \in \mathcal{P}$,

(b) $q_1, \ldots, q_n \in V \Rightarrow p \in V$, whenever $n > 0$ and $\{\neg q_1, \ldots, \neg q_n, p\} \in \mathcal{P}$.

Of all subsets $V \subseteq PV$ satisfying (a) and (b) there is obviously a smallest one, namely $V_{\mathcal{P}} := \bigcap\{V_w \mid w \vDash \mathcal{P}\}$. The $\mathcal{P}$-model corresponding to $V_{\mathcal{P}}$ is denoted by $w_{\mathcal{P}}$ and called the *minimal* $\mathcal{P}$*-model*. We may define $V_{\mathcal{P}}$ also as follows: Let $V_0 = \{p \in PV \mid \{p\} \in \mathcal{P}\}$ and $V_{k+1} = V_k \cup \big\{p \in PV \mid \{\neg q_1, \ldots, \neg q_n, p\} \in \mathcal{P}$ for some $q_1, \ldots, q_n \in V_k\big\}$. Then $V_{\mathcal{P}} = \bigcup_{k\in\mathbb{N}} V_k$. Indeed, $V_k \subseteq V_w$ for all k and all $w \vDash \mathcal{P}$. Hence, $\bigcup_{k\in\mathbb{N}} V_k \subseteq V_{\mathcal{P}}$. Also $V_{\mathcal{P}} \subseteq \bigcup_{k\in\mathbb{N}} V_k$ holds, because $w \vDash \mathcal{P}$ provided $V_w = \bigcup_{k\in\mathbb{N}} V_k$.

The minimal m with $p \in V_m$ is termed the $\mathcal{P}$*-rank* of p, denoted by $\rho_{\mathcal{P}} p$. Those p with $\{p\} \in \mathcal{P}$ are of $\mathcal{P}$-rank 0. The variables arising from these by applying (b) have $\mathcal{P}$-rank 1 if not already in V_0, and so on.

Lemma 2.3. *Let $\mathcal{P}$ be a set of positive Horn clauses and $q_0, \ldots, q_k \in V_{\mathcal{P}}$. Then holds $\mathcal{P}, N \vdash^{HR} \square$, where $N = \{\neg q_0, \ldots, \neg q_k\}$.*

Proof. For variables $r_0, \ldots, r_n \in V_{\mathcal{P}}$ set $\rho_{\mathcal{P}}(r_0, \ldots, r_n) := \max\{\rho_{\mathcal{P}} r_0, \ldots, \rho_{\mathcal{P}} r_n\}$. Let $\mu(r_0, \ldots, r_n)$ be the number of $i \leqslant n$ such that $\rho_{\mathcal{P}} r_i = \rho_{\mathcal{P}}(r_0, \ldots, r_n)$. The claim is proved inductively on $\rho := \rho_{\mathcal{P}}(q_0, \ldots, q_k)$ and $\mu := \mu(q_0, \ldots, q_k)$. First suppose $\rho = 0$, i.e., $\{q_0\}, \ldots, \{q_k\} \in \mathcal{P}$. Then there certainly exists an H-resolution for $\mathcal{P}, N$, namely $(\{q_i\}, \{\neg q_i, \ldots, \neg q_k\})_{i\leqslant k}$. Now take $\rho > 0$ and w.l.o.g. $\rho = \rho_{\mathcal{P}} q_0$. Then there exist $q_{k+1}, \ldots, q_m \in V_{\mathcal{P}}$ such that $P := \{\neg q_{k+1}, \ldots, \neg q_m, q_0\} \in \mathcal{P}$ and $\rho_{\mathcal{P}}(q_{k+1}, \ldots, q_m) < \rho$. Thus, $\rho_{\mathcal{P}}(q_1, \ldots, q_k, q_{k+1}, \ldots, q_m)$ is $< \rho$, or it is $= \rho$ so that $\mu(q_1, \ldots, q_m) < \mu$. By the induction hypothesis, in both cases $\mathcal{P}, N_1 \vdash^{HR} \square$ for $N_1 := \{\neg q_1, \ldots, \neg q_m\}$. Hence, an H-resolution $(P_i, N_i)_{1\leqslant i\leqslant \ell}$ for $\mathcal{P}, N_1$ exists. But then $(P_i, N_i)_{i\leqslant \ell}$ with $P_0 := P$ and $N_0 := N$ is just an H-resolution for $\mathcal{P}, N$. ❏

Theorem 2.4 (on Horn resolution). *A set $\mathcal{K}$ of Horn clauses is satisfiable if and only if $\mathcal{K} \nvdash^{HR} \square$.*

Proof. The condition $\mathcal{K} \nvdash^{HR} \square$ is certainly necessary if $\mathcal{K}$ is satisfiable. For the converse assume $\mathcal{K}$ is unsatisfiable, $\mathcal{K} = \mathcal{P} \cup \mathcal{N}$, all $P \in \mathcal{P}$ are positive, and all $N \in \mathcal{N}$ negative. Since $w_{\mathcal{P}} \vDash \mathcal{P}$ but $w_{\mathcal{P}} \nvDash \mathcal{P} \cup \mathcal{N}$ there is some $N = \{\neg q_0, \ldots, \neg q_k\} \in \mathcal{N}$ such that $w_{\mathcal{P}} \nvDash N$. Consequently, $w_{\mathcal{P}} \vDash q_0, \ldots, q_k$ and so $q_0, \ldots, q_k \in V_{\mathcal{P}}$. By Lemma 2.3 we then obtain $\mathcal{P}, N \vdash^{HR} \square$, and a fortiori $\mathcal{K} \vdash^{HR} \square$. ❏

Corollary 2.5. *Let $\mathcal{K} = \mathcal{P} \cup \mathcal{N}$ be a set of Horn clauses, all $P \in \mathcal{P}$ positive, and all $N \in \mathcal{N}$ negative. Then the following conditions are equivalent:*

(i) *$\mathcal{K}$ is unsatisfiable,* (ii) *$\mathcal{P}, N$ is unsatisfiable for some $N \in \mathcal{N}$.*

Proof. (i) implies $\mathcal{K} \vdash^{HR} \square$ by Theorem 2.4. Hence, there is some $N \in \mathcal{N}$ and some H-Resolution for $\mathcal{P}, N$, whence $\mathcal{P}, N$ is unsatisfiable. (ii)⇒(i) is trivial. ❏

Thus, the investigation of sets of Horn clauses as regard satisfiability can completely be reduced to the case of just a single negative clause.

The hitherto illustrated techniques can without further ado be carried over to quantifier-free formulas of a first-order language $\mathcal{L}$, in that one thinks of the propositional variables to be replaced by prime formulas of $\mathcal{L}$. Clauses are then finite sets of literals in $\mathcal{L}$. By Remark 1 in **4.1** a set of $\mathcal{L}$-formulas is satisfiably equivalent to a set of quantifier-free formulas, which w.l.o.g. are given in conjunctive normal form. Splitting into conjuncts provides a satisfiably equivalent set of disjunctions of literals. Converting these disjunctions into clauses, one obtains a set of clauses for which, by the remark just cited, a consistency condition can be stated propositionally. Now, because predicate-logical proofs are always reducible to the demonstration of certain inconsistencies by virtue of the equivalence of $X \vdash \alpha$ with the inconsistency of $X, \neg\alpha$, these proofs can basically also be carried out by resolution.

To sum up, resolution by Theorems 2.2 and 2.4 is not at all restricted to propositional logic but includes application to sets of literals of first-order languages. The predicate logic version of of Theorem 2.2, Theorem 5.3, will essentially be reduced to the former. Moreover, questions concerning predicate logic resolution can often directly be treated propositionally, as indicated by the exercises below.

Before elaborating on this, we consider an additional aid to automated proof procedures, namely unification. This will later be combined with resolution, and it is this combination that makes automated proof procedures fast enough for modern computers, equipped with efficient interpreters of PROLOG.

Exercises

1. Prove that the satisfiable set of clauses $\mathcal{P} = \{\{p_3\}, \{\neg p_3, p_1, p_2\}\}$ does not have a smallest model (the second clause in $\mathcal{P}$ is not a Horn clause).

2. Let $p_{m,n,k}$ for $m, n, k \in \mathbb{N}$ be propositional variables, $\mathtt{S}$ the successor function, and $\mathcal{P}$ the set of all clauses belonging to the following Horn formulas:

$$p_{m,0,m} \;\; ; \;\; p_{m,n,k} \to p_{m,\mathtt{S}n,\mathtt{S}k} \qquad (m, n, k \in \mathbb{N}).^{2}$$

   Let the standard model w_{St} be defined by $w_{St} \vDash p_{m,n,k} \;\Leftrightarrow\; m + n = k$. Show that the minimal model $w_{\mathcal{P}}$ coincides with w_{St}.

3. Let $\mathcal{P}$ be the set of Horn clauses of Exercise 2. Prove that

   (a) $\mathcal{P}, \neg p_{n,m,n+m} \vdash^{HR} \square$, (b) $\mathcal{P}, \neg p_{n,m,k} \vdash^{HR} \square \;\Rightarrow\; k = n + m$.

   (a) and (b) together can be stated as (c) $\mathcal{P}, \neg p_{n,m,k} \vdash^{HR} \square \;\Leftrightarrow\; k = n + m$.

[2] In **4.4** these formulas will be interpreted as the ground instances of a logic program for computing the sum of two natural numbers.

4.3 Unification

A decisive aid in logic programming is unification. This notion is meaningful for any set of formulas, but we confine ourself to ¬-*free* clauses $K \neq \square$ of an identity-free language. K contains only unnegated prime formulas, each starting with a relation symbol. Such a clause K is called *unifiable* if a substitution σ exists, a so-called *unifier* of K, such that $K^\sigma := \{\lambda^\sigma \mid \lambda \in K\}$ contains exactly one element; in other words, K^σ is a *singleton*. Here σ is the easiest be understood as a simultaneous substitution, that is, σ is globally defined and $x^\sigma = x$ for almost all variables x. Simultaneous substitutions form a semigroup with respect to composition, with the neutral element ι (see page 48), a fact we will heavily make use of.

Example 1. Consider $K = \{rxfxz, rfyzu\}$, r and f binary. Here $\omega = \frac{fyz}{x}\,\frac{ffyzz}{u}$ is a unifier: $K^\omega = \{rfyzffyzz\}$, as is readily confirmed. Clearly, ω as a composition of simple substitutions can also be understood as a global substitution.

Obviously, a clause containing prime formulas that start with distinct relation symbols is not unifiable. A further obstacle to unification is highlighted by

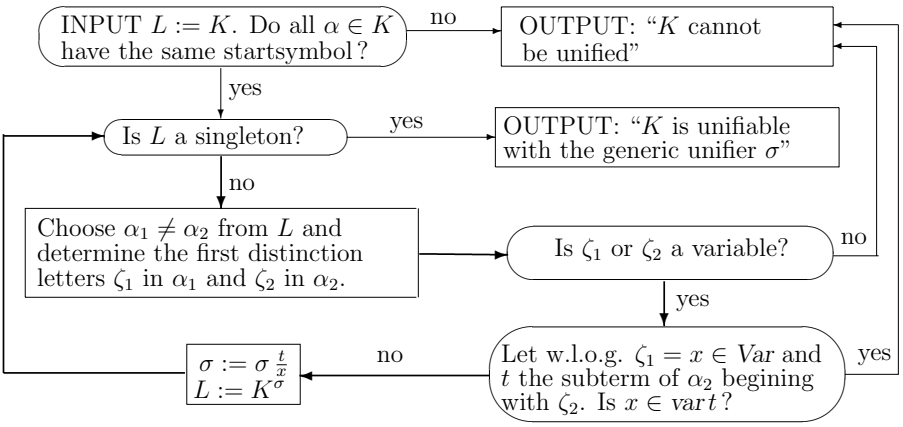
Example 2. Let $K = \{rx, rfx\}$ (r, f unary). Assume $(rx)^\sigma = (rfx)^\sigma$. This clearly implies $rx^\sigma = rfx^\sigma$ and hence $x^\sigma = fx^\sigma$, which is impossible, since no term can be a proper subterm of itself. Hence, K is not unifiable.

If σ is a unifier then so too is $\sigma\tau$ for any substitution τ. Call ω a *generic* or a *most general* unifier of K if any other unifier τ of K has a representation $\tau = \omega\sigma$ for some substitution σ. By Theorem 3.1 below, each unifiable clause has a generic unifier. For instance, it will turn out below that ω in Example 1 *is* generic.

A *renaming of variables*, a *renaming* for short, is for the sake of simplicity a substitution ρ such that $\rho^2 = \iota$. This definition could be rendered more generally, but it suffices for our purposes. ρ is necessarily bijective and maps variables to variables. If $x_i^\rho = y_i$ ($\neq x_i$) and hence $y_i^\rho = x_i$ for $i = 1, \ldots, n$, and $z^\rho = z$ otherwise, that is, if ρ swaps the variables x_i and y_i, we shall write $\rho = \binom{x_1 \cdots x_n}{y_1 \cdots y_n}$.

If ω is a generic unifier of K then so too is $\omega' = \omega\rho$, for any renaming ρ. Indeed, for any given unifier τ of K there is some σ such that $\tau = \omega\sigma$. For $\sigma' := \rho\sigma$ then $\tau = \omega\rho^2\sigma = (\omega\rho)(\rho\sigma) = \omega'\sigma'$. Choosing in Example 1 for instance $\rho := \binom{y\ z}{u\ v}$, we obtain the generic unifier $\omega' = \omega\rho$ for K, with $K^{\omega'} = \{rfuvffuvv\}$.

We now consider a procedure in the form of a flow diagram, the *unification algorithm*, denoted by $\mathfrak{U}$. It checks each nonempty clause K of prime formulas of an identity-free language for unifiability, and in the positive case it produces a generic unifier. $\mathfrak{U}$ uses a variable σ for substitutions and a variable L for clauses, with initial values ι and K, respectively. Later on, L contains K^σ for suitable σ, which depends on the state of the procedure.



The *first distinction letters* of two strings are the first symbols, read from the left, that distinguish the strings. The first letter of $\alpha \in L$ is a relation symbol. By Exercise 1 in **2.2**, any further symbol ζ in α determines uniquely at each position of its occurrence a subterm of α whose initial symbol is ζ. The diagram has just one (thick-lined) loop that starts and ends in the test "Is L a singleton?". It runs through the operation $\sigma := \sigma \frac{t}{x}$, $L := K^\sigma$ which first assigns a new value to σ and then to L. This reduces the number of variables in L since $x \notin var\,L$ because of $x \notin var\,t$. Therefore, $\mathfrak{U}$ stops in any case and halts in one of the two OUTPUT boxes of $\mathfrak{U}$. But we do not yet know whether $\mathfrak{U}$ always ends up in the "right" box, i.e., whether $\mathfrak{U}$ answers correctly. The final value σ is printed in the lower OUTPUT box. Let $\sigma_0 := \iota$ and σ_i for $i > 0$ the value of σ after the ith run through the loop.

Example 3. Let $\mathfrak{U}$ be executed on K from Example 1. The first distinction letters of the two members $\alpha_1, \alpha_2 \in K$ are $\zeta_1 = x$ and $\zeta_2 = f$ at the second position. The subterm beginning with ζ_2 in α_2 is $t = fyz$. Hence $\sigma_1 = \frac{fyz}{x}$, and after the first run through the loop with $\sigma := \sigma_1$, we have $K^\sigma = \{rfyzffyzz, rfyzu\}$. Here the first distinction letters are f and u. The subterm beginning with f (at position 5) is $t = ffyzz$. Since $u \notin var\,ffyzz$, the loop is run through once again and we obtain $\sigma_2 = \sigma_1 \frac{ffyzz}{u} = \frac{fyz}{x}\frac{ffyzz}{u}$. This is a unifier, and $\mathfrak{U}$ comes to a halt with OUTPUT "K is unifiable with the generic unifier $\frac{fyz}{x}\frac{ffyzz}{u}$" according to Theorem 3.1.

Theorem 3.1. *The unification algorithm $\mathfrak{U}$ is sound, i.e., upon input of a negation-free clause K it always answers correctly.*[3] *$\mathfrak{U}$ unifies with a generic unifier.*

Proof. This is obvious if two elements of K are already distinguished by the first letter. Assume therefore that all $\alpha \in K$ begin with the same letter. If $\mathfrak{U}$ stops

[3] The proof will be a paradigm for a correctness proof of an algorithm. It almost always has to be carried out inductively on the number of runs through a loop occurring in the algorithm.

with the output "K is unifiable...," K is in fact unifiable, since it must have been previously verified that $L = K^\sigma$ is a singleton. Converely, it has to be shown that $\mathfrak{U}$ also halts with the correct output provided K *is* unifiable. The latter will be our assumption till the end of the proof, along with the choice of an arbitrary unifier τ of K. Let i $(= 0, \ldots, m)$ denote the moment after the ith run through the loop has been finished. $i = 0$ before any run. $i = m$ after the last run is complete, in which $\mathfrak{U}$ gets again the question "Is L a singleton?". We will show inductively on i that

$(*)$ there exists a substitution τ_i with $\sigma_i\tau_i = \tau \quad (i = 0, \ldots, m)$.

This is trivial for $i = 0$: choose simply $\tau_0 = \tau$, so that $\sigma_0\tau_0 = \iota\tau = \tau$. Suppose $(*)$ holds for $i < m$. Then $K^{\sigma_i\tau_i} = K^\tau$ is a singleton, but K^{σ_i} still contains two distinct formulas α_1, α_2 with x at some position in α_1 and t a term in α_2 starting at the same position in α_2. From $\alpha_1^{\tau_i} = \alpha_2^{\tau_i}$ (note that $\alpha_1^{\tau_i}, \alpha_2^{\tau_i} \in K^{\sigma_i\tau_i} = K^\tau$ and K^τ is a singleton) we get $x^{\tau_i} = t^{\tau_i}$. Hence, $x \notin var\,t$, for otherwise x^{τ_i} would be a proper subterm of itself. Set $\tau_{i+1} := \frac{t}{x}\,\tau_i$. Then $\frac{t}{x}\,\tau_{i+1} = \tau_i$. Indeed, for $y \neq x$ we obtain $y^{\frac{t}{x}\,\tau_{i+1}} = y^{\tau_{i+1}} = y^{\frac{t}{x}\,\tau_i} = y^{\tau_i}$, but in view of $x^{\tau_i} = t^{\tau_i}$ we have also

$$x^{\frac{t}{x}\,\tau_{i+1}} = t^{\tau_{i+1}} = t^{\frac{t}{x}\,\tau_i} \quad = t^{\tau_i} \quad \text{(since } x \notin var\,t\text{)}$$
$$= x^{\tau_i}.$$

$\frac{t}{x}\,\tau_{i+1} = \tau_i$ and $\sigma_{i+1} = \sigma_i\,\frac{t}{x}$ yield the induction claim $\sigma_{i+1}\tau_{i+1} = \sigma_i\,\frac{t}{x}\,\tau_{i+1} = \sigma_i\tau_i = \tau$.

Next we show that $L = K^{\sigma_m}$ is a singleton. Assume that this is not the case and choose $\alpha_1, \alpha_2 \in L$ as in the diagram. Then the upper right test is answered "yes" since otherwise ζ_1, ζ_2 would be distinct function symbols or constants not removable by any substitution. This contradicts $\alpha_1^{\tau_m} = \alpha_2^{\tau_m}$. However, the lower right question is answered "no," because α_1 starts at the first distinction position with x, hence $\alpha_1^{\tau_m}$ with x^{τ_m}, and $\alpha_2^{\tau_m}$ $(= \alpha_1^{\tau_m})$ with t^{τ_m}. Therefore, $x^{\tau_m} = t^{\tau_m}$ which implies $x \notin var\,t$. Thus, the loop runs through once again which contradicts the definition of m; hence σ_m is indeed a unifier, that is, $\mathfrak{U}$ terminates correctly in the unfiable case as well. Moreover, σ_m is a generic unifier, because $\sigma_m\tau_m = \tau$ by $(*)$, with the arbitrarily chosen unifier τ. This completes the proof. ❏

Exercises

1. Show that for prime formulas α, β without shared variables are equivalent
   (i) $\{\alpha, \beta\}$ is unifiable, (ii) there are substitutions σ, τ such that $\alpha^\sigma = \beta^\tau$.

2. Show: $\sigma = \frac{\vec{t}}{\vec{x}}$ is *idempotent* (which is to mean $\sigma^2 = \sigma$) if and only if $x_i \notin var\,t_j$, for all i, j with $1 \leqslant i, j \leqslant n$.

3. For clauses K_0, K_1 we term ρ a *separator* of K_0, K_1 if ρ is a renaming such that $var\,K_0^\rho \cap var\,K_1 = \emptyset$. Let K_0, K_1 be negation-free. Show that if $K_0 \cup K_1$ is unifiable then so is $K_0^\rho \cup K_1$, but not conversely, in general.

4.4 Logic Programming

A rather general starting point in dealing with systems of artificial intelligence consists in using computers to draw consequences φ from certain data and facts given in the form of a set of formulas X, that is, proving $X \vdash \varphi$ mechanically. That this is possible in theory was the subject of **3.5**. In practice, however, such a project is in general realizable only under certain limitations regarding the pattern of the formulas in X and φ. These limitations refer to any first-order language $\mathcal{L}$ adapted to the needs of the particular problem. For logic programming the following restrictions are characteristic:

- $\mathcal{L}$ is identity-free and contains at least one constant symbol,
- each $\alpha \in X$ is a positive universal Horn sentence,
- φ is a sentence of the form $\exists\vec{x}(\gamma_0 \wedge \cdots \wedge \gamma_k)$ with prime formulas γ_i.

Note that $\neg\varphi$ is equivalent to $\forall\vec{x}(\neg\gamma_0 \vee \cdots \vee \neg\gamma_k)$ and hence a negative universal Horn sentence. Because $\forall$-quantifiers can be distributed among conjunctions, we may assume w.l.o.g. that each sentence $\alpha \in X$ is of the form

$$(*) \quad (\beta_1 \wedge \cdots \wedge \beta_m \to \beta)^{\mathtt{G}} \quad (\beta, \beta_1, \ldots, \beta_m \text{ prime formulas}, m \geqslant 0).$$

A finite set of sentences of this type is called a *logic program* and will henceforth be denoted by the letter $\mathcal{P}$. The availability of a constant symbol just ensures the existence of a Herbrand model for $\mathcal{P}$. In the programming language PROLOG, $(*)$ is written without quantifiers in the following way:

$$\beta :\!- \beta_1, \ldots, \beta_m \quad (\text{or just } \beta :\!- \text{ in case } m = 0).$$

$:\!-$ symbolizes converse implication mentioned in **1.1**. For $m = 0$ such program clauses are called *facts*, and for $m > 0$ *rules*. In the following we make no distinction between a logic program as a set of formulas and its transcript in PROLOG. The sentence $\varphi = \exists\vec{x}(\gamma_0 \wedge \cdots \wedge \gamma_k)$ in the last item above is also called a *query* to $\mathcal{P}$. In PROLOG it is mostly denoted by $:\!- \gamma_0, \ldots, \gamma_k.$[4] $\exists\vec{x}$ may also be empty. The origin of this notation lies in the equivalence of the kernel $\neg\gamma_0 \vee \cdots \vee \neg\gamma_k$ of $\neg\varphi \equiv \forall\vec{x}(\neg\gamma_0 \vee \cdots \vee \neg\gamma_k)$ to $\bot \leftarrow \gamma_0 \wedge \cdots \wedge \gamma_k$, omitting the writing of $\bot$.

Using rules one proceeds from given facts to arrive not only at new facts but also at answers to queries. The restriction as regards the formulas in $\mathcal{P}$ and the abstinence from $=$ is not really essential. This will become clear in Examples 1 and 4 and in the considerations of this section. Whenever required, $=$ can be treated as a binary relation symbol by adjoining the Horn sentences (4) in **4.1**.

[4] Sometimes also $?\!- \gamma_0, \ldots, \gamma_k$. Like many programming languages, PROLOG also has numerous "dialects." We shall therefore not consistently stick to a particular syntax. We also disregard many details, for instance that variables always begin with capital letters and that PROLOG recognizes certain unchanging predicates like *read*, ..., to provide a convenient user interface.

Program clauses and negated queries can equally well be written as Horn clauses: $\beta :- \beta_0, \ldots, \beta_m$ as $\{\neg\beta_1, \ldots, \neg\beta_n, \beta\}$, and $:- \gamma_0, \ldots, \gamma_k$ as $\{\neg\gamma_0, \ldots, \neg\gamma_k\}$. For a logic program $\mathcal{P}$, the corresponding set of positive Horn clauses is denoted by $\mathcal{P}$. Confusing $\mathcal{P}$ and $\mathcal{P}$ is nearly always harmless, because the two can almost always be identified. To justify this semantically, let $\mathcal{A} \vDash K$ for an $\mathcal{L}$-structure $\mathcal{A}$ and $K = \{\lambda_0, \ldots, \lambda_k\}$ simply mean $\mathcal{A} \vDash \bigvee_{i \leqslant k} \lambda_i$ which is equivalent to $\mathcal{A} \vDash (\bigvee_{i \leqslant k} \lambda_i)^{\mathfrak{G}}$. For $\mathcal{L}$-models $\mathcal{M}$, let $\mathcal{M} \vDash K$ have its ordinary meaning $\mathcal{M} \vDash \bigvee_{i \leqslant k} \lambda_i$.

The empty clause corresponds to $\bot$, so that always $\mathcal{A} \nvDash \square$. If an $\mathcal{A} \vDash K$ exists for all $K \in \mathcal{K}$, then $\mathcal{A}$ is called a *model for* $\mathcal{K}$ and $\mathcal{K}$ is called *satisfiable* or *consistent*, since this is equivalent to the consistency of the sets of sentences corresponding to $\mathcal{K}$. Further let $\mathcal{K} \vDash H$ if every model for $\mathcal{K}$ also satisfies H. Evidently $\mathcal{K} \vDash K^\sigma$ for $K \in \mathcal{K}$ and arbitrary substitutions σ, since $\mathcal{A} \vDash K \Rightarrow \mathcal{A} \vDash K^\sigma$. The clause K^σ is also termed an *instance* of K, in particular a ground instance if $var\, K^\sigma = \emptyset$.

A logic program $\mathcal{P}$, considered as a set of positive Horn formulas, is always consistent. All facts and rules of $\mathcal{P}$ are valid in the minimal Herbrand model $\mathcal{C}_\mathcal{P}$, which should be thought of as the model of a domain of objects about which one wishes to express properties by means of sentences using $\mathcal{P}$. A logic program $\mathcal{P}$ is always written such that a real situation is modeled as precisely as possible by $\mathcal{C}_\mathcal{P}$.

Suppose $\mathcal{P} \vdash \exists\vec{x}\gamma$. Then a central goal is obtaining solutions of the latter, in particular in $\mathcal{C}_\mathcal{P}$ which by Theorem 1.4 always exist. Here $\gamma\,\frac{\vec{t}}{\vec{x}}$ is called a *solution* of $\mathcal{P} \vdash \exists\vec{x}\gamma$ whenever $\mathcal{P} \vdash \gamma\,\frac{\vec{t}}{\vec{x}}$. One also speaks of the solutions $\frac{\vec{t}}{\vec{x}}$ or $\vec{x} := \vec{t}$.

Logic programming follows the strategy of proving $\mathcal{P} \vdash \varphi$ for a query φ by establishing the inconsistency of $\mathcal{P}, \neg\varphi$. To verify this we know from Theorem 1.1 that an inconsistency proof of $\mathrm{GI}(\mathcal{P}, \neg\varphi)$ suffices. The resolution theorem shows that for this proof in turn, it suffices to derive the empty clause from the set of clauses $\mathrm{GI}(\mathcal{P}, N)$ corresponding to $\mathrm{GI}(\mathcal{P}, \neg\varphi)$. Here $\mathrm{GI}(\mathcal{K})$ generally denotes the set of all ground instances of members of a set $\mathcal{K}$ of clauses, and $N = \{\neg\gamma_1, \ldots, \neg\gamma_n\}$ is the negative clause corresponding to the query φ, the so-called *goal clause*.

As a matter of fact, we proceed somewhat more artfully and work not only with ground instances but also with arbitrary instances. Nor does the search for resolutions take place coincidentally or arbitrarily, but rather with the most sparing use of substitutions possible for the purpose of unification. Before the general formulation of Theorem 4.2, we exhibit this method of "unified resolution" by means of two easy examples. In the first of these, sum denotes the graph of addition in $\mathbb{N}$.

Example 1. We consider the following logic program $\mathcal{P} = \mathcal{P}_+$ in $\mathcal{L}\{0, \mathtt{S}, \mathsf{sum}\}$:

$$\forall x\, \mathsf{sum}\, x0x \quad ; \quad \forall x \forall y \forall z (\mathsf{sum}\, xyz \to \mathsf{sum}\, x\mathtt{S}y\mathtt{S}z).$$

In PROLOG one may write this program slightly shorter as follows:

$$\mathsf{sum}\, x0x :- \quad ; \quad \mathsf{sum}\, x\mathtt{S}y\mathtt{S}z :- \mathsf{sum}\, xyz.$$

The first program clause is a "fact," the second one is a "rule." The set of Horn clauses corresponding to $\mathcal{P}$ is $\mathcal{P} = \{\{\mathsf{sum}\, x0x\}, \{\neg\mathsf{sum}\, xyz, \mathsf{sum}\, x\mathsf{S}y\mathsf{S}z\}\}$. $\mathcal{P}$ describes indeed sum = graph + in $\mathbb{N}$; more precisely, $\mathcal{C}_\mathcal{P} \simeq \mathcal{N} := (\mathbb{N}, 0, \mathsf{S}, \mathsf{sum})$, that is, $\mathcal{C}_\mathcal{P} \models \mathsf{sum}\,\underline{m}\,\underline{n}\,\underline{k} \Leftrightarrow \mathcal{N} \models \mathsf{sum}\,\underline{m}\,\underline{n}\,\underline{k}$ $(\Leftrightarrow m + n = k)$. This is deduced similarly as in Example 5 on page 111, but more directly from Exercise 2 in **4.2**. By replacing therein $p_{m,n,k}$ with $\mathsf{sum}\,\underline{m}\,\underline{n}\,\underline{k}$, the set of formulas of this exercise corresponds precisely to the ground instances of $\mathcal{P}_+$.

Examples of queries to $\mathcal{P}$ are $\exists u \exists v\, \mathsf{sum}\, u\underline{1}v$, $\exists u\, \mathsf{sum}\, uu\underline{6}$, and $\mathsf{sum}\,\underline{n}\,\underline{2}\,\underline{n+2}$ (here the $\exists$-prefix is empty). For each of these three queries φ clearly holds $\mathcal{C}_\mathcal{P} \models \varphi$. Hence, $\mathcal{P} \vdash \varphi$ by Theorem 1.4. But how can this be confirmed by a computer?

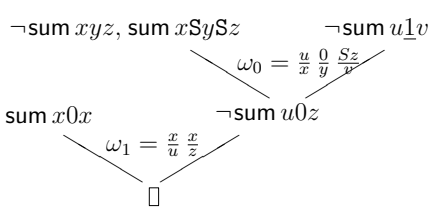
As an illustration, let $\varphi := \exists u \exists v\, \mathsf{sum}\, u\underline{1}v$. Clearly, $(u, v) := (\underline{n}, \mathsf{S}\underline{n})$ is a solution of

$$(*) \quad \mathcal{P} \vdash \exists u \exists v\, \mathsf{sum}\, u\underline{1}v.$$

We will show that $\mathcal{P} \vdash \mathsf{sum}\, x\underline{1}\mathsf{S}x$ where x occurs free in the last formula, is the general solution of $(*)$. The inconsistency proof of $\mathcal{P}, \neg\varphi$ results by deriving $\square$ from suitable instances of $\mathcal{P}, N$ which will be constructed by certain substitutions. $N := \{\neg\mathsf{sum}\, u\underline{1}v\}$ is the goal clause corresponding to φ.

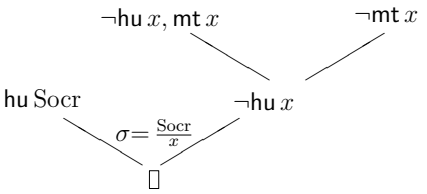
The resolution rule is not directly applicable to $\mathcal{P}, N$. But with $\omega_0 := \frac{u}{x}\,\frac{0}{y}\,\frac{Sz}{v}$ it is applicable to $P^{\omega_0}, N^{\omega_0}$, with the Horn clause $P := \{\neg\mathsf{sum}\, xyz, \mathsf{sum}\, x\mathsf{S}y\mathsf{S}z\} \in \mathcal{P}$. Indeed, one easily confirms $P^{\omega_0} = \{\neg\mathsf{sum}\, u0z, \mathsf{sum}\, u\underline{1}\mathsf{S}z\}$ and $N^{\omega_0} = \{\neg\mathsf{sum}\, u\underline{1}\mathsf{S}z\}$. The resolvent of the pair of Horn clauses $P^{\omega_0}, N^{\omega_0}$ is $N_1 := \{\neg\mathsf{sum}\, u0z\}$. This can be stated as follows: Application of RR became possible thanks to the unifiability of the clause $\{\mathsf{sum}\, x\mathsf{S}y\mathsf{S}z, \mathsf{sum}\, u\underline{1}v\}$, where $\neg\mathsf{sum}\, x\mathsf{S}y\mathsf{S}z$ belongs to P and $\neg\mathsf{sum}\, u\underline{1}v$ to N. But we have still to continue to try to get the empty clause.

Let $P_1 := \{\mathsf{sum}\, x0x\} \in \mathcal{P}$. Then P_1, N_1 can be brought to resolution by unification with $\omega_1 := \frac{x}{u}\,\frac{x}{z}$. This is because $P_1^{\omega_1} = \{\mathsf{sum}\, x0x\}$ and $N_1^{\omega_1} = \{\neg\mathsf{sum}\, x0x\}$. Now, simply apply RR to this pair of clauses to obtain $\square$. The figure on the left renders this description more intuitive. The set braces of the clauses have been omitted in the figure. This resolution can certainly be produced by a computer; the computer has just to look for appropriate unifiers!



With the above, $(*)$ is proved by Theorem 4.2(a) below. At the same time, by Theorem 4.2(b), we got a solution of $(*)$, namely $(\mathsf{sum}\, u\underline{1}v)^{\omega_0\omega_1} = \mathsf{sum}\, x\underline{1}\mathsf{S}x$. The latter is in our example a most general solution, because by substitution one can obtain from $\mathsf{sum}\, x\underline{1}\mathsf{S}x$ all individual solutions, namely all sentences $\mathsf{sum}\,\underline{n}\,\underline{1}\,\underline{\mathsf{S}n}$.

Example 2. The logic program $\mathcal{P} = \{\forall x(\mathsf{hu}\, x \to \mathsf{mt}\, x), \mathsf{hu}\, \mathrm{Socr}\}$ formalizes the two premises of the old classical Aristotelian syllogism *All humans are mortal; Socrates is a human. Hence, Socrates is mortal.* Here $\mathcal{C}_\mathcal{P}$ is the one-point model {Socr} because Socr is the only constant and no functions occur. The figure on the right shows a resolution of the query $:- \mathsf{mt}\, x$, with the solution $x := \mathrm{Socr}$; see also Theorem 4.2(b). The predicate logic argument would run as follows: $\forall x(\mathsf{hu}\, x \to \mathsf{mt}\, x)$ implies $\mathsf{hu}\, \mathrm{Socr} \to \mathsf{mt}\, \mathrm{Socr}$ by specification. Thus, since $\mathsf{hu}\, \mathrm{Socr}$, MP yields $\mathsf{mt}\, \mathrm{Socr}$. Proofs using MP can therefore also be gained by resolution.



Of course, the above examples are far too simple to display the efficiency of logic programming. Here we are interested only in illustrating the methods involved.

Following these preliminary considerations we now generalize these and start with the following definition. Its complicated look is no hindrance for programming.

Definition of the derivation rules UR and UHR of unified resolution and of unified Horn resolution, respectively. Suppose K_0, K_1 are clauses and ω is any substitution. Define $K \in U_\omega R(K_0, K_1)$ if there are clauses H_0, H_1 and $\neg$-free clauses $G_0, G_1 \neq \square$ such that after a possible swapping of the two indices,

(a) $K_0 = H_0 \cup G_0$ and $K_1 = H_1 \cup \overline{G_1}$ $\quad (\overline{G_1} = \{\overline{\lambda} \mid \lambda \in G_1\})$,

(b) ω is a generic unifier of $G_0 \cup G_1$ and $K = H_0^\omega \cup H_1^\omega$.

K is called a *U-resolvent* of K_0, K_1 or an *application of the rule UR to* K_0, K_1 if $K \in U_\omega R(K_0^\rho, K_1)$ for some ω and separator ρ of K_0, K_1.[5] The restriction of UR to Horn clauses K_0, K_1 (K_0 positive, K_1 negative) is denoted by UHR and $U_\omega R(K_0, K_1)$ by $U_\omega HR(K_0, K_1)$. The resolvent K is then termed a *UH-resolvent* of K_0, K_1.

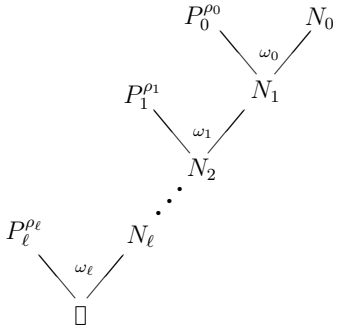
Note that by (b), $G_0^\omega = \{\pi\} = G_1^\omega$ for some prime formula π; hence K results from applying RR or HR, respectively. Applying UR or UHR to K_0, K_1 always includes a choice of ω and ρ. In the examples we used UHR. In the first resolution step of Example 1, $\neg\mathsf{sum}\, u0z \in U_{\omega_0}HR(P^\rho, N)$ (with $\rho = \iota$). The splitting of K_0 and K_1 according (a) above reads $H_0 = \{\neg\mathsf{sum}\, xyz\}$, $G_0 = \{\mathsf{sum}\, x\mathtt{S}y\mathtt{S}z\}$, and $H_1 = \emptyset$, $G_1 = \{\mathsf{sum}\, u\underline{1}v\}$. UHR was used again in the second resolution step, as well as in Example 2, strictly following the above definition instructions.

We write $\mathcal{K} \vdash^{UR} H$ if H is derivable from the set of clauses $\mathcal{K}$ using UR. Accordingly let $\mathcal{K} \vdash^{UHR} H$ be defined for sets of Horn clauses $\mathcal{K}$, where only UHR is used. Just as in propositional logic, derivations in $\vdash^{UR}$ or $\vdash^{UHR}$ can be visualized by means of trees.

[5] Using a separator is more general than demanding just $K \in U_\omega R(K_0, K_1)$; see Exercise 1.

A (successful) *U-Resolution* for $\mathcal{K}$ is just a U-resolution tree with leaves in $\mathcal{K}$ and root $\square$, the empty clause.

A *UH-resolution* is defined similarly; it may as well be regarded as a sequence $(P_i^{\rho_i}, N_i, \omega_i)_{i\leqslant\ell}$ with $N_{i+1} \in U_{\omega_i} HR(P_i^{\rho_i}, N_i)$ for $i < \ell$ and $\square \in U_{\omega_\ell} HR(P_\ell^{\rho_\ell}, N_\ell)$. If $\mathcal{P}$ is a set of positive clauses and N a negative clause, and if further $P_i \in \mathcal{P}$ holds for all $i \leqslant \ell$ and $N_0 = N$, one speaks of a *UH-resolution for* $\mathcal{P}, N$. In general, $\mathcal{P}$ consists of the clauses of some logic program and N is a given goal clause. In place of *UH*-resolution one may also speak of *SLD-resolution* (***L**inear resolution with **S**election function for **D**efinite clauses*). This name has nothing to do with some special strategy for searching a successful resolution, implemented in PROLOG. For details on this matter see for instance [Ll]. The figure on the left illustrates a *UH*-resolution $(P_i^{\rho_i}, N_i, \omega_i)_{i\leqslant\ell}$ for $\mathcal{P}, N$. It obviously generalizes the diagrams in the Examples 1 and 2, which are *UH*-resolutions as we know.



First of all we prove the soundness of the calculus $\vdash^{UR}$. Note that this also covers the calculus $\vdash^{UHR}$ of unified Horn resolution with its more special clauses.

Lemma 4.1 (Soundness lemma). $\mathcal{K} \vdash^{UR} H$ *implies* $\mathcal{K} \models H$.

Proof. It suffices to show that $K_0, K_1 \models H$ if H is a U-resolvent of K_0, K_1. Let $H \in U_\omega R(K_0^\rho, K_1)$, $K_0^\rho = H_0 \cup G_0$, $K_1 = H_1 \cup \overline{G_1}$, $G_0^\omega = \{\pi\} = G_1^\omega$, $H = H_0^{\rho\omega} \cup H_1^\omega$, and $\mathcal{A} \models K_0, K_1$, so that $\mathcal{A} \models K_0^{\rho\omega}, K_1^\omega$ as well. Further, let $w\colon Var \to A$, with $\mathcal{M} := (\mathcal{A}, w) \models K_0^{\rho\omega} = H_0^{\rho\omega} \cup \{\pi\}$, $\mathcal{M} \models H_1^\omega \cup \{\neg\pi\}$. If $\mathcal{M} \not\models \pi$ then evidently $\mathcal{M} \models H_0^{\rho\omega}$. Otherwise $\mathcal{M} \not\models \neg\pi$, hence $\mathcal{M} \models H_1^\omega$. So $\mathcal{M} \models H_0^\omega \cup H_1^\omega = H$ in any case. This states that $\mathcal{A} \models H$, because w was arbitrary. ❑

With respect to the calculus $\vdash^{UHR}$ this lemma serves the proof of (a) in

Theorem 4.2 (Main theorem of logic programming). *Let $\mathcal{P}$ be a logic program, $\exists\vec{x}\gamma$ a query, $\gamma = \gamma_0 \wedge \cdots \wedge \gamma_k$, and $N = \{\neg\gamma_0, \ldots, \neg\gamma_k\}$. Then the following hold:*

(a) *$\mathcal{P} \vdash \exists\vec{x}\gamma$ iff $\mathcal{P}, N \vdash^{UHR} \square$ (Adequacy),*

(b) *Let $(P_i^{\rho_i}, N_i, \omega_i)_{i\leqslant\ell}$ be any UH-resolution for $\mathcal{P}, N$ and $\omega := \omega_0 \cdots \omega_\ell$, then $\mathcal{P} \vdash \gamma^\omega$ (Solution soundness),*

(c) *Let $\mathcal{P} \vdash \gamma\frac{\vec{t}}{\vec{x}}$ with $\vec{t} \in \mathcal{T}_0^n$. Then there exists a UH-resolution $(K_i^{\rho_i}, N_i, w_i)_{i\leqslant\ell}$ and some τ such that $x_i^{\omega\tau} = t_i$ for $i = 1, \ldots, n$, where $\omega := \omega_0 \cdots \omega_\ell$ (Solution completeness).*

The proof, based on a substantial number of substitutions, is undertaken in **4.5**. Here are just a few comments. Since $\neg\exists\vec{x}\gamma \equiv \forall\vec{x}\neg\gamma$ it is obvious that

$(*)\quad \mathcal{P} \vdash \exists\vec{x}\gamma$

is equivalent to the inconsistency of $\mathcal{P}, \forall\vec{x}\neg\gamma$, hence also to that of the corresponding set of Horn clauses $\mathcal{P}, N$. Theorem 4.2(a) states that this is equivalent to $\mathcal{P}, N \vdash^{UHR} \square$, which is not obvious. (b) tells us how to achieve a solution of $(*)$ by a successful resolution. Since γ^ω in (b) may still contain free variables (like $(\mathsf{sum}\, u\underline{1}v)^\omega = \mathsf{sum}\, x\underline{1}\mathsf{S}x$ for $\omega = \omega_1\omega_2$ in Example 1) and since $\mathcal{P} \vdash \gamma^\omega \Rightarrow \mathcal{P} \vdash \gamma^{\omega\tau}$ for any τ, one often obtains whole families of solutions of $(*)$ in the Herbrand model $\mathcal{C}_\mathcal{P}$ by substituting ground terms. By (c), all solutions in $\mathcal{C}_\mathcal{P}$ are gained in this way. However, the theorem makes no claim as to whether and in what circumstances $(*)$ is solvable.

Logic programming is also expedient for purely theoretical purposes. For instance, it can be used to make the notion of computable functions on $\mathbb{N}$ more precise. The definition below provides just one of many similarly styled, intuitively illuminating possibilities. We will construct an undecidable problem (Theorem 4.3 below) that explains the difficulties surrounding a general answer to the question $\mathcal{P} \vdash \exists\vec{x}\gamma$. Because in **6.1** computable functions are equated with recursive functions, we keep things fairly brief here.

Definition. $f : \mathbb{N}^n \to \mathbb{N}$ is called *computable*[6] if there is a logic program $\mathcal{P}$ $(= \mathcal{P}_f)$ in a language that, in addition to 0 and S, contains only relation symbols, including a symbol denoted by r_f (to mean graph f), such that for all $\vec{k}$ and m,

(1) $\quad \mathcal{P} \vdash r_f\vec{\underline{k}}\,\underline{m} \;\Leftrightarrow\; f\vec{k} = m \qquad \big(\vec{\underline{k}} = (\underline{k_1}, \ldots, \underline{k_n})\big)$.

The domain of the Herbrand model $\mathcal{C}_\mathcal{P}$ is $\mathbb{N}$ and $\mathcal{P} \vdash r_f\vec{\underline{k}}\,\underline{m} \;\Leftrightarrow\; \mathcal{C}_\mathcal{P} \vDash r_f\vec{\underline{k}}\,\underline{m}$ by Theorem 1.4, so that (1) holds when just the following claim has been proved:

(2) $\quad \mathcal{C}_{\mathcal{P}_f} \vDash r_f\vec{\underline{k}}\,\underline{m} \;\Leftrightarrow\; f\vec{k} = m, \quad$ for all $\vec{k}, m$.

A function $f : \mathbb{N}^n \to \mathbb{N}$ satisfying (1) is certainly computable in the intuitive sense: a deduction machine is set to list all formulas provable from $\mathcal{P}$, and one simply has to wait until a sentence $r\vec{\underline{k}}\,\underline{m}$ appears. Then $m = f\vec{k}$ is computed. By Theorem 4.2(a), the left-hand side of (1) is equivalent to $\mathcal{P}, \{\neg r\vec{\underline{k}}\,\underline{m}\} \vdash^{UHR} \square$. Therefore, f is basically also computable with the resolution calculus.

Example 3. The program $\mathcal{P} = \mathcal{P}_+$ in Example 1 computes $+$, or more precisely, graph $+$. Indeed, $\mathcal{C}_\mathcal{P} \vDash \mathsf{sum}\ \underline{k}\,\underline{n}\,\underline{m} \;\Leftrightarrow\; k + n = m$ as was shown there. So (2) holds and hence also (1). A logic program $\mathcal{P}_\times$ for computing $\mathsf{prd} := \text{graph} \cdot$ arises from $\mathcal{P}_+$ by adding to the program of $\mathcal{P}_+$ the following two program clauses:

$$\mathsf{prd}\, x00 \;:\!-\; ; \quad \mathsf{prd}\, x\mathsf{S}yu \;:\!-\; \mathsf{prd}\, xyz, \mathsf{sum}\, zxu.$$

[6] By grounding the notion of computability in different terms one could f provisionally call *LP-computable*. Our definition is equivalent to that in **6.1**, but we will not make full use of this.

Example 4. The program $\mathcal{P}_{\mathtt{S}}$, consisting of the single fact $r_{\mathtt{S}}\, x\mathtt{S}x :-$, computes the graph $r_{\mathtt{S}}$ of the successor function. Clearly, $\mathcal{P}_{\mathtt{S}} \vdash r_{\mathtt{S}}\, \underline{n}\mathtt{S}\underline{n}$, since $\mathcal{P}_{\mathtt{S}}, \{\neg r_{\mathtt{S}}\, \underline{n}\mathtt{S}\underline{n}\} \vdash^{UHR} \square$ (notice that $\square$ is a resolvent of $\{r_{\mathtt{S}}\, x\mathtt{S}x\}^{\sigma}$ and $\{\neg r_{\mathtt{S}}\, \underline{n}\mathtt{S}\underline{n}\}^{\sigma}$ with $\sigma = \frac{\mathtt{S}^n 0}{x}$). Let $m \neq \mathtt{S}n$. Then $(\mathbb{N}, 0, \mathtt{S}, \operatorname{graph}\mathtt{S}) \vDash \mathcal{P}_{\mathtt{S}}, \neg r_{\mathtt{S}}\, \underline{n}\, \underline{m}$; hence $\mathcal{P}_{\mathtt{S}} \nvdash r_{\mathtt{S}}\, \underline{n}\, \underline{m}$. This proves (1).

It is not difficult to recognize that each recursive function f can be computed by a logic program $\mathcal{P}_f$ in the above sense in a language that in addition to some relation symbols contains only the operation symbols $0, \mathtt{S}$. Exercises 2 and 3 are steps in the proof, which proceeds by induction on the generating operations ***Oc***, ***Op***, and ***Oμ*** of recursive functions from **6.1**. Example 4 confirms this for the recursive initial function $\mathtt{S}$. The interested reader should study **6.1** to some extend to understand what is going on. Thus, the concept of logic programming is very comprehensive. On the other hand, this has the consequence that the question $\mathcal{P} \vdash \exists \vec{x}\gamma$ is, in general, not effectively decidable. Indeed, this is the assertion of our next theorem.

Theorem 4.3. *A logic program $\mathcal{P}$ exists whose signature contains at least a binary relation symbol r, but no operation symbols other than $0, \mathtt{S}$, so that no algorithm answers the question $\mathcal{P} \vdash \exists x\, rx\underline{k}$ for each k.*

Proof. Let $f : \mathbb{N} \to \mathbb{N}$ be recursive, but $\mathit{ran}\, f = \{m \in \mathbb{N} \mid \exists k\, fk = m\}$ nonrecursive. Such a function f exists; see Exercise 4 in **6.5**. Then we get for $\mathcal{P} := \mathcal{P}_f$,

$$\begin{aligned}
\mathcal{P} \vdash \exists x\, rx\underline{m} &\Leftrightarrow \mathcal{C}_{\mathcal{P}} \vDash \exists x\, rx\underline{m} && \text{(Theorem 1.4, } r \text{ stands for } r_f\text{)} \\
&\Leftrightarrow \mathcal{C}_{\mathcal{P}} \vDash r\underline{k}\, \underline{m} \text{ for some } k && (\mathcal{C}_{\mathcal{P}} \text{ has the domain } \mathbb{N}) \\
&\Leftrightarrow fk = m \text{ for some } k && \big(\text{by (2)}\big) \\
&\Leftrightarrow m \in \mathit{ran}\, f.
\end{aligned}$$

Thus, if the question $\mathcal{P} \vdash \exists x\, rx\underline{m}$ were decidable then so too would be the question $m \in \mathit{ran}\, f$, and this is a contradiction to the choice of f. ❑

Exercises

1. Let $H \in U_{\omega}R(K_0, K_1)$. Show that H is a U-resolvent of K_0, K_1 according to the definition, that is, there exists a (generic) ω' and a separator ρ of K_0, K_1 such that $H \in U_{\omega'}R(K_0^{\rho}, K_1)$. The converse need not hold.

2. Let $g : \mathbb{N}^n \to \mathbb{N}$ and $h : \mathbb{N}^{n+2} \to \mathbb{N}$ be computable by means of the logic programs $\mathcal{P}_g$ and $\mathcal{P}_h$, and let $f : \mathbb{N}^{n+1} \to \mathbb{N}$ arise from g, h by primitive recursion, i.e., $f(\vec{a}, 0) = g\vec{a}$ and $f(\vec{a}, k+1) = h(\vec{a}, k, f(\vec{a}, k))$ for all $\vec{a} \in \mathbb{N}^n$. Provide a logic program for computing (the graph of) f.

3. Let $\mathcal{P}_h$ and $\mathcal{P}_{g_i}$ be logic programs for computing $h : \mathbb{N}^m \to \mathbb{N}$ and $g_i : \mathbb{N}^n \to \mathbb{N}$ $(i = 1, \ldots, m)$. Further let f be defined by $f\vec{a} = h(g_1\vec{a}, \ldots, g_m\vec{a})$ for all $\vec{a} \in \mathbb{N}^n$. Give a logic program for computing f.

4.5 Proof of the Main Theorem

While we actually require the following Lemma and Theorem 5.3 below only for the unified Horn resolution, the proofs are carried out here for the more general U-resolution. The calculi $\vdash^{RR}$ and $\vdash^{HR}$ from **4.2** are given henceforth with respect to variable-free clauses of a fixed identity-free language.

Lemma 5.1. *Let K_0, K_1 be clauses with separator ρ and let $K_0^{\sigma_0}$, $K_1^{\sigma_1}$ be variable-free. Suppose K is a resolvent of $K_0^{\sigma_0}, K_1^{\sigma_1}$. Then there exist substitutions ω, τ and some $H \in U_\omega R(K_0^\rho, K_1)$ such that $H^\tau = K$, i.e., K is a ground instance of some U-resolvent of K_0, K_1. Further, for a given finite set V of variables, ω, τ can be selected such that $x^{\omega\tau} = x^{\sigma_1}$ for all $x \in V$. The same holds for Horn resolution.*

Proof. Suppose w.l.o.g. $K_0^{\sigma_0} = L_0, \pi$ and $K_1^{\sigma_1} = L_1, \neg\pi$ for some prime formula π, and $K = L_0 \cup L_1$. Let $H_i := \{\alpha \in K_i \mid \alpha^{\sigma_i} \in L_i\}$, $G_0 := \{\alpha \in K_0 \mid \alpha^{\sigma_0} = \pi\}$ and $G_1 := \{\beta \in \overline{K_1} \mid \beta^{\sigma_1} = \pi\}$, $i = 0, 1$. Then $K_0 = H_0 \cup G_0$, $K_1 = H_1 \cup \overline{G_1}$, $H_i^{\sigma_i} = L_i$, $G_i^{\sigma_i} = \{\pi\}$. Let ρ be a separator of K_0, K_1 and define σ by $x^\sigma = x^{\rho\sigma_0}$ in case $x \in var\, K_0^\rho$, and $x^\sigma = x^{\sigma_1}$ otherwise. Then $K_0^{\rho\sigma} = K_0^{\rho\rho\sigma_0} = K_0^{\sigma_0}$ (consider $\rho^2 = \iota$), along with $K_1^\sigma = K_1^{\sigma_1}$. This leads to $(G_0^\rho \cup G_1)^\sigma = G_0^{\rho\sigma} \cup G_1^\sigma = G_0^{\sigma_0} \cup G_1^{\sigma_1} = \{\pi\}$, that is, σ unifies $G_0^\rho \cup G_1$. Let ω be a generic unifier of this clause so that $\sigma = \omega\tau$ for suitable τ. Then $H := H_0^{\rho\omega} \cup H_1^\omega \in U_\omega R(K_0^\rho, K_1)$ by definition of the rule UR. Furthermore $H^\tau = K$, since $K_0^{\rho\sigma} = K_0^{\sigma_0}$. Then $\omega\tau = \sigma$ and $K_1^\sigma = K_1^{\sigma_1}$ yield

$$H^\tau = H_0^{\rho\omega\tau} \cup H_1^{\omega\tau} = H_0^{\rho\sigma} \cup H_1^\sigma = H_0^{\sigma_0} \cup H_1^{\sigma_1} = L_0 \cup L_1 = K.$$

V being finite, ρ can be chosen such that $V \cap var\, K_0^\rho = \emptyset$. By definition of σ and by virtue of $\sigma = \omega\tau$ it then follows that $x^{\omega\tau} = x^\sigma = x^{\sigma_1}$ also for $x \in V$. ❏

Lemma 5.2 (Lifting lemma). *Suppose* $\mathrm{GI}(\mathcal{K}) \vdash^{RR} \square$ *for some set of clauses $\mathcal{K}$. Then also $\mathcal{K} \vdash^{UR} \square$. If $\mathcal{K}$ consists of Horn clauses only, then $\mathcal{K} \vdash^{UHR} \square$.*

Proof. We prove the claim *If* $\mathrm{GI}(\mathcal{K}) \vdash^{RR} K$ *then exist H and σ with $\mathcal{K} \vdash^{UR} H$ and $K = H^\sigma$*. For $K = \square$ is this the lemma (remember that $\square^\sigma = \square$). Our claim follow straightforwardly by induction on $\mathrm{GI}(\mathcal{K}) \vdash^{RR} K$; it is clear for $K \in \mathrm{GI}(\mathcal{K})$, and for the inductive step ($\mathrm{GI}(\mathcal{K}) \vdash^{RR} K_0, K_1$ and K is a resolvent of K_0, K_1) one merely requires Lemma 5.1. The case for Horn clauses is completely similar. ❏

Theorem 5.3 (U-resolution theorem). *A set of clauses $\mathcal{K}$ is inconsistent iff $\mathcal{K} \vdash^{UR} \square$; a set of Horn clauses $\mathcal{K}$ is inconsistent iff $\mathcal{K} \vdash^{UHR} \square$.*

Proof. If $\mathcal{K} \vdash^{UR} \square$ then $\mathcal{K} \models \square$ by Lemma 4.1, hence $\mathcal{K}$ is inconsistent. Suppose now the latter, so that the set U of $\forall$-sentences corresponding to $\mathcal{K}$ is inconsistent as well. Then $\mathrm{GI}(U)$ is inconsistent according to Theorem 1.1, hence $\mathrm{GI}(\mathcal{K})$ as well. Thus, $\mathrm{GI}(\mathcal{K}) \vdash^{RR} \square$ by Theorem 2.2 and so $\mathcal{K} \vdash^{UR} \square$ by Lemma 5.2. For sets of Horn clauses the proof runs similarly using the above lemma and Theorem 2.4. ❏

Proof of Theorem 4.2. (a): $\mathcal{P} \vdash \exists \vec{x}\gamma$ is equivalent to the inconsistency of $\mathcal{P}, \forall\vec{x}\neg\gamma$ or of $\mathcal{P}, N$. But this, by Theorem 5.3, is the same as saying $\mathcal{P}, N \vdash^{UHR} \square$.

(b): Proof by induction on the length ℓ of a successful *UH*-resolution $(P_i^{\rho_i}, N_i, \omega_i)_{i\leqslant\ell}$ for $\mathcal{P}, N$. Let $\ell = 0$, so that $\square \in U_\omega HR(P_0^\rho, N)$ for suitable ρ, ω. Then ω unifies $P_0^\rho \cup N = P_0^\rho \cup \{\gamma_0, \ldots, \gamma_k\}$, i.e., $P_0^{\rho\omega} = \{\pi\} = \gamma_i^\omega$ for some prime formula π and all $i \leqslant k$. By virtue of $P_0 \in \mathcal{P}$ we obtain $\mathcal{P} \vdash \gamma_i^\omega\ (= \pi)$ for each $i \leqslant k$, and so $\mathcal{P} \vdash \gamma_0^\omega \wedge \cdots \wedge \gamma_k^\omega = \gamma^\omega$ as claimed. Now let $\ell > 0$. Then $(P_i^{\rho_i}, N_i, \omega_i)_{1\leqslant i\leqslant\ell}$ is a *UH*-resolution for $\mathcal{P}, N_1$ as well. By the induction hypothesis,

(1) $\mathcal{P} \vdash \alpha^{\omega_1\cdots\omega_\ell}$ whenever $\neg\alpha \in N_1$.

It suffices to verify that $\mathcal{P} \vdash \gamma_i^\omega$ for all $i \leqslant k$. To this end we distinguish two cases for given i: if $\neg\gamma_i^{\omega_0} \in N_1$ then $\mathcal{P} \vdash (\gamma_i^{\omega_0})^{\omega_1\cdots\omega_\ell}$ by (1), hence $\mathcal{P} \vdash \gamma_i^\omega$. Now suppose $\neg\gamma_i^{\omega_0} \notin N_1$. Then $\gamma_i^{\omega_0}$ disappears in the resolution step from $P_0^{\rho_0}, N_0\ (= N)$ to N_1. So P_0 takes the form $P_0 = \{\neg\beta_1, \ldots, \neg\beta_m, \beta\}$ where $\beta^{\rho_0\omega_0} = \gamma_i^{\omega_0}$ and $\neg\beta_j^{\rho_0\omega_0} \in N_1$ for $j = 1, \ldots, m$. Thus (1) evidently yields $\mathcal{P} \vdash (\beta_j^{\rho_0\omega_0})^{\omega_1\cdots\omega_\ell}$, hence $\mathcal{P} \vdash \bigwedge_{j=1}^m \beta_j^{\rho_0\omega}$. At the same time $\mathcal{P} \vdash \bigwedge_{j=1}^m \beta_j^{\rho_0\omega} \to \beta^{\rho_0\omega}$ because of $\mathcal{P} \vDash P_0^{\rho_0\omega}$. Using MP we then obtain $\mathcal{P} \vdash \beta^{\rho_0\omega}$. From $\beta^{\rho_0\omega_0} = \gamma_i^{\omega_0}$ and an application of $\omega_1 \cdots \omega_\ell$ to both sides we obtain $\beta^{\rho_0\omega} = \gamma_i^\omega$, thus proving $\mathcal{P} \vdash \gamma_i^\omega$ also in the second case.

(c): Let $\mathcal{P} \vdash \gamma^\sigma$ such that $\sigma := \frac{\vec{t}}{\vec{x}}$. Then $\mathcal{P}, \neg\gamma^\sigma$ is inconsistent, and by Theorem 1.1 so too is $\mathcal{P}', \neg\gamma^\sigma$ where $\mathcal{P}' := \mathrm{GI}(\mathcal{P})$ (consider $\mathrm{GI}(\neg\gamma^\sigma) = \{\neg\gamma^\sigma\}$). According to Theorem 2.4, there is an *H*-resolution $\boldsymbol{B} = (P_i', Q_i)_{i\leqslant\ell}$ for $\mathcal{P}', N^\sigma$, that is, $Q_0 = N^\sigma$. Here let, say, $P_i' = P_i^{\sigma_i}$ for appropriate $P_i \in \mathcal{P}$ and σ_i. From this we obtain

(2) *for finite* $V \subseteq$ *Var there exist* $\rho_i, N_i, \omega_i, \tau$ *such that* $(P_i^{\rho_i}, N_i, \omega_i)_{i\leqslant\ell}$ *is a UH-resolution for* $\mathcal{P}, N$, *and* $x^{\omega\tau} = x^\sigma$ *for* $\omega := \omega_0 \cdots \omega_\ell$ *and all* $x \in V$.

This completes our reasoning, since (2) yields (for $V = \{x_1, \ldots, x_n\}$) $x_i^{\omega\tau} = x_i^\sigma = t_i$ for $i = 1, \ldots, n$, whence (c). For the inductive proof of (2) look at the first resolution step $Q_1 = HR(P_0', Q_0)$ in $\boldsymbol{B}$. By Lemma 5.1 choose $\omega_0, \rho_0, \tau_0, H$ (where $K_0 := P_0$, $K_1 := N_0 = N$, $\sigma_1 := \sigma$) in such a way that $H \in U_\omega HR(P_0^{\rho_0}, N_0)$ and $H^{\tau_0} = Q_1$, as well as $x^{\omega_0\tau_0} = x^\sigma$ for all $x \in V$. If $\ell = 0$, that is, $Q_1 = \square$, then also $H = \square$ and (2) is proved with $\tau = \tau_0$. Now suppose $\ell > 0$. For the *H*-resolution $(P_i', Q_i)_{1\leqslant i\leqslant\ell}$ for $\mathcal{P}', Q_1$ and for $V' := var\{x^{\omega_0} \mid x \in V\}$ there exist by the induction hypothesis ρ_i, N_i, ω_i for $i = 1, \ldots, \ell$ and some τ, such that $(P_i^{\rho_i}, N_i, \omega_i)_{1\leqslant i\leqslant\ell}$ is a *UH*-resolution for $\mathcal{P}, H$ and simultaneously $y^{\omega_1\cdots\omega_\ell\tau} = y^{\tau_0}$ for all $y \in V'$ (instead of $Q_0 = N^\sigma$ here $Q_1 = H^{\tau_0}$). Because of $var\, x^{\omega_0} \subseteq V'$ and $x^{\omega_0\tau_0} = x^\sigma$ for $x \in V$ we get

(3) $x^{\omega\tau} = (x^{\omega_0})^{\omega_1\cdots\omega_\ell\tau} = x^{\omega_0\tau_0} = x^\sigma$, for all $x \in V$.

$(P_i^{\rho_i}, N_i, \omega_i)_{i\leqslant\ell}$ is certainly a *UH*-resolution. Moreover, by virtue of (3), in addition $x_i^{\omega\tau} = x_i^\sigma$ for $i = 1, \ldots, n$. This proves (2), hence (c), and completes the proof of the main theorem.

Chapter 5

Elements of Model Theory

Model theory can be seen as applied mathematical logic. Here the techniques developed in mathematical logic are combined with construction methods of other areas (such as algebra and analysis) to their mutual benefit. The following demonstrations can provide only a first glimpse in this respect, a deeper understanding being gained, for instance, from [CK] or [Ho]. For further-ranging topics, such as saturated models, stability theory, and the model theory of languages other than elementary ones, we refer to the special literature, [Bue], [Mar], [Pz], [Rot], [Sa], [Sh].

The theorems of Löwenheim and Skolem were first formulated in the generality given in **5.1** by Tarski. These and the compactness theorem form the basis of model theory, a now wide-ranging discipline that arose around 1950. Key concepts of model theory are elementary equivalence and elementary extension. These are not only interesting in themselves but also have multiple applications to model constructions in set theory, nonstandard analysis, algebra, geometry and elsewhere.

Complete axiomatizable theories are decidable; see **3.5**. The question of decidability and completeness of mathematical theories and the development of well-honed methods that solve these questions have always been a driving force for the further development of mathematical logic. Of the numerous methods, we introduce here the most important: Vaught's test, Ehrenfeucht's game, Robinson's method of model completeness, and quantifier elimination. For more involved cases, such as the theories of algebraically closed and real closed fields, model-theoretical criteria are developed and applied. For a complete understanding of the material in **5.5** the reader should to some extent be familiar with basic constructions in classical algebra, mainly concerning the theory of fields.

Chapter **2** should have been read. From Chapter **3** we require a certain amount of material for applications of model theory to the solution of decision problems, and from Chapter **4** just the notion of a Horn formula.

5.1 Elementary Extensions

In **3.3** nonstandard models were obtained using a method that we now generalize. For given $\mathcal{L}$ and a set A let $\mathcal{L}A$ denote the language resulting from $\mathcal{L}$ by adjoining new constant symbols $\boldsymbol{a}$ for all $a \in A$. The symbol $\boldsymbol{a}$ should depend only on a, so that $\mathcal{L}A \subseteq \mathcal{L}B$ whenever $A \subseteq B$. To simplify notation we write from Theorem 1.3 onwards just a rather than $\boldsymbol{a}$; there will be no risk of misunderstanding.

Let $\mathcal{B}$ be an $\mathcal{L}$-structure and $A \subseteq B$ (the domain of $\mathcal{B}$). Then the $\mathcal{L}A$-expansion in which $\boldsymbol{a}$ is interpreted by $a \in A$ will be denoted by $\mathcal{B}_A$. According to Exercise 3 in **2.3** holds for arbitrary $\alpha = \alpha(\vec{x}) \in \mathcal{L}$ and $\vec{a} \in A^n$, with $\alpha(\vec{\boldsymbol{a}}) := \alpha\, \frac{\boldsymbol{a}_1}{x_1} \cdots \frac{\boldsymbol{a}_n}{x_n}$,

(1) $\mathcal{B} \vDash \alpha\,[\vec{a}] \Leftrightarrow \mathcal{B}_A \vDash \alpha(\vec{\boldsymbol{a}})$.

Clearly, every sentence from $\mathcal{L}A$ is of the form $\alpha(\vec{\boldsymbol{a}})$ for suitable $\alpha(\vec{x}) \in \mathcal{L}$ and $\vec{a} \in A^n$. Instead of $\mathcal{B}_A \vDash \alpha(\vec{\boldsymbol{a}})$ (which is equivalent to $\mathcal{B} \vDash \alpha\,[\vec{a}]$) we later write just $\mathcal{B}_A \vDash \alpha(\vec{a})$ or even $\mathcal{B} \vDash \alpha(\vec{a})$, as in Theorem 1.3. Thus, $\mathcal{B}$ may also denote a constant expansion of $\mathcal{B}$ if it is not the distinction that is to be emphasized. This notation is somewhat sloppy but points up the ideas behind the constructions.

Note that for an $\mathcal{L}$-structure $\mathcal{A}$, the $\mathcal{L}A$-expansion $\mathcal{A}_A$ receives a new constant symbol for *every* $a \in A$, even if some elements of $\mathcal{A}$ already possess names in $\mathcal{L}$. The set of all variable-free literals $\lambda \in \mathcal{L}A$ such that $\mathcal{A}_A \vDash \lambda$ is called the *diagram* $D\mathcal{A}$ of $\mathcal{A}$. For instance, $D(\mathbb{R}, <)$ contains for all $a, b \in \mathbb{R}$ the literals $\boldsymbol{a} = \boldsymbol{b}$, $\boldsymbol{a} \neq \boldsymbol{b}$, $\boldsymbol{a} < \boldsymbol{b}$, $\boldsymbol{a} \not< \boldsymbol{b}$, depending on whether indeed $a{=}b$, $a{\neq}b$, $a{<}b$, or $a{\not<}b$ for the reals a, b. Diagrams are important for various constructions of model extensions.

The notion of an embedding $\imath\colon \mathcal{A} \to \mathcal{B}$ as defined in **2.1** (that is, the image of $\mathcal{A}$ under $\imath$ is an isomorphic copy of $\mathcal{A}$), embraces the notion of a substructure. For $\mathcal{A} \subseteq \mathcal{B}$, the mapping $\imath = id_A$ is the trivial *identical embedding* of $\mathcal{A}$ into $\mathcal{B}$.

Let $\mathcal{L}_0 \subseteq \mathcal{L}$. In this chapter, the embeddability of an $\mathcal{L}_0$-structure $\mathcal{A}$ into an $\mathcal{L}$-structure $\mathcal{B}$ often means the embeddability of $\mathcal{A}$ into the $\mathcal{L}_0$-reduct $\mathcal{B}_0$ of $\mathcal{B}$, and we shall write $\mathcal{A} \subseteq \mathcal{B}$ also in this case. In this sense the group $\mathbb{Z}$, for example, is embeddable into the field $\mathbb{Q}$. Our first statement is

Theorem 1.1. *Suppose $\mathcal{L}_0 \subseteq \mathcal{L}$ and let $\mathcal{A}$ be an $\mathcal{L}_0$-structure. An $\mathcal{L}A$-structure $\mathcal{B}$ satisfies $D\mathcal{A}$ if and only if $\imath : a \mapsto \boldsymbol{a}^{\mathcal{B}}$ is an embedding of $\mathcal{A}$ in $\mathcal{B}$.*

Proof. Let $\mathcal{B} \vDash D\mathcal{A}$ and $a, b \in A$, $a \neq b$. Then $\boldsymbol{a} \neq \boldsymbol{b} \in D\mathcal{A}$. Hence $\mathcal{B} \vDash \boldsymbol{a} \neq \boldsymbol{b}$, or equivalently, $\boldsymbol{a}^{\mathcal{B}} \neq \boldsymbol{b}^{\mathcal{B}}$. Thus $\imath$ is injective. For $r \in L_0$ and $\vec{a} \in A^n$ it holds that

$$\begin{aligned} r^{\mathcal{A}}\vec{a} \;&\Leftrightarrow\; r\vec{\boldsymbol{a}} \in D\mathcal{A} \quad\Leftrightarrow\; \mathcal{B} \vDash r\vec{\boldsymbol{a}} \quad \big(\text{since } \mathcal{B} \vDash D\mathcal{A}\big) \\ &\Leftrightarrow\; r^{\mathcal{B}}\imath\vec{a} \qquad \big(\imath\vec{a} := (\imath a_1, \ldots, \imath a_n)\big). \end{aligned}$$

Similarly $\imath f^{\mathcal{A}}\vec{a} = f^{\mathcal{B}}\imath\vec{a}$ is obtained, for note that whenever $\vec{a} \in A^n$ and $b \in A$ then $f^{\mathcal{A}}\vec{a} = b \Leftrightarrow f\vec{\boldsymbol{a}} = \boldsymbol{b} \in D\mathcal{A} \Leftrightarrow \mathcal{B} \vDash f\vec{\boldsymbol{a}} = \boldsymbol{b} \Leftrightarrow f^{\mathcal{B}}\imath\vec{a} = \imath b$. Thus, $\imath$ is indeed an embedding. Now suppose the latter. For variable-free terms t in $\mathcal{L}_0 A$ one easily

verifies $\imath t^{\mathcal{A}} = t^{\mathcal{B}}$, where here and elsewhere $t^{\mathcal{A}}, t^{\mathcal{B}}$ are to mean more precisely $t^{\mathcal{A}_A}$ and $t^{\mathcal{B}_A}$. Since $\imath$ is injective it follows for variable-free equations in $\mathcal{L}_0 A$ that

$$t_1 = t_2 \in D\mathcal{A} \;\Leftrightarrow\; t_1^{\mathcal{A}} = t_2^{\mathcal{A}} \;\Leftrightarrow\; \imath t_1^{\mathcal{A}} = \imath t_2^{\mathcal{A}} \;\Leftrightarrow\; t_1^{\mathcal{B}} = t_2^{\mathcal{B}} \;\Leftrightarrow\; \mathcal{B} \vDash t_1 = t_2.$$

In the same way $t_1 \neq t_2 \in D\mathcal{A} \Leftrightarrow \mathcal{B} \vDash t_1 \neq t_2$ is verified, and prime sentences of the form $r\vec{t}$ are dealt with analogously. This proves $\mathcal{B} \vDash D\mathcal{A}$. ❑

Corollary 1.2. *Let $\mathcal{A}$ be an $\mathcal{L}$-structure. $\mathcal{B} \vDash D\mathcal{A}$ iff $\mathcal{A}$ is embeddable into $\mathcal{B}$. Moreover, $\mathcal{B} \vDash D\mathcal{A} \;\Leftrightarrow\; \mathcal{A} \subseteq \mathcal{B}$ provided $A \subseteq B$.*

Indeed, by the theorem with $\mathcal{L}_0 = \mathcal{L}$, the mapping $\imath : a \mapsto \boldsymbol{a}^{\mathcal{B}}$ realizes the embedding, and also the converse of the claim is obvious. $\imath$ is the identical mapping in case $A \subseteq B$, which verifies the "Moreover" part. Frequent use will be made of this corollary, without mentioning it explicitly. Taking an (algebraic) *prime model* for a theory T to mean a model embeddable into every T-model, the corollary states that $\mathcal{A}_A$ is a prime model for $D\mathcal{A}$, understood as a theory. We will use the concept of a prime model only in this sense.[1]

Probably the most important concept in model theory, for which a first example appears on the next page, is given by the following

Definition. Let $\mathcal{A}, \mathcal{B}$ be $\mathcal{L}$-structures. $\mathcal{A}$ is called an *elementary substructure* of $\mathcal{B}$, and $\mathcal{B}$ an *elementary extension* of $\mathcal{A}$, in symbols $\mathcal{A} \preccurlyeq \mathcal{B}$, if $A \subseteq B$ and

(2) $\quad \mathcal{A} \vDash \alpha\,[\vec{a}] \;\Leftrightarrow\; \mathcal{B} \vDash \alpha\,[\vec{a}]$, for all $\alpha = \alpha(\vec{x}) \in \mathcal{L}$ and $\vec{a} \in A^n$.

Obviously, $\mathcal{A} \preccurlyeq \mathcal{B}$ implies $\mathcal{A} \subseteq \mathcal{B}$. Terming $D_{el}\mathcal{A} := \{\alpha \in \mathcal{L}A^0 \mid \mathcal{A}_A \vDash \alpha\}$ the *elementary diagram* of $\mathcal{A}$, $\mathcal{A} \preccurlyeq \mathcal{B}$ is equivalent to $A \subseteq B$ and $\mathcal{B}_A \vDash D_{el}\mathcal{A}$. Namely, (2) already holds given only $\mathcal{A} \vDash \alpha\,[\vec{a}] \;\Rightarrow\; \mathcal{B} \vDash \alpha\,[\vec{a}]$, for all $\alpha = \alpha(\vec{x}) \in \mathcal{L}$, $\vec{a} \in A^n$.

(2) is equivalent to $\mathcal{A}_A \vDash \alpha(\vec{\boldsymbol{a}}) \Leftrightarrow \mathcal{B}_A \vDash \alpha(\vec{\boldsymbol{a}})$, by (1). And since every $\alpha \in \mathcal{L}A$ is of the form $\alpha(\vec{\boldsymbol{a}})$ for appropriate $\alpha(\vec{x}) \in \mathcal{L}$, $\vec{a} \in A^n$ and $n \geqslant 0$, the property $\mathcal{A} \preccurlyeq \mathcal{B}$ is also characterized by $A \subseteq B$ and $\mathcal{A}_A \equiv \mathcal{B}_A$ (elementary equivalence in $\mathcal{L}A$).

In general, $\mathcal{A} \preccurlyeq \mathcal{B}$ means much more than $\mathcal{A} \subseteq \mathcal{B}$ and $\mathcal{A} \equiv \mathcal{B}$. For instance, let $\mathcal{A} = (\mathbb{N}_+, <)$ and $\mathcal{B} = (\mathbb{N}, <)$. Then certainly $\mathcal{A} \subseteq \mathcal{B}$, and since $\mathcal{A} \simeq \mathcal{B}$, also $\mathcal{A} \equiv \mathcal{B}$. But $\mathcal{A} \preccurlyeq \mathcal{B}$ is false. For example, $\exists x\, x < \mathbf{1}$ holds in $\mathcal{B}_A$, but obviously not in $\mathcal{A}_A$. The following theorem will prove to be very useful for, among other things, the provision of nontrivial examples for $\mathcal{A} \preccurlyeq \mathcal{B}$:

Theorem 1.3 (Tarski's criterion). *For $\mathcal{L}$-structures $\mathcal{A}, \mathcal{B}$ with $\mathcal{A} \subseteq \mathcal{B}$ the following conditions are equivalent:*

(i) $\mathcal{A} \preccurlyeq \mathcal{B}$,

(ii) $\mathcal{B} \vDash \exists y \varphi(\vec{a}, y) \;\Rightarrow\; \mathcal{B} \vDash \varphi(\vec{a}, a)$ *for some* $a \in A \qquad \big(\varphi(\vec{x}, y) \in \mathcal{L},\ \vec{a} \in A^n\big)$.

[1] It must be distinguished from the concept of an *elementary* prime model for T, which means that $\mathcal{A}$ is elementarily embeddable into every $\mathcal{B} \vDash T$ in the sense of Exercise 2.

Proof. (i)⇒(ii): Let $\mathcal{A} \preccurlyeq \mathcal{B}$ and $\mathcal{B} \vDash \exists y \varphi(\vec{a}, y)$, so that also $\mathcal{A} \vDash \exists y \varphi(\vec{a}, y)$. Then clearly $\mathcal{A} \vDash \varphi(\vec{a}, a)$ for some $a \in A$. But $\mathcal{A} \preccurlyeq \mathcal{B}$; hence $\mathcal{B} \vDash \varphi(\vec{a}, a)$. (ii)⇒(i): Since $\mathcal{A} \subseteq \mathcal{B}$, (2) certainly holds for prime formulas. The induction steps for $\wedge, \neg$ are obvious. Only the quantifier step needs a closer look:

$$\begin{aligned} \mathcal{A} \vDash \forall y \varphi(\vec{a}, y) &\Leftrightarrow \mathcal{A} \vDash \varphi(\vec{a}, a) \text{ for all } a \in A \\ &\Leftrightarrow \mathcal{B} \vDash \varphi(\vec{a}, a) \text{ for all } a \in A \quad \text{(induction hypothesis)} \\ &\Leftrightarrow \mathcal{B} \vDash \forall y \varphi(\vec{a}, y) \qquad \text{(see below)}. \end{aligned}$$

We prove the direction ⇒ in the last equivalence indirectly: Assume $\mathcal{B} \nvDash \forall y \varphi(\vec{a}, y)$. Then $\mathcal{B} \vDash \exists y \neg \varphi(\vec{a}, y)$. Hence $\mathcal{B} \vDash \neg \varphi(\vec{a}, a)$ for some $a \in A$ according to (ii). Thus, $\mathcal{B} \vDash \varphi(\vec{a}, a)$ cannot hold for all $a \in A$. ❑

Interesting examples for $\mathcal{A} \preccurlyeq \mathcal{B}$ are provided in a surprisingly simple way by the following Theorem which, unfortunately, is applicable only if $\mathcal{B}$ has "many automorphisms" as is the case in the example below, and in geometry, for instance.

Theorem 1.4. *Let $\mathcal{A} \subseteq \mathcal{B}$. Suppose that for all n, all $\vec{a} \in A^n$, and all $b \in B$ there is an automorphism $\imath : \mathcal{B} \to \mathcal{B}$ such that $\imath\vec{a} = \vec{a}$, and $\imath b \in A$. Then $\mathcal{A} \preccurlyeq \mathcal{B}$.*

Proof. It suffices to verify condition (ii) in Theorem 1.3. Let $\mathcal{B} \vDash \exists y \varphi(\vec{a}, y)$, or equivalently $\mathcal{B} \vDash \varphi(\vec{a}, b)$ for some $b \in B$. Then $\mathcal{B} \vDash \varphi(\imath\vec{a}, \imath b)$ according to Theorem 2.3.4, and since $\imath\vec{a} = \vec{a}$, we obtain $\mathcal{B} \vDash \varphi(\vec{a}, a)$ with $a := \imath b \in A$. This proves (ii). ❑

Example. It is readily shown that for given $a_1, \ldots, a_n \in \mathbb{Q}$ and $b \in \mathbb{R}$ there exists an automorphism of $(\mathbb{R}, <)$ that maps b to a rational number and leaves $a_1, \ldots, a_n$ fixed (Exercise 3). Thus, $(\mathbb{Q}, <) \preccurlyeq (\mathbb{R}, <)$. In particular $(\mathbb{Q}, <) \equiv (\mathbb{R}, <)$.

Here an outlook at less simple examples of elementary extensions, considered more closely in **5.5**. Let $\mathcal{A} = (\mathbb{A}, 0, 1, +, \cdot)$ denote the *field of algebraic numbers* and $\mathcal{C}$ the field of complex numbers. The domain $\mathbb{A}$ consists of all complex numbers that are zeros of (monic) polynomials with rational coefficients. Then $\mathcal{A} \preccurlyeq \mathcal{C}$. Similarly, $\mathcal{A}_r \preccurlyeq \mathcal{R}$ where $\mathcal{A}_r$ denotes the field of all *real algebraic numbers* and $\mathcal{R}$ is the field of all reals. Both these facts follow from the model completeness of the theory of algebraically closed and real closed fields, respectively, proven on page 154.

Before continuing we will acquaint ourselves somewhat with transfinite cardinal numbers. It is possible to assign a set-theoretical object denoted by $|M|$ not only to finite sets but in fact to arbitrary sets M in such a way that

(3) $M \sim N \Leftrightarrow |M| = |N|$ ($\sim$ means equipotency, see page 87).

$|M|$ is called the *cardinal number* or *cardinality* of M. For a finite set M, $|M|$ is just the number of elements in M; for an infinite set M, $|M|$ is called a *transfinite cardinal number*, or briefly a transfinite cardinal.

At this stage it is unimportant just how $|M|$ is defined in detail. Significant are (4) and (5), taken as granted, from which (6) and (7) straightforwardly follow.

(4) The cardinal numbers are well-ordered according to size, i.e., each nonempty collection of them possesses a smallest element. Here let $|N| \leqslant |M|$ if there is an injection from N to M.[2] In particular, $|\mathbb{N}| \leqslant |M|$ for any infinite M.

(5) $|M \cup N| = |M \times N| = \max\{|M|, |N|\}$ for any sets M and N of which at least one is infinite.

We first prove that $M^* := \bigcup_{n>0} M^n$ has the same cardinality as M for infinite M (M^* is the set of all nonempty finite sequences of elements of M). In short,

(6) $|M^*| = |M| \qquad (M \text{ infinite})$.

Indeed, $|M^1| = |M|$, and the hypothesis $|M^n| = |M|$ yields $|M^{n+1}| = |M^n \times M| = |M|$ by (5). Thus $|M^n| = |M|$ for all n. Therefore $|M^*| = |\bigcup_{n>0} M^n| = |M \times \mathbb{N}| = |M|$. One similarly obtains from (4), (5) for every transfinite cardinal κ the property

(7) If $A_0, A_1 \ldots$ are any sets and $|A_n| \leqslant \kappa$ for all $n \in \mathbb{N}$ then $|\bigcup_{n\in\mathbb{N}} A_n| \leqslant \kappa$.

The smallest transfinite cardinal number is that of the countably infinite sets, denoted by $\aleph_0$. The next one is called $\aleph_1$. Then follows $\aleph_2$ etc. The Cantor–Bernstein theorem readily shows that the power set $\mathfrak{P}\mathbb{N}$ and the set $\mathbb{R}$ have the same cardinality, denoted by $2^{\aleph_0}$. Certainly $\aleph_0 < 2^{\aleph_0}$, and so clearly $\aleph_1 \leqslant 2^{\aleph_0}$. Cantor's *continuum hypothesis* (CH) states that $\aleph_1 = 2^{\aleph_0}$. CH is provably independent in ZFC; see e.g. [Ku]. While there are axioms extending beyond ZFC that decide CH one way or another, none of these is sufficiently plausible to be regarded as "true." In the last decades some evidence has been collected that suggests that $2^{\aleph_0} = \aleph_2$, but this is seemingly not yet enough to convince the majority of mathematicians.

The cardinality of a structure $\mathcal{A}$ is always that of its domain, that is, $|\mathcal{A}| := |A|$. The following theorem generalizes Theorem 3.4.1 page 87 essentially. The additive "downwards" prevents a mix up of these theorems. For $|\mathcal{B}| \geqslant |\mathcal{L}|$, Theorem 1.5 ensures the existence of some $\mathcal{A} \preccurlyeq \mathcal{B}$ (in particular $\mathcal{A} \equiv \mathcal{B}$) such that $|\mathcal{A}| \leqslant |\mathcal{L}|$.

Theorem 1.5 (Löwenheim–Skolem theorem downwards). *Suppose $\mathcal{B}$ is an $\mathcal{L}$-structure such that $|\mathcal{L}| \leqslant |\mathcal{B}|$ and let $A_0 \subseteq B$ be arbitrary. Then $\mathcal{B}$ has an elementary substructure $\mathcal{A}$ of cardinality $\leqslant \max\{|A_0|, |\mathcal{L}|\}$ such that $A_0 \subseteq A$.*

Proof. We construct a sequence $A_0 \subseteq A_1 \subseteq \cdots \subseteq B$ as follows. Let A_k be given. For every $\alpha = \alpha(\vec{x}, y)$ and $\vec{a} \in A_k^n$ such that $\mathcal{B} \vDash \exists y \alpha(\vec{a}, y)$ we select some $b \in B$ with $\mathcal{B} \vDash \alpha(\vec{a}, b)$ and adjoin b to A_k, thus getting A_{k+1}. In particular, if α is $f\vec{x} = y$ then certainly $\mathcal{B} \vDash \exists y\, f\vec{a} = y$. Since $\mathcal{B} \vDash \exists! y\, f\vec{a} = y$, there is no alternative selection, hence $f^{\mathcal{B}}\vec{a} \in A_{k+1}$. Thus, $A := \bigcup_{k\in\mathbb{N}} A_k$ is closed under the operations of $\mathcal{B}$, and

[2] With this definition $|M| \leqslant |N|$ & $|N| \leqslant |M| \;\Rightarrow\; |M| = |N|$ is derivable without AC , called the *Cantor–Bernstein Theorem*. Actually, the first proof without AC (more elegant than Bernstein's) is due to Dedekind who left it unpublished in his diary from 1887.

therefore defines a substructure $\mathcal{A} \subseteq \mathcal{B}$. We prove $\mathcal{A} \preccurlyeq \mathcal{B}$ by Tarski's criterion. Let $\mathcal{B} \vDash \exists y \alpha(\vec{a}, y)$ for $\alpha = \alpha(\vec{x}, y)$ and $\vec{a} \in A^n$. Then $\vec{a} \in A_k^n$ for some k. Therefore, there is some $a \in A_{k+1}$ (hence $a \in A$) such that $\mathcal{B} \vDash \alpha(\vec{a}, a)$. This proves (ii) in Theorem 1.3 and so $\mathcal{A} \preccurlyeq \mathcal{B}$. It remains to prove $|A| \leqslant \kappa := \max\{|A_0|, |\mathcal{L}|\}$. There are at most κ formulas and κ finite sequences of elements in A_0. Thus, by definition of A_1, at most κ new elements are adjoined to A_0. Hence $|A_1| \leqslant \kappa$. Similarly, $|A_n| \leqslant \kappa$ is verified for each $n > 0$. By (7) we thus get $|\bigcup_{n\in\mathbb{N}} A_n| \leqslant \kappa$. ❑

Combined with the compactness theorem, the above theorem yields

Theorem 1.6 (Löwenheim–Skolem theorem upwards). *Let $\mathcal{C}$ be any infinite $\mathcal{L}$-structure and $\kappa \geqslant \max\{|\mathcal{C}|, |\mathcal{L}|\}$. Then there exists an $\mathcal{A} \succcurlyeq \mathcal{C}$ with $|\mathcal{A}| = \kappa$.*

Proof. Let $D \supseteq C$ where $|D| = \kappa$. From (6) it follows that $|\mathcal{L}D| = \kappa$, because the alphabet of $\mathcal{L}D$ has cardinality κ. Because $|C| \geqslant \aleph_0$, by the compactness theorem, $D_{el}\mathcal{C} \cup \{c \neq d \mid c, d \in D,\ c \neq d\}$ has a model $\mathcal{B}$. Since $d \mapsto d^{\mathcal{B}}$ $(d \in D)$ is injective we may assume $d^{\mathcal{B}} = d$ for all $d \in D$, i.e., $D \subseteq B$. By Theorem 1.5 with $\mathcal{L}D$ for $\mathcal{L}$ and D for A_0, there is some $\mathcal{A} \preccurlyeq \mathcal{B}$ with $D \subseteq A$ and $\kappa \leqslant |D| \leqslant |A| \leqslant \max\{|\mathcal{L}D|, |D|\} = \kappa$. Hence $|A| = \kappa$. From $C \subseteq D$ and $\mathcal{A} \equiv_{\mathcal{L}D} \mathcal{B} \vDash D_{el}\mathcal{C}$ it follows that $\mathcal{A} \vDash D_{el}\mathcal{C}$. Since also $C \subseteq D \subseteq A$, the $\mathcal{L}$-reduct of $\mathcal{A}$ is an elementary extension of $\mathcal{C}$. ❑

These theorems show in particular that a countable theory T with at least one infinite model also has models in every infinite cardinality. Further, $\vdash_T \alpha$ already holds when merely $\mathcal{A} \vDash \alpha$ for all T-models $\mathcal{A}$ of a *single* infinite cardinal number κ, provided T has only infinite models, because under this assumption every T-model is elementarily equivalent to a T-model of cardinality κ.

Exercises

1. Let $\mathcal{A} \preccurlyeq \mathcal{C}$ and $\mathcal{B} \preccurlyeq \mathcal{C}$, where $A \subseteq B$. Prove that $\mathcal{A} \preccurlyeq \mathcal{B}$.

2. An embedding $\imath : \mathcal{A} \to \mathcal{B}$ is termed *elementary* if $\imath\mathcal{A} \preccurlyeq \mathcal{B}$, where $\imath\mathcal{A}$ denotes the image of $\mathcal{A}$ under $\imath$. Show similarly to Theorem 1.1 that an $\mathcal{L}A$-structure $\mathcal{B}$ is a model of $D_{el}\mathcal{A}$ iff $\mathcal{A}$ is elementarily embeddable into $\mathcal{B}$.

3. Let $a_1, \ldots, a_n \in \mathbb{Q}$ and $b \in \mathbb{R}$. Show that there is an automorphism of $(\mathbb{R}, <)$ that maps b to a rational number and leaves all a_i fixed.

4. Let $\mathcal{A} \equiv \mathcal{B}$. Construct a structure $\mathcal{C}$ such that $\mathcal{A}, \mathcal{B}$ are both elementarily embeddable into $\mathcal{C}$.

5. Let $\mathcal{A}$ be an $\mathcal{L}$-structure generated from $G \subseteq A$ and $\mathcal{T}_G$ the set of ground terms in $\mathcal{L}G$. Prove that (a) for every $a \in A$ there is some $t \in \mathcal{T}_G$ such that $a = t^{\mathcal{A}}$, (b) if $\mathcal{A} \vDash T$ and $D\mathcal{A} \vdash_T \alpha$ $(\in \mathcal{L}G)$ then $D_G\mathcal{A} \vdash_T \alpha$. Here $D_G\mathcal{A} := D\mathcal{A} \cap \mathcal{L}G$.

5.2 Complete and κ-Categorical Theories

According to the definition on page 82, a theory $T \subseteq \mathcal{L}^0$ is complete if it is consistent and each extended theory $T' \supset T$ in $\mathcal{L}^0$ is inconsistent. A complete theory need not be maximally consistent in the whole of $\mathcal{L}$. For instance, even if T is complete, in general neither $\vdash_T x = y$ nor $\vdash_T x \neq y$. Some equivalent formulations of completeness, whose usefulness depend on the situation at hand, are presented by

Theorem 2.1. *For a consistent theory T the following conditions are equivalent:*[3]

(i) *T is complete,*
(ii) *$T = Th\,\mathcal{A}$ for every $\mathcal{A} \models T$,*
(iii) *$\mathcal{A} \equiv \mathcal{B}$ for all $\mathcal{A}, \mathcal{B} \models T$,*
(iv) *$\vdash_T \alpha \vee \beta \;\Rightarrow\; \vdash_T \alpha$ or $\vdash_T \beta$ $(\alpha, \beta \in \mathcal{L}^0)$,*
(v) *$\vdash_T \alpha$ or $\vdash_T \neg\alpha$ $(\alpha \in \mathcal{L}^0)$.*

Proof. (i) $\Rightarrow$ (ii): Since $T \subseteq Th\,\mathcal{A}$ for each model $\mathcal{A} \models T$, it must be that $T = Th\,\mathcal{A}$. (ii) $\Rightarrow$ (iii): For $\mathcal{A}, \mathcal{B} \models T$ we have by (ii) $Th\,\mathcal{A} = T = Th\,\mathcal{B}$, and therefore $\mathcal{A} \equiv \mathcal{B}$. (iii) $\Rightarrow$ (iv): Let $\vdash_T \alpha \vee \beta$, $\mathcal{A} \models T$, and $\mathcal{A} \models \alpha$, say. Then $\mathcal{B} \models \alpha$ for all $\mathcal{B} \models T$ by (iii), hence $\vdash_T \alpha$. (v) is a special case of (iv) because $\vdash_T \alpha \vee \neg\alpha$, for arbitrary $\alpha \in \mathcal{L}^0$. (v) $\Rightarrow$ (i): Let $T' \supset T$ and $\alpha \in T' \backslash T$. Then $\vdash_T \neg\alpha$ by (v); hence also $\vdash_{T'} \neg\alpha$. But then T' is inconsistent. Hence, by the above definition, T is complete. ❏

We now present various methods by which conjectured completeness can be confirmed. The completeness question is important for many reasons. For example, according to Theorem 3.5.2, a complete axiomatizable theory is decidable whatever the means of proving completeness might have been.

An elementary theory with at least one infinite model, even if it is complete, has many different infinite models. For instance, according to Theorem 1.6, the theory possesses models of arbitrarily high cardinality. However, sometimes it happens that all of its models of a given finite or infinite cardinal number κ are isomorphic. The following definition bears this circumstance in mind.

Definition. A theory T is *κ-categorical* if there exists up to isomorphism precisely one T-model of cardinality κ.

Example 1. The theory $\mathsf{Taut}_=$ of tautological sentences in $\mathcal{L}_=$ is κ-categorical for every cardinal κ. Indeed, here models $\mathcal{A}, \mathcal{B}$ of cardinality κ are naked sets and these are trivially isomorphic under any bijection from A onto B.

The theory DO of *densely ordered sets* results from the theory of ordered sets (formalized in **2.3**; see also **2.1**) by adjoining the axioms

$$\exists x \exists y\, x \neq y\,; \quad \forall x \forall y \exists z (x < y \to x < z \wedge z < y).$$

[3] All these conditions are also equivalent (they all hold) if the inconsistent theory is taken to be complete, which is not the case here as we already agreed upon in **3.3**.

It is easily seen that a densely ordered set is infinite. DO can be extended by adjoining the axioms $\mathsf{L} := \exists x \forall y\, x \leqslant y$ and $\mathsf{R} := \exists x \forall y\, y \leqslant x$ to the theory DO_{11} of densely ordered sets with edge elements. Replacing R by $\neg\mathsf{R}$ results in the theory DO_{10} of densely ordered sets with left but without right edge element. Accordingly DO_{01} denotes the theory with right but without left, and DO_{00} that of dense orders without any edge elements. The paradigm for DO_{00} is $(\mathbb{Q}, <)$.

Example 2. DO_{00} is $\aleph_0$-categorical (Exercise 1 treats the other DO_{ij}). The following proof is due to Cantor. A function f with $dom\, f \subseteq M$ and $ran\, f \subseteq N$ is said to be a *partial function* from M to N. Let $A = \{a_0, a_1, \dots\}$ and $B = \{b_0, b_1, \dots\}$ be countable DO_{00}-models. Define f_0 by $f_0 a_0 = b_0$ so that $dom\, f_0 = \{a_0\}$, $ran\, f_0 = \{b_0\}$ (construction step 0). Assume that in the nth step a partial function f_n *from* A to B with finite domain was constructed with $a < a' \Leftrightarrow f_n a < f_n a'$, for all $a, a' \in dom\, f_n$ (a so-called *partial isomorphism*), and that $\{a_0, \dots, a_n\} \subseteq dom\, f_n$ and $\{b_0, \dots, b_n\} \subseteq ran\, f_n$. These conditions are trivially satisfied for f_0. Let m be minimal with $a_m \in A \setminus dom\, f_n$. Choose $b \in B \setminus ran\, f_n$ such that $g_n := f_n \cup \{(a_m, b)\}$ is also a partial isomorphism. This is possible thanks to the denseness of B. Now let m be minimal with $b_m \in B \setminus ran\, g_n$. Choose a suitable $a \in A \setminus dom\, g_n$ such that $f_{n+1} := g_n \cup \{(a, b_m)\}$ is a partial isomorphism too. This "to and fro" construction provides both for $a_{n+1} \in dom\, f_{n+1}$ and $b_{n+1} \in ran\, f_{n+1}$ as is readily seen. Claim: $f = \bigcup_{n\in\mathbb{N}} f_n$ is an isomorphism from A onto B. Indeed, f is a function. Moreover, $dom\, f = A$ and $ran\, f = B$. The isomorphism condition $x < y \Leftrightarrow fx < fy$ is obvious.

Example 3. The *successor theory* T_{suc} in $\mathcal{L}\{0, \mathsf{S}\}$ has the axioms

$$\forall x\, 0 \neq \mathsf{S}x, \quad \forall xy(\mathsf{S}x = \mathsf{S}y \to x = y), \quad (\forall x \neq 0)\exists y\, x = \mathsf{S}y,$$

$$\forall x_0 \cdots x_n \Big(\bigwedge\nolimits_{i<n} \mathsf{S}x_i = x_{i+1} \to x_0 \neq x_n\Big) \quad (n = 1, 2, \dots, \text{ there are no "circles"}).$$

T_{suc} is not $\aleph_0$-categorical, but it is $\aleph_1$-categorical. Indeed, each model $\mathcal{A} \vDash T_{\mathrm{suc}}$ with $|\mathcal{A}| = \aleph_1$ consists up to isomorphism of the (countable) standard model $(\mathbb{N}, 0, \mathsf{S})$ and $\aleph_1$ many "threads" of isomorphism type $(\mathbb{Z}, \mathsf{S})$ where $\mathsf{S}: z \mapsto z + 1$. For if there were only countably many such threads then the entire model would be countable. It now easily follows that any two T_{suc}-models of cardinality $\aleph_1$ are isomorphic.

Example 4. The theory ACF_p of a.c. fields of given characteristic p (page 82) is $\aleph_1$-categorical. We sketch here a proof very briefly because ACF_p is analyzed in **5.5** in a different way. The claim follows from the facts that each field is embeddable into an a.c. field (cf. Example 1 of **5.5**) and that a *transcendental extension* $\mathcal{K}'$ of a field $\mathcal{K}$ (that is, every $a \in K' \setminus K$ is transcendental on K) has a transcendence basis B, that is, a maximal system of algebraically independent elements in $K' \setminus K$. The isomorphism type of $\mathcal{K}'$ is completely determined by the cardinality of B.

It is fairly plausible that in Examples 3 and 4 κ-categoricity holds for every cardinal $\kappa > \aleph_0$. This is no coincidence. It is explained by the following theorem.

Morley's theorem. *If a countable theory T is κ-categorical for some $\kappa > \aleph_0$ then it is κ-categorial for all $\kappa > \aleph_0$.*

The proof makes use of extensive methods and must be passed over here. On the other hand, the proof of the following theorem requires but little effort.

Theorem 2.2 (Vaught's test). *A countable consistent theory T without finite models is complete provided it is κ-categorical for some κ.*

Proof. Note first that $\kappa \geqslant \aleph_0$ because T possesses no finite models. Assume that T is incomplete. Choose some $\alpha \in \mathcal{L}^0$ with $\nvdash_T \alpha$ and $\nvdash_T \neg\alpha$. Then T, α and $T, \neg\alpha$ are consistent. These sets have countable infinite models by Theorem 1.5, and according to Theorem 1.6 there are also models $\mathcal{A}$ and $\mathcal{B}$ of cardinal κ. Since $\mathcal{A}, \mathcal{B} \vDash T$, by hypothesis $\mathcal{A} \simeq \mathcal{B}$, hence $\mathcal{A} \equiv \mathcal{B}$, which contradicts $\mathcal{A} \vDash \alpha$ and $\mathcal{B} \vDash \neg\alpha$. ❑

Example 5. (a) The theory DO_{00} of densely ordered sets without edge elements has only infinite models and is $\aleph_0$-categorical by Example 2. Hence it is complete by Vaught's test, confirming $(\mathbb{Q}, <) \equiv (\mathbb{R}, <)$ once again. Each DO_{ij} is a complete theory (Exercise 1). This clearly implies $\mathcal{A} \equiv \mathcal{B}$ for $\mathcal{A}, \mathcal{B} \vDash \mathsf{DO}$ iff $\mathcal{A}, \mathcal{B}$ have "the same edge configuration." Each of the DO_{ij}, being a complete axiomatizable theory, is hence decidable. Therefore, by Exercise 3 in **3.5**, the same is true for DO.
(b) The successor theory T_{suc} is $\aleph_1$-categorical (Example 3) and has only infinite models. Hence it is complete and as an axiomatizable theory thus decidable.
(c) ACF_p is $\aleph_1$-categorical by Example 4. Each a.c. field $\mathcal{A}$ is infinite. For assume the converse, that is, $A = \{a_0, \ldots, a_n\}$. Then the polynomial $1 + \prod_{i \leqslant n}(x - a_i)$ would have no root. Hence, by Vaught's test ACF_p is complete and decidable (since it is axiomatizable). This result will be derived by quite different methods in **5.5**.

The model classes of sentences are called *elementary classes*. These clearly include the model classes of finitely axiomatizable elementary theories. For any theory T, $\operatorname{Md} T = \bigcap_{\alpha \in T} \operatorname{Md} \alpha$ is an intersection of elementary classes, also termed an Δ-*elementary class*. Thus, the class of all fields is elementary, and that of all a.c. fields is Δ-elementary. On the other hand, the class of all finite fields is not Δ-elementary because its theory evidently has infinite models. An algebraic characterization of elementary and Δ-elementary classes will be provided in **5.7**.

The model classes of complete theories are called *elementary types*. $\operatorname{Md} T$ is the union of the elementary types belonging to the completions of a theory T. For instance, DO has just the four completions DO_{ij} determined by the edge configuration, that is, by those of the sentences $\mathsf{L}, \mathsf{R}, \neg\mathsf{L}, \neg\mathsf{R}$, valid in the respective completion. For this case, the next theorem provides more information.

Let $X \subseteq \mathcal{L}$ be nonempty and T a theory. Take $\langle X \rangle$ to denote the set (still dependent on T) of all formulas equivalent in T to Boolean combinations of formulas in X. Clearly, $\top \in \langle X \rangle$ since $\top \equiv_T \alpha \vee \neg\alpha$ for $\alpha \in X$. Therefore, $T \subseteq \langle X \rangle$, because

$\alpha \equiv_T \top$ whenever $\alpha \in T$. Call $X \subseteq \mathcal{L}^0$ a *Boolean basis for* $\mathcal{L}^0$ *in* T if *every* $\alpha \in \mathcal{L}^0$ belongs to $\langle X \rangle$. $\mathcal{A} \equiv_X \mathcal{B}$ is to mean $\mathcal{A} \models \alpha \Leftrightarrow \mathcal{B} \models \alpha$, for all $\alpha \in X$. Example 6(b) below indicates how useful a Boolean base for decision problems can be.

Theorem 2.3 (Basis theorem for sentences). *Let T be a theory and $X \subseteq \mathcal{L}^0$ a set of sentences such that $\mathcal{A} \equiv_X \mathcal{B} \Rightarrow \mathcal{A} \equiv \mathcal{B}$, for all $\mathcal{A}, \mathcal{B} \models T$.*[4] *Then X is a Boolean basis for $\mathcal{L}^0$ in T.*

Proof. Suppose $\alpha \in \mathcal{L}^0$ and $Y_\alpha := \{\beta \in \langle X \rangle \mid \alpha \vdash_T \beta\}$. We claim that $(*)$: $Y_\alpha \vdash_T \alpha$. Otherwise let $\mathcal{A} \models T, Y_\alpha, \neg\alpha$. Then $T_X\mathcal{A} := \{\gamma \in \langle X \rangle \mid \mathcal{A} \models \gamma\} \vdash \neg\alpha$; indeed for any $\mathcal{B} \models T_X\mathcal{A}$ we have $\mathcal{B} \equiv_X \mathcal{A}$ and hence $\mathcal{B} \equiv \mathcal{A}$. Therefore $\gamma \vdash_T \neg\alpha$ for some $\gamma \in T_X\mathcal{A}$, because $\langle X \rangle$ is closed under conjunctions. This yields $\alpha \vdash_T \neg\gamma$, i.e., $\neg\gamma \in Y_\alpha$. Thus $\mathcal{A} \models \neg\gamma$, in contradiction to $\mathcal{A} \models \gamma$. So $(*)$ holds. Hence there are $\beta_0, \ldots, \beta_m \in Y_\alpha$ such that $\beta := \bigwedge_{i \leqslant m} \beta_i \vdash_T \alpha$. We know $\alpha \vdash_T \beta_i$ and so $\alpha \vdash_T \beta$ as well. This and $\beta \vdash_T \alpha$ confirms $\alpha \equiv_T \beta$, and since $\beta \in \langle X \rangle$, also $\alpha \in \langle X \rangle$. ❏

Example 6. (a) For $T = \mathsf{DO}$ and $X = \{\mathsf{L}, \mathsf{R}\}$ it holds that $\mathcal{A} \equiv_X \mathcal{B} \Rightarrow \mathcal{A} \equiv \mathcal{B}$, for all $\mathcal{A}, \mathcal{B} \models T$. Indeed, $\mathcal{A} \equiv_X \mathcal{B}$ states that $\mathcal{A}, \mathcal{B}$ possess the same edge configuration. But then $\mathcal{A} \equiv \mathcal{B}$, because the DO_{ij} are all complete; see Example 5(a). Therefore, L and R form a Boolean basis for $\mathcal{L}^0_<$ in DO. This theory has four completions, and so by Exercise 3 in **3.6**, exactly 15 $(= 2^4 - 1)$ consistent extensions.
(b) Let $T = \mathsf{ACF}$ and $X = \{\mathtt{char}_p \mid p \text{ prime}\}$. Again, $\mathcal{A} \equiv_X \mathcal{B} \Rightarrow \mathcal{A} \equiv \mathcal{B}$, for all $\mathcal{A}, \mathcal{B} \models T$, because by Example 5(c) ACF_p is complete for each p (including $p = 0$). Hence, by Theorem 2.3, the $\mathtt{char}_p$ constitute a Boolean basis for sentences modulo ACF. This implies the decidability of ACF: let $\alpha \in \mathcal{L}^0$ be given; just wait in an enumeration process of the theorems of ACF until a sentence of the form $\alpha \leftrightarrow \beta$ appears, where β is a Boolean combination of the $\mathtt{char}_p$. Such a sentence definitely appears. Then test whether $\beta \equiv_{\mathsf{ACF}} \top$, for example by converting β into a CNF.

Corollary 2.4. *Let $T \subseteq \mathcal{L}^0$ be a theory with arbitrarily large finite models, such that all finite T-models with the same number of elements and all infinite T-models are elementarily equivalent. Then it holds that*

(a) *the sentences $\exists_n$ form a Boolean basis for $\mathcal{L}^0$ in T,*

(b) *T is decidable provided T is finitely axiomatizable.*

Proof. Let $X := \{\exists_k \mid k \in \mathbb{N}\}$. Then by hypothesis, $\mathcal{A} \equiv_X \mathcal{B} \Rightarrow \mathcal{A} \equiv \mathcal{B}$, for all $\mathcal{A}, \mathcal{B} \models T$. Thus, (a) follows by Theorem 2.3. By Exercise 4 in **2.3** each α compatible with T is therefore equivalent in T to a formula of the form $\bigvee_{\nu \leqslant n} \exists_{=k_\nu}$ or $\bigvee_{\nu \leqslant n} \exists_{=k_\nu} \vee \exists_m$. Then both sentences clearly have a finite T-model. In other words, T has the finite model property. Thus, (b) holds by Exercise 3 in **3.6**. ❏

[4] This assumption is equivalent to the assertion $\{\gamma \in \langle X \rangle \mid \mathcal{A} \models \gamma\}$ is complete; see the subsequent proof. For refinements of the theorem we refer to [HR].

Simple examples of applications are the theories $\mathsf{Taut}_=$ and the theory FO of all finite ordered sets. It is proved in the next section that the latter theory satisfies the hypothesis of the corollary. The equivalent formulas mentioned in the proof also permit a complete description of the elementary classes of $\mathcal{L}_=$. These are finite unions of classes determined by sentences of the form $\exists_{=k}$ and $\exists_m$. The elementary classes of FO-models admit a description of similar simplicity.

These examples make the following sufficiently clear: If we know the elementary types of a theory T then we also know their elementary classes. As a rule the type classification, that is, finding an appropriate set X satisfying the hypothesis of Theorem 2.3, is successful only in particular cases. The required work tends to be extensive. We mention in this regard the theories of abelian groups, of Boolean algebras, and of other locally finite varieties; see for instance [MV]. The above examples are just the simplest ones.

Easy to deal with is the case of an incomplete theory T that has finitely many completions. Example 6(a) is just a special case. According to Exercise 3 in **3.5**, T then has finitely many extensions. Moreover, all these are *finite* extensions. Indeed, if $T + \{\alpha_i \mid i \in \mathbb{N}\}$ is a nonfinite extension then w.l.o.g. $\bigwedge_{i<n} \alpha_i \nvdash_T \alpha_n$, which obviously implies that T has infinitely many completions, contradicting our hypothesis. Thus, we may assume that $T_1, \dots, T_m$ are the completions of T and that $T_i = T + \alpha_i$ for some $\alpha_i \in \mathcal{L}^0$. Then $\{\alpha_1, \dots, \alpha_m\}$ is a Boolean basis for $\mathcal{L}^0$ in T. Exercise 4 provides a canonical axiomatization of all consistent extensions of T.

Exercises

1. Prove that also DO_{10}, DO_{11}, and DO_{01} are $\aleph_0$-categorical and hence complete. In addition, verify that these and DO_{00} are the only completions of DO.

2. Prove that T_{suc} (page 138) is also completely axiomatized by the first two given axioms plus IS: $\varphi\frac{0}{x} \wedge \forall x(\varphi \to \varphi\frac{\mathtt{S}x}{x}) \to \forall x\varphi$; here φ runs over all formulas of the language $\mathcal{L}\{0, \mathtt{S}\}$ (the "induction schema" for $\mathcal{L}\{0, \mathtt{S}\}$).

3. Show that the theory T of torsion-free divisible abelian groups is $\aleph_1$-categorical and complete, hence decidable. This shows, in particular, the elementary equivalence of the groups $(\mathbb{R}, 0, +)$ and $(\mathbb{Q}, 0, +)$.

4. Let $T + \alpha_1, \dots, T + \alpha_m$ be all completions of T. Prove that $T + \bigvee_{1 \leqslant \nu \leqslant n} \alpha_{i_\nu}$ are all consistent extensions of T. Here $1 \leqslant n \leqslant m$ and $1 \leqslant i_0 < \dots < i_n \leqslant m$.

5. Show that an $\aleph_0$-categorical theory T with no finite models has an elementary prime model. Example: $(\mathbb{Q}, <)$ is an elementary prime model for DO_{00}.

5.3 Ehrenfeucht's game

Unfortunately, Vaught's criterion has only limited applications because many complete theories are not categorical in any transfinite cardinality. Let SO denote the *theory of discretely ordered sets*, i.e., of all $(M, <)$ such that every $a \in M$ has an immediate successor provided a is not the right edge element, and likewise an immediate predecessor provided a is not a left edge element. "SO" is intended to recall "step order," because the word "discrete" in connection with orders often has the stronger sense "each cut is a jump." SO_{ij} $(i.j \in \{0, 1\})$ is defined analogously to DO_{ij} (see page 138). For instance, SO_{10} is the theory of discretely ordered sets with left and without right edge element. Clearly $(\mathbb{N}, <)$ is a prime model for SO_{10}. The models of SO_{10} arise from arbitrary orders $(M, <)$ with a left edge element by replacing the latter by $(\mathbb{N}, <)$ and every other element of M by a specimen of $(\mathbb{Z}, <)$. From this it follows that SO_{10} cannot be κ-categorical for any $\kappa \geqslant \aleph_0$. Yet this theory is complete will be shown, and the same applies to SO_{00} and SO_{01}. Only SO_{11} is incomplete and is the only one of the four theories that has finite models. It coincides with the elementary theory of all finite ordered sets, Exercise 3.

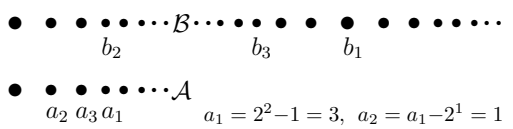
We prove the completeness of SO_{10} game-theoretically using a two-person game with players I and II, *Ehrenfeucht's game* $\Gamma_k(\mathcal{A}, \mathcal{B})$, which is played in k rounds, $k \geqslant 0$. The $\mathcal{A}, \mathcal{B}$ be given $\mathcal{L}$-structures and $\mathcal{L}$ a *relational language*, i.e., $\mathcal{L}$ does not contain any constant or operation symbols. With regard to our goal this presents no real loss of generality because each structure can be converted into a relational one by replacing its operations by the corresponding graphs. Another advantage of relational structures used in the sequel is that there is a bijective correspondence between subsets and substructures.

We now describe the game $\Gamma_k(\mathcal{A}, \mathcal{B})$. Player I chooses in each of the k rounds one of the two structures $\mathcal{A}$ and $\mathcal{B}$. If this is $\mathcal{A}$, he selects some $a \in A$. Then player II has to answer with some element $b \in B$. If player I chooses $\mathcal{B}$ and some b from B then player II must answer with some element $a \in A$. This is the entire game. After k rounds elements $a_1, \ldots, a_k \in A$ and $b_1, \ldots, b_k \in B$ have been selected, where a_i, b_i denote the elements selected in round i. Player II wins if the mapping $a_i \mapsto b_i$ $(i = 1, \ldots, k)$ is a partial isomorphism from $\mathcal{A}$ to $\mathcal{B}$; in other words, if the substructure of $\mathcal{A}$ with the domain $\{a_1, \ldots, a_k\}$ is isomorphic to the substructure of $\mathcal{B}$ with the domain $\{b_1, \ldots, b_k\}$. Otherwise, player I is the winner.

We write $\mathcal{A} \sim_k \mathcal{B}$ if player II has a winning strategy in the game $\Gamma_k(\mathcal{A}, \mathcal{B})$, that is, in every round player II can answer any move from player I such that at the end player II is the winner. For the "zero-round game" let $\mathcal{A} \sim_0 \mathcal{B}$ by definition.

Example. Let $\mathcal{A} = (\mathbb{N}, <)$ be a proper initial segment of $\mathcal{B} \vDash \mathsf{SO}_{10}$. We show that $\mathcal{A} \sim_k \mathcal{B}$ for arbitrary $k > 0$. Player II plays as follows: If player I chooses some b_1 in

B in the first round then player II answers with $a_1 = 2^{k-1}-1$ if $d(0,b_1) \geqslant 2^{k-1}-1$; otherwise with $a_1 = d(0,b_1)$.[5] The procedure is similar if player I begins with $\mathcal{A}$. If player I now selects some $b_2 \in B$ such that $d(0,b_2), d(b_1,b_2) \geqslant 2^{k-2}-1$, then player II answers with $a_2 = a_1 \pm 2^{k-2}$ depending on whether $b_2 > b_1$ or $b_2 < b_1$, and otherwise with the element of the same distance from 0 or a_1 as that of b_2 from 0 in B respectively from b_1. Similarly in the third round etc. The figure shows the course of a 3-round game played in the described way, in which player I has chosen from $\mathcal{B}$ only. With this strategy player II wins every game as can be shown by induction on k.



In contrast to the example, for $\mathcal{A} = (\mathbb{N}, <)$ and $\mathcal{B} = (\mathbb{Z}, <)$ player II's chances have already dropped in $\Gamma_2(\mathcal{A}, \mathcal{B})$ if player I selects $0 \in A$ in the first round. Player II will loose already in the 2nd round. This has to do with the fact that the existence of an edge element is expressible by a sentence of quantifier rank 2. We write $\mathcal{A} \equiv_k \mathcal{B}$ for $\mathcal{L}$-structures $\mathcal{A}, \mathcal{B}$ if $\mathcal{A} \vDash \alpha \Leftrightarrow \mathcal{B} \vDash \alpha$, for all $\alpha \in \mathcal{L}^0$ with $\operatorname{qr}\alpha \leqslant k$. It is always the case that $\mathcal{A} \equiv_0 \mathcal{B}$ for all $\mathcal{A}, \mathcal{B}$, because in relational languages there are no sentences of quantifier rank 0. Below we will prove the following remarkable

Theorem 3.1. *$\mathcal{A} \sim_k \mathcal{B}$ implies $\mathcal{A} \equiv_k \mathcal{B}$. Hence, $\mathcal{A} \equiv \mathcal{B}$ provided $\mathcal{A} \sim_k \mathcal{B}$ for all k.*

For finite signatures a somewhat weaker version of the converse of the theorem is valid as well, though we do not discuss this here. Before proving Theorem 3.1 we demonstrate its applicability. The theorem and the above example yield $(\mathbb{N}, <) \equiv_k \mathcal{B}$ for all k and hence $(\mathbb{N}, <) \equiv \mathcal{B}$ for every $\mathcal{B} \vDash \mathsf{SO}_{10}$, because $(\mathbb{N}, <)$ is a prime model for SO_{10}. Therefore SO_{10} is evidently complete. For reasons of symmetry the same holds for SO_{01}, and likewise for SO_{00}. On the other hand, SO_{11} has the finite model property according to Exercise 3. This readily implies that SO_{11} coincides with the theory FO of all finite ordered sets.

For the proof of Theorem 3.1 we first consider a minor generalization of $\Gamma_k(\mathcal{A}, B)$, *the game* $\Gamma_k(\mathcal{A}, \mathcal{B}, \vec{a}, \vec{b})$ *with prior moves* $\vec{a} \in A^n, \vec{b} \in B^n$. In the first round player I selects some $a_{n+1} \in A$ or $b_{n+1} \in B$ and player II answers with b_{n+1} or a_{n+1}, etc. The game protocol consists of sequences $(a_1, \ldots, a_{n+k})$ and $(b_1, \ldots, b_{n+k})$ at the end. Player II has won if $a_i \mapsto b_i$ $(i = 1, \ldots, n+k)$ is a partial isomorphism. Clearly, for $n = 0$ we obtain precisely the original game $\Gamma_k(\mathcal{A}, \mathcal{B})$.

This adjustment brings about an inductive characterization of a winning strategy for player II independent of more general concepts as follows:

[5] The "distance" $d(a,b)$ between elements a, b of some SO-model is 0 for $a = b$, 1 + the number of elements lying between a and b if it is finite, and $d(a,b) = \infty$ otherwise.

Definition. Player II has a winning strategy in $\Gamma_0(\mathcal{A}, \mathcal{B}, \vec{a}, \vec{b})$ provided $a_i \mapsto b_i$ for $i = 1, \ldots, n$ is a partial isomorphism. Player II has a winning strategy in $\Gamma_{k+1}(\mathcal{A}, \mathcal{B}, \vec{a}, \vec{b})$ if for every $a \in A$ there is some $b \in B$, and for every $b \in B$ some $a \in A$, such that player II has a winning strategy in $\Gamma_k(\mathcal{A}, \mathcal{B}, \vec{a}{}^\frown a, \vec{b}{}^\frown b)$. Here $\vec{c}{}^\frown c$ denotes the operation of appending the element c to the sequence $\vec{c}$.

We shall write $(\mathcal{A}, \vec{a}) \sim_k (\mathcal{B}, \vec{b})$ if player II has a winning strategy in $\Gamma_k(\mathcal{A}, \mathcal{B}, \vec{a}, \vec{b})$. In particular $\mathcal{A} \sim_k \mathcal{B}$, which is the case $\vec{a} = \vec{b} = \emptyset$, is now precisely defined.

Lemma 3.2. *Let $(\mathcal{A}, \vec{a}) \sim_k (\mathcal{B}, \vec{b})$ where $\vec{a} \in A^n$ and $\vec{b} \in B^n$. Then for all $\varphi = \varphi(\vec{x})$ with $\operatorname{qr}\varphi \leqslant k$ holds the equivalence $(*)$: $\mathcal{A} \vDash \varphi(\vec{a}) \Leftrightarrow \mathcal{B} \vDash \varphi(\vec{b})$.*

Proof by induction on k. Let $k = 0$. Since $a_i \mapsto b_i$ $(i = 1, \ldots, n)$ is a partial isomorphism, $(*)$ is valid for prime formulas and since the induction steps in the proof of $(*)$ for $\neg$, $\wedge$ are obvious; it is valid also for all formulas φ with $\operatorname{qr}\varphi = 0$. Now let $(\mathcal{A}, \vec{a}) \sim_{k+1} (\mathcal{B}, \vec{b})$. The only interesting case is $\varphi = \forall y \alpha(\vec{x}, y)$ such that $\operatorname{qr}\varphi = k + 1$, because every other formula of quantifier rank $k + 1$ is a Boolean combination of such formulas and formulas of quantifier rank $\leqslant k$ (Exercise 5 in **2.2**), and induction over $\neg$, $\wedge$ in proving $(*)$ is harmless. Assume $\mathcal{A} \vDash \forall y \alpha(\vec{a}, y)$ and $b \in B$. Then Player II chooses some $a \in A$ with $(\mathcal{A}, \vec{a}{}^\frown a) \sim_k (\mathcal{B}, \vec{b}{}^\frown b)$, so that according to the induction hypothesis, $\mathcal{A} \vDash \alpha(\vec{a}, a) \Leftrightarrow \mathcal{B} \vDash \alpha(\vec{b}, b)$. Clearly, the latter is supposed to hold for sequences $\vec{a}, \vec{b}$ of elements of arbitrary length. Because of $\mathcal{A} \vDash \alpha(\vec{a}, a)$, also $\mathcal{B} \vDash \alpha(\vec{b}, b)$. Since b was arbitrary we obtain $\mathcal{B} \vDash \forall y \alpha(\vec{b}, y)$. For reasons of symmetry, $\mathcal{B} \vDash \forall y \alpha(\vec{b}, y) \;\Rightarrow\; \mathcal{A} \vDash \forall y \beta(\vec{a}, y)$ holds as well. ❏

Theorem 3.1 is the application of the lemma for the case $n = 0$ and is therefore proved. The method illustrated is wide-ranging and has many generalizations.

Exercises

1. Let $\mathcal{A}, \mathcal{B}$ be two infinite densely ordered sets with the same edge configuration. Prove that $\mathcal{A} \sim_k \mathcal{B}$ for all k. Hence $\mathcal{A}, \mathcal{B}$ are elementarily equivalent.

2. Let $\mathcal{A}, \mathcal{B} \vDash \mathsf{SO}_{11}$, $k > 0$ and $|A|, |B| \geqslant 2^k - 1$. Prove that $\mathcal{A} \sim_k \mathcal{B}$, so that $\mathcal{A} \equiv_k \mathcal{B}$ according to Theorem 3.1.

3. Infer from Exercise 2 that SO_{11} has the finite model property and coincides with the elementary theory FO of all finite ordered sets.

4. Show that L, R, $\exists_1$, $\exists_2$, ... constitute a Boolean basis modulo SO and use this to prove the decidability of SO.[6]

[6] Moreover, the theory of all linear orders is decidable (Ehrenfeucht), and thus each of its finite extensions; but the proof is incomparably more difficult than for DO or SO.

5.4 Embedding and Characterization Theorems

Many of the foregoing theories, for instance those of orders, of groups in $\cdot, e, ^{-1}$, and of rings, are universal or $\forall$-theories. In other words, they possess axiom systems of $\forall$-sentences. We already know that for every theory T of this kind $\mathcal{A} \subseteq \mathcal{B} \vDash T$ implies $\mathcal{A} \vDash T$, in short T is **S**-*invariant*. DO obviously does not have this property, and so there cannot exist an axiom system of $\forall$-sentences for it. According to Theorem 4.3 the $\forall$-theories are completely characterized by the property of **S**-invariance. This presents a particularly simple example of the model-theoretical characterization of certain syntactic forms of axiom systems.

$T^\forall := \{\alpha \in T \mid \alpha \text{ is an } \forall\text{-sentence}\}$ is called the *universal part* of a theory T. Note the distinction between the set $T^\forall$ and the $\forall$-theory $T^\forall$, which of course contains more than just $\forall$-sentences. For $\mathcal{L}_0 \subseteq \mathcal{L}$ put $T_0^\forall := \mathcal{L}_0 \cap T^\forall$. If $\mathcal{A}$ is an $\mathcal{L}_0$-structure and $\mathcal{B}$ an $\mathcal{L}$-structure then $\mathcal{A} \subseteq \mathcal{B}$ or "$\mathcal{A}$ is a substructure of $\mathcal{B}$" will often mean in this section that $\mathcal{A}$ is a substructure of the $\mathcal{L}_0$-reduct of $\mathcal{B}$. The phrase "$\mathcal{A}$ is embeddable into $\mathcal{B}$" introduced in **5.1** is to be understood similarly. Examples will be found below. First we state the following

Lemma 4.1. *Every $T_0^\forall$-model $\mathcal{A}$ is embeddable into some T-model.*

Proof. It is enough to prove $(*)$: $T, D\mathcal{A}$ is consistent, because if $\mathcal{B} \vDash T, D\mathcal{A}$ then $\mathcal{A}$ is embeddable into $\mathcal{B}$ by Theorem 1.1. Assume $(*)$ is false. Then there is a conjunction $\varkappa(\vec{a})$ of sentences in $D\mathcal{A}$ such that $\varkappa(\vec{a}) \vdash_T \bot$, or equivalently, $\vdash_T \neg\varkappa(\vec{a})$. Here let $\vec{a}$ embrace all the constants of $\mathcal{L}A$ that appear in the members of $\varkappa$ but not in T. By the rule $(\forall)$ of constant-quantification from **3.2**, $\vdash_T \forall\vec{x}\neg\varkappa(\vec{x})$. Hence $\forall\vec{x}\neg\varkappa(\vec{x}) \in T_0^\forall$ and thus $\mathcal{A} \vDash \forall\vec{x}\neg\varkappa(\vec{x})$, contradicting $\mathcal{A} \vDash \varkappa(\vec{a})$. ❏

Lemma 4.2. $\operatorname{Md} T^\forall$ *consists of precisely the substructures of all T-models.*

Proof. Every substructure of a T-model is of course a $T^\forall$-model. Furthermore, each $\mathcal{A} \vDash T^\forall$ is (by Lemma 4.1 for $\mathcal{L}_0 = \mathcal{L}$) embeddable into some $\mathcal{B} \vDash T$, and this is surely equivalent to $\mathcal{B}' \simeq \mathcal{B}$ and $\mathcal{A} \subseteq \mathcal{B}'$ for some $\mathcal{B}' \vDash T$, because $\operatorname{Md} T$ is always closed under isomorphic images. ❏

Example. (a) Let AG be the theory of abelian groups in $\mathcal{L}\{\circ\}$. A substructure of $\mathcal{A} \vDash \mathsf{AG}$ is obviously a commutative regular semigroup. Conversely, it is not hard to prove that every such semigroup is embeddable into an abelian group. Therefore, the theory $\mathsf{AG}^\forall$ coincides with the theory of the commutative regular semigroups. *Warning*: noncommutative regular semigroups need not be embeddable into groups.

(b) Substructures of fields in $\mathcal{L}\{0, 1, +, -, \cdot\}$ are integral domains. Conversely, according to a well-known construction every integral domain is embeddable into a field, its *quotient field*. It is constructed similarly to the field $\mathbb{Q}$ from the ring $\mathbb{Z}$.

By Lemma 4.2, the theories T_J of integral domains and T_F of fields have the same universal part which is axiomatized by the axioms for T_J, i.e., the axioms for commutative rings with 1 and without zero-divisors. Also ACF has the same universal part, because every field is embeddable into some algebraically closed field.

Theorem 4.3. *T is a universal theory if and only if T is* **S***-invariant.*

Proof. This follows immediately from Lemma 4.2, since for an **S**-invariant theory T holds that $\operatorname{Md} T = \operatorname{Md} T^{\forall}$. In other words, T is axiomatized by $T^{\forall}$. ❑

This theorem is reminiscent of the **HSP** theorem cited on page 104. However, the latter concerns identities only. It has a different proof that is akin to the proof of the following remarkable theorem. It concerns universal Horn theories introduced in **4.1**. Call T **SP***-invariant* if $\operatorname{Md} T$ is closed under direct products and substructures. Always remember that a statement like $\mathcal{A} \models \varphi(\vec{a})$ with $\vec{a} \in A^n$ is to mean $\mathcal{A}_A \models \varphi(\vec{a})$, or equivalently, $\mathcal{A} \models \varphi(\vec{x})\,[\vec{a}]$.

Theorem 4.4. *T is a universal Horn theory if and only if T is* **SP***-invariant.*

Proof. $\Rightarrow$: Exercise 1 in **4.1**. $\Leftarrow$: Trivial if $\vdash_T \forall xy\, x = y$, for then T is axiomatized by $\forall xy\, x = y$. Let T be nontrivial. Put $U = \{\alpha \in T \mid \alpha \text{ a universal Horn sentence}\}$. We shall prove $\operatorname{Md} T = \operatorname{Md} U$. Only $\operatorname{Md} U \subseteq \operatorname{Md} T$ is not obvious. Let $\mathcal{A} \models U$. To verify $\mathcal{A} \models T$ it suffices to show $(*)$: $T \cup D\mathcal{A} \nvdash \bot$, since for $\mathcal{B} \models T, D\mathcal{A}$ w.l.o.g. $\mathcal{A} \subseteq \mathcal{B}$, so $\mathcal{A} \models T$ thanks to **S**-invariance. Let $P := \{\pi \in D\mathcal{A} \mid \pi \text{ prime}\}$, so that $D\mathcal{A} = P \cup \{\neg\pi_i \mid i \in I\}$ for some $I \neq \emptyset$, all π_i prime. We first show $\binom{*}{*}$: $P \nvdash_T \pi_i$ for all $i \in I$. Indeed, otherwise $\vdash_T \varkappa(\vec{a}) \to \pi_i(\vec{a})$ for some conjunction $\varkappa(\vec{a})$ of sentences in P, with the tuple $\vec{a}$ of constants not in T. Therefore $\vdash_T \alpha := \forall\vec{x}(\varkappa(\vec{x}) \to \pi_i(\vec{x}))$. Hence $\alpha \in U$, for α is a universal Horn sentence, whence $\mathcal{A} \models \alpha$. But this contradicts $\mathcal{A} \models \varkappa(\vec{a}) \wedge \neg\pi_i(\vec{a})$ and confirms $\binom{*}{*}$. From $\binom{*}{*}$ follows $(*)$ because if $\mathcal{A}_i \models T, P, \neg\pi_i$, then we have $\mathcal{B} := \prod_{i\in I} \mathcal{A}_i \models T$ and $\mathcal{B} \models P \cup \{\neg\pi_i \mid i \in I\} = D\mathcal{A}$. ❑

The following application of Lemma 4.1 aims in a somewhat different direction.

Theorem 4.5. *Let $\mathcal{L}_0 \subseteq \mathcal{L}$ and $\mathcal{A}$ be an $\mathcal{L}_0$-structure. For $T \subseteq \mathcal{L}^0$ are equivalent:*

(i) *$\mathcal{A}$ is embeddable into some T-model,*

(ii) *any finitely generated substructure of $\mathcal{A}$ is embeddable into a T-model,*

(iii) *$\mathcal{A} \models T_0^{\forall}$ $(= \mathcal{L}_0 \cap T^{\forall})$.*

Proof. (i)$\Rightarrow$(ii): Trivial. (ii)$\Rightarrow$(iii): Let $\forall\vec{x}\alpha \in T_0^{\forall}$ with $\alpha = \alpha(\vec{x})$ quantifier-free, $\vec{x} = (x_1, \ldots, x_n)$ w.l.o.g. $\neq \emptyset$. Let $\mathcal{A}_0$ for $\vec{a} = (a_1, \ldots, a_n) \in A^n$ be the substructure in $\mathcal{A}$ generated from $a_1, \ldots, a_n$. By (ii), $\mathcal{A}_0 \subseteq \mathcal{B}$ for some model $\mathcal{B} \models T$. Since $\mathcal{B} \models \forall\vec{x}\alpha$, it holds that $\mathcal{A}_0 \models \forall\vec{x}\alpha$; therefore $\mathcal{A}_0 \models \alpha(\vec{a})$, so that $\mathcal{A} \models \alpha(\vec{a})$ by Theorem 2.3.2. Since both $\forall\vec{x}\alpha \in T_0^{\forall}$ and $\vec{a} \in A^n$ were choosen arbitrarily, $\mathcal{A} \models \forall\vec{x}\alpha$. (iii)$\Rightarrow$(i): This is exactly the claim of Lemma 4.1. ❑

Examples of applications. (a) Let T be the theory of ordered abelian groups in $\mathcal{L} = \mathcal{L}\{0, +, -, <\}$. Such a group is clearly torsion-free, which is expressed by a schema of $\forall$-sentences in $\mathcal{L}_0 = \mathcal{L}\{0, +, -\}$. Conversely, Theorem 4.5 implies that a torsion-free abelian group (the $\mathcal{A}$ in the theorem) is orderable, or what amounts to the same thing, is embeddable into an ordered abelian group. One needs to show only that every finitely generated torsion-free abelian group G is orderable. By a well-known result from group theory, $G \simeq \mathbb{Z}^n$ for some $n > 0$. But $\mathbb{Z}^n$ can be ordered lexicographically as is easily seen by induction on n. For nonabelian groups, the conditions corresponding to torsion-freeness are somewhat more involved.

(b) Without needing algebraic methods we know that there exists a set of universal sentences in $0, 1, +, -, \cdot$, whose adoption to the theory of fields characterizes the orderable fields. Sufficient for this, by Theorem 4.5, is the set of all $\forall$-sentences in $0, 1, +, -, \cdot$ provable from the axioms for ordered fields. Indeed even the schema of sentences '-1 is not a sum of squares' is enough (E. Artin).

Not just $\forall$-theories but also $\forall$-formulas can be characterized model-theoretically. Call $\alpha(\vec{x})$ **S**-*persistent* or simply *persistent* in T if for all $\mathcal{A}, \mathcal{B} \vDash T$ with $\mathcal{A} \subseteq \mathcal{B}$

$$\text{(sp)} \quad \mathcal{B} \vDash \alpha(\vec{a}) \Rightarrow \mathcal{A} \vDash \alpha(\vec{a}), \text{ for all } \vec{a} \in A^n.$$

This property characterizes the $\forall$-formulas up to equivalence according to

Theorem 4.6. *If $\alpha = \alpha(\vec{x})$ is persistent in T then α is equivalent to some $\forall$-formula α' in T, which can be chosen in such a way that free$\,\alpha' \subseteq$ free$\,\alpha$.*

Proof. Let Y be the set of all formulas of the form $\forall\vec{y}\beta(\vec{x}, \vec{y})$ with $\alpha \vdash_T \forall\vec{y}\beta(\vec{x}, \vec{y})$ where β is quantifier-free; here the tuples $\vec{x}$ and $\vec{y}$ may be of length $n \geqslant 0$ and $m \geqslant 0$, respectively. We prove (a): $Y \vdash_T \alpha(\vec{x})$. This would complete the proof because there then exists, thanks to free$\,Y \subseteq \{x_1, \ldots, x_n\}$, a conjunction $\varkappa = \varkappa(\vec{x})$ of formulas from Y with $\varkappa \vdash_T \alpha$. Since also $\alpha \vdash_T \varkappa$, we have $\alpha \equiv_T \varkappa$, and since a conjunction of $\forall$-formulas $\varkappa$ is again equivalent to an $\forall$-formula, $\varkappa \in Y$. For proving (a) assume $(\mathcal{A}, \vec{a}) \vDash T, Y$ (or $\mathcal{A} \vDash T, Y\,[\vec{a}]$) with $\vec{a} \in A^n$. We need to show that $(\mathcal{A}, \vec{a}) \vDash \alpha$. This follows from (b): $T, \alpha(\vec{a}), D\mathcal{A}$ is consistent, for if $\mathcal{B} \vDash T, \alpha(\vec{a}), D\mathcal{A}$, then w.l.o.g. $\mathcal{A} \subseteq \mathcal{B}$; hence $\mathcal{A} \vDash \alpha(\vec{a})$ since α is persistent. If (b) were false then $\alpha(\vec{a}) \vdash_T \neg\varkappa(\vec{a}, \vec{b})$ for some conjunction $\varkappa(\vec{a}, \vec{b})$ of sentences from $D\mathcal{A}$ with the m-tuple $\vec{b}$ of constants of $\varkappa$ from $A \setminus \{a_1, \ldots, a_n\}$. Thus $\alpha(\vec{a}) \vdash_T \forall\vec{y}\neg\varkappa(\vec{a}, \vec{y})$. Since the $a_1, \ldots, a_n$ do not appear in T, we get $\alpha(\vec{x}) \vdash_T \forall\vec{y}\neg\varkappa(\vec{x}, \vec{y}) \in Y$. Therefore, $(\mathcal{A}, \vec{a}) \vDash \forall\vec{y}\neg\varkappa(\vec{x}, \vec{y})$, or equivalently $\mathcal{A} \vDash \forall\vec{y}\neg\varkappa(\vec{a}, \vec{y})$, in contradiction to $\mathcal{A} \vDash \varkappa(\vec{a}, \vec{b})$. ❏

Remark. Let T be countable and all T-models infinite. Then α is already equivalent in T to an $\forall$-formula, provided α is κ-*persistent*; this means that (sp) holds for all T-models $\mathcal{A}, \mathcal{B}$ of some fixed cardinal $\kappa \geqslant \aleph_0$. For in this case each T-model is elementarily equivalent to a model of cardinality κ by the Löwenheim–Skolem theorems. Hence, it suffices to verify (a) in the above proof by considering only models $\mathcal{A}, \mathcal{B}$ of cardinality κ.

Sentences of the form $\forall\vec{x}\,\exists\vec{y}\alpha$ with kernel α are called $\forall\exists$-*sentences*. Many theories, for instance of fields, of real or algebraically closed fields, of divisible groups, are $\forall\exists$-*theories*, i.e., they possess axiom systems of $\forall\exists$-sentences. We are going to characterize the $\forall\exists$-theories semantically. A *chain K of structures* is simply a set K of $\mathcal{L}$-structures such that $\mathcal{A} \subseteq \mathcal{B}$ or $\mathcal{B} \subseteq \mathcal{A}$ for all $\mathcal{A}, \mathcal{B} \in K$. Chains are often given as sequences $\mathcal{A}_0 \subseteq \mathcal{A}_1 \subseteq \mathcal{A}_2 \subseteq \cdots$ of structures. No matter how K is given, a structure $\mathcal{C} := \bigcup K$ can be defined in a natural way: Let $C := \bigcup\{A \mid \mathcal{A} \in K\}$ be its domain. Further let $r^{\mathcal{C}}\vec{a} \Leftrightarrow r^{\mathcal{A}}\vec{a}$ for $\vec{a} \in C^n$, where $\mathcal{A} \in K$ is chosen so that $\vec{a} \in A^n$. Such an $\mathcal{A} \in K$ exists: Let $\mathcal{A}$ simply be the maximum of the chain members containing $a_1, , \ldots, a_n$, respectively. The definition of $r^{\mathcal{C}}$ is independent on the choice of $\mathcal{A}$. Indeed, let $\mathcal{A}' \in K$ and $a_1, \ldots, a_n \in A'$. Since $\mathcal{A} \subseteq \mathcal{A}'$ or $\mathcal{A}' \subseteq \mathcal{A}$, it holds that $r^{\mathcal{A}}\vec{a} \Leftrightarrow r^{\mathcal{A}'}\vec{a}$ in either case. Finally, for function symbols f let $f^{\mathcal{C}}\vec{a} = f^{\mathcal{A}}\vec{a}$, where $\mathcal{A} \in K$ is chosen such that $\vec{a} \in A^n$. Here too the choice of $\mathcal{A} \in K$ is irrelevant. $\mathcal{C}$ was just defined in such a way that each $\mathcal{A} \in K$ is a substructure of $\mathcal{C}$.

Example 1. Let $\mathcal{D}_n$ be the additive group of n-place decimal numbers (with at most n decimals after the decimal point). Since $\mathcal{D}_n \subseteq \mathcal{D}_{n+1}$, the $\mathcal{D}_n$ form a chain. Here $\mathcal{D} = \bigcup_{n\in\mathbb{N}} \mathcal{D}_n$ is just the additive group of finite decimal numbers. The corresponding holds if the $\mathcal{D}_n$ are understood as ordered sets. Because then $\mathcal{D} \vDash \mathsf{DO}$, while $\mathcal{D}_n \vDash \mathsf{SO}$ for all n, $\operatorname{Md}\mathsf{SO}$ is not closed under union of chains.

It is easy to see that an $\forall\exists$-sentence $\alpha = \forall x_1 \ldots x_n \exists y_1 \ldots y_m \beta(\vec{x}, \vec{y})$ valid in all members $\mathcal{A}$ of a chain K of structures is also valid in $\mathcal{C} = \bigcup K$. For let $\vec{a} \in C^n$. Then $\vec{a} \in A^n$ for some $\mathcal{A} \in K$. Hence, there is some $\vec{b} \in A^m$ with $\mathcal{A} \vDash \beta(\vec{a}, \vec{b})$. Since $\mathcal{A} \subseteq \mathcal{C}$ and $\beta(\vec{x}, \vec{y})$ is open, it follows that $\mathcal{C} \vDash \beta(\vec{a}, \vec{b})$. Therefore, $\mathcal{C} \vDash \exists\vec{y}\beta(\vec{a}, \vec{y})$. Now, $\vec{a}$ is arbitrary here so that indeed $\mathcal{C} \vDash \forall\vec{x}\exists\vec{y}\beta(\vec{x}, \vec{y})$.

Thus, if T is an $\forall\exists$-theory, $\operatorname{Md} T$ is always closed under union of chains, or as it is said, T is *inductive*. Just this property is characteristic for $\forall\exists$-theories. However, the proof of this is no longer simple. It requires the notion of an *elementary chain*. This is a set K of $\mathcal{L}$-structures such that $\mathcal{A} \preccurlyeq \mathcal{B}$ or $\mathcal{B} \preccurlyeq \mathcal{A}$, for all $\mathcal{A}, \mathcal{B} \in K$. Clearly, K is then also a chain in the ordinary sense.

Lemma 4.7 (Tarski's chain lemma). *Let K be an elementary chain and put $\mathcal{C} = \bigcup K$. Then $\mathcal{A} \preccurlyeq \mathcal{C}$ for every $\mathcal{A} \in K$.*

Proof. We have to show that $\mathcal{A} \vDash \alpha(\vec{a}) \;\Leftrightarrow\; \mathcal{C} \vDash \alpha(\vec{a})$, with $\vec{a} \in A^n$. This follows by induction on $\alpha = \alpha(\vec{x})$ and is clear for prime formulas. The induction steps over $\wedge, \neg$ are also straightforward. Let $\mathcal{A} \vDash \forall y\alpha(y, \vec{a})$ and $a_0 \in C$ arbitrary. There is certainly some $\mathcal{B} \in K$ such that $a_0, \ldots, a_n \in \mathcal{B}$ and $\mathcal{A} \preccurlyeq \mathcal{B}$. Thus, $\mathcal{B} \vDash \forall y\alpha(y, \vec{a})$ and hence $\mathcal{B} \vDash \alpha(a_0, \vec{a})$. By the induction hypothesis (which is supposed to hold for *any* chain members) so too $\mathcal{C} \vDash \alpha(a_0, \vec{a})$. Since $a_0 \in C$ was arbitrary, $\mathcal{C} \vDash \forall y\alpha(y, \vec{a})$. The converse $\mathcal{C} \vDash \forall y\alpha(y, \vec{a}) \Rightarrow \mathcal{A} \vDash \forall y\alpha(y, \vec{a})$ follows similarly. ❑

We require yet another useful concept, found in many of the examples in **5.5**. Let $\mathcal{A} \subseteq \mathcal{B}$. Then $\mathcal{A}$ is termed *existentially closed in* $\mathcal{B}$, in symbols $\mathcal{A} \subseteq_{ec} \mathcal{B}$, provided

$$(*) \quad \mathcal{B} \vDash \exists \vec{x} \varphi(\vec{x}, \vec{a}) \Rightarrow \mathcal{A} \vDash \exists \vec{x} \varphi(\vec{x}, \vec{a}) \quad (\vec{a} \in A^n),$$

where $\varphi = \varphi(\vec{x}, \vec{a})$ runs through all conjunctions of literals from $\mathcal{L}A$. $(*)$ then holds automatically for all open $\varphi \in \mathcal{L}A$. One sees this straight away by converting φ into a disjunctive normal form and distributing $\exists \vec{x}$ over the disjuncts.

Clearly $\mathcal{A} \preccurlyeq \mathcal{B} \Rightarrow \mathcal{A} \subseteq_{ec} \mathcal{B} \Rightarrow \mathcal{A} \subseteq \mathcal{B}$. Moreover, $\subseteq_{ec}$ satisfies a readily proved chain lemma as well: If K is a chain of structures such that $\mathcal{A} \subseteq_{ec} \mathcal{B}$ or $\mathcal{B} \subseteq_{ec} \mathcal{A}$ for all $\mathcal{A}, \mathcal{B} \in K$, then $\mathcal{A} \subseteq_{ec} \bigcup K$ for every $\mathcal{A} \in K$. This is an easy exercise.

The next lemma presents various characterizations of $\mathcal{A} \subseteq_{ec} \mathcal{B}$. Let $D_\forall \mathcal{A}$ denote the *universal diagram* of $\mathcal{A}$, which is the set of all $\forall$-sentences of $\mathcal{L}A$ valid in $\mathcal{A}$. Clearly $D_\forall \mathcal{A} \subseteq D_{el} \mathcal{A}$. In (iii) the indexing of $\mathcal{B}$ with A is omitted to ease legibility.

Lemma 4.8. *Let $\mathcal{A}, \mathcal{B}$ be $\mathcal{L}$-structures and $\mathcal{A} \subseteq \mathcal{B}$. Then are equivalent*

(i) $\mathcal{A} \subseteq_{ec} \mathcal{B}$, (ii) *there is an* $\mathcal{A}' \supseteq \mathcal{B}$ *such that* $\mathcal{A} \preccurlyeq \mathcal{A}'$, (iii) $\mathcal{B} \vDash D_\forall \mathcal{A}$.

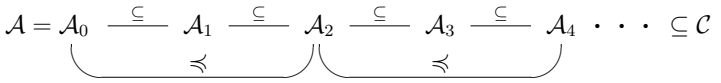
Proof. (i)⇒(ii): Let $\mathcal{A} \subseteq_{ec} \mathcal{B}$. We obtain some $\mathcal{A}' \supseteq \mathcal{B}$ such that $\mathcal{A} \preccurlyeq \mathcal{A}'$ as a model of $D_{el} \mathcal{A} \cup D\mathcal{B}$ (more precisely, as the $\mathcal{L}$-reduct of such a model), so that it remains only to show the consistency. Suppose the opposite, so that $D_{el} \mathcal{A} \vdash \neg \varkappa(\vec{b})$ for some conjunction $\varkappa(\vec{b})$ of members from $D\mathcal{B}$ with the n-tuple $\vec{b}$ of all constants of $B \setminus A$ in $\varkappa$. Since $b_1, \ldots, b_n$ do not occur in $D_{el} \mathcal{A}$, we get $D_{el} \mathcal{A} \vdash \forall \vec{x} \neg \varkappa(\vec{x})$. Thus $\mathcal{A} \vDash \forall \vec{x} \neg \varkappa(\vec{x})$. On the other hand $\mathcal{B} \vDash \varkappa(\vec{b})$; hence $\mathcal{B} \vDash \exists \vec{x} \varkappa(\vec{x})$. With (i) and $\varkappa(\vec{x}) \in \mathcal{L}A$ also $\mathcal{A} \vDash \exists \vec{x} \varkappa(\vec{x})$, in contradiction to $\mathcal{A} \vDash \forall \vec{x} \neg \varkappa(\vec{x})$. (ii)⇒(iii): Since $\mathcal{A} \preccurlyeq \mathcal{A}'$, we have $\mathcal{A}' \vDash D_{el} \mathcal{A} \supseteq D_\forall \mathcal{A}$. Since $\mathcal{B} \subseteq \mathcal{A}' \vDash D_\forall \mathcal{A}$, evidently $\mathcal{B} \vDash D_\forall \mathcal{A}$. (iii)⇒(i): By (iii), $\mathcal{A} \vDash \alpha \Rightarrow \mathcal{B} \vDash \alpha$, for all $\forall$-sentences α of $\mathcal{L}A$. The latter is equivalent to $\mathcal{B} \vDash \alpha \Rightarrow \mathcal{A} \vDash \alpha$, for all $\exists$-sentences of $\mathcal{L}A$ and hence to (i). ❏

Theorem 4.9. *A theory T is an $\forall\exists$-theory if and only if T is inductive.*

Proof. As already shown, an $\forall\exists$-theory T is inductive. Conversely let T be inductive. We show that $\operatorname{Md} T = \operatorname{Md} T^{\forall\exists}$, where $T^{\forall\exists}$ denotes the set of all $\forall\exists$-theorems provable in T. The nontrivial part is the verification of $\operatorname{Md} T^{\forall\exists} \subseteq \operatorname{Md} T$. So let $\mathcal{A} \vDash T^{\forall\exists}$. **Claim:** $T \cup D_\forall \mathcal{A}$ is consistent. Otherwise $\vdash_T \neg \varkappa$ for some conjunction $\varkappa = \varkappa(\vec{a})$ of sentences of $D_\forall \mathcal{A}$ with the tuple $\vec{a}$ of constants in A appearing in $\varkappa$ but not in T. Hence $\vdash_T \forall \vec{x} \neg \varkappa(\vec{x})$. Now, $\varkappa(\vec{x})$ is equivalent to an $\forall$-formula, and so $\neg \varkappa(\vec{x})$ to an $\exists$-formula. Thus, $\forall \vec{x} \neg \varkappa(\vec{x})$ belongs up to equivalence to $T^{\forall\exists}$. Therefore $\mathcal{A} \vDash \forall \vec{x} \neg \varkappa(\vec{x})$, which contradicts $\mathcal{A} \vDash \varkappa(\vec{a})$. This proves the claim.

Now let $\mathcal{A}_1 \vDash T \cup D_\forall \mathcal{A}$ and w.l.o.g. $\mathcal{A}_1 \supseteq \mathcal{A}$. Then also $\mathcal{A} \subseteq_{ec} \mathcal{A}_1$ in view of Lemma 4.8. By the same lemma there exists an $\mathcal{A}_2 \supseteq \mathcal{A}_1$ with $\mathcal{A}_0 := \mathcal{A} \preccurlyeq \mathcal{A}_2$, so that $\mathcal{A}_2 \vDash T^{\forall\exists}$ as well. We now repeat this construction with $\mathcal{A}_2$ in place of $\mathcal{A}_0$ and obtain structures $\mathcal{A}_3, \mathcal{A}_4$ such that $\mathcal{A}_2 \subseteq_{ec} \mathcal{A}_3 \vDash T$, $\mathcal{A}_3 \subseteq \mathcal{A}_4$ and $\mathcal{A}_2 \preccurlyeq \mathcal{A}_4$.

Continuing this construction produces a sequence $\mathcal{A}_0 \subseteq \mathcal{A}_1 \subseteq \mathcal{A}_2 \subseteq \cdots$ of structures with the inclusion relation illustrated in the following figure:



Let $\mathcal{C} := \bigcup_{i\in\mathbb{N}} \mathcal{A}_i$. Clearly also $\mathcal{C} = \bigcup_{i\in\mathbb{N}} \mathcal{A}_{2i}$, and because $\mathcal{A} = \mathcal{A}_0 \preccurlyeq \mathcal{A}_2 \preccurlyeq \cdots$ we get $\mathcal{A} \preccurlyeq \mathcal{C}$ by the chain lemma. At the same time we also have $\mathcal{C} = \bigcup_{i\in\mathbb{N}} \mathcal{A}_{2i+1}$, and since by construction $\mathcal{A}_{2i+1} \models T$ for all i, it holds that $\mathcal{C} \models T$, for T is inductive. But then too $\mathcal{A} \models T$ because $\mathcal{A} \preccurlyeq \mathcal{C}$. This is what we had to prove. ❑

A decent application of the theorem is that SO_{10} cannot be axiomatized by $\forall\exists$-axioms, for SO_{10} is not inductive according to Example 1. SO_{10} is an $\forall\exists\forall$-theory, and we see now that at least one $\forall\exists\forall$-axiom is needed in its axiomatization.

The "sandwich" construction in the proof of Theorem 4.9 can still be generalized. We will not elaborate on this but rather add some words about so-called model compatibility. Let $T_0 + T_1$ be the smallest theory containing T_0 and T_1. From the consistency of T_0 and T_1 we cannot infer that $T_0 + T_1$ is consistent, even if T_0 and T_1 are *model compatible* in the following sense: every T_0-model is embeddable into some T_1-model and vice versa. This property is equivalent to $T_0^\forall = T_1^\forall$ by Lemma 4.1, hence is an equivalence relation. Thus, the class of consistent $\mathcal{L}$-theories splits into disjoint classes of pairwise model compatible theories. That model compatible theories need not be compatible in the ordinary sense is shown by the following

Example 2. DO and SO are model compatible (Exercise 2) but $\mathsf{DO}+\mathsf{SO}$ is clearly inconsistent. Since DO is inductive, we get another argument that SO is not inductive: if it were inductive, $\mathsf{DO} + \mathsf{SO}$ would be consistent according to Exercise 3.

Exercises

1. Let X be a set of *positive* sentences, i.e., the $\alpha \in X$ are constructed from prime formulas by means of $\wedge$, $\vee$, $\forall$, $\exists$ only. Prove $\mathcal{A} \models X \Rightarrow \mathcal{B} \models X$, whenever $\mathcal{B}$ is a homomorphic image of $\mathcal{A}$, that is, $\operatorname{Md} X$ is closed under homomorphic images. Once again the converse holds (Lyndon's theorem; see [CK]).

2. Show that the theories DO and SO are model compatible.

3. Suppose T_0 and T_1 are model compatible and inductive. Show that $T_0 + T_1$ is an inductive theory which, in addition, is model compatible with T_0 and T_1.

4. For inductive T show that of all inductive extensions model compatible with T there exists a largest one, the *inductive completion* of T. For instance, this is ACF for the theory T_F of fields.

5.5 Model Completeness

After [Ro1], a theory T is called *model complete* if for every model $\mathcal{A} \models T$ the theory $T + D\mathcal{A}$ is complete in $\mathcal{L}A$. For $\mathcal{A}, \mathcal{B} \models T$ where $\mathcal{A} \subseteq \mathcal{B}$ (hence $\mathcal{B}_A \models D\mathcal{A}$) the completeness of $T + D\mathcal{A}$ obviously means the same as $\mathcal{A}_A \equiv \mathcal{B}_A$, or equivalently, $\mathcal{A} \preccurlyeq \mathcal{B}$. In short, a model complete theory T has the property

$$(*)\ \mathcal{A} \subseteq \mathcal{B} \Rightarrow \mathcal{A} \preccurlyeq \mathcal{B}, \text{ for all } \mathcal{A}, \mathcal{B} \models T.$$

Conversely, if $(*)$ is satisfied then $T + D\mathcal{A}$ is also complete. Indeed, let $\mathcal{B} \models T, D\mathcal{A}$ so that w.l.o.g. $\mathcal{A} \subseteq \mathcal{B}$ and hence $\mathcal{A} \preccurlyeq \mathcal{B}$. But then all these $\mathcal{B}$ are elementarily equivalent in $\mathcal{L}A$ to $\mathcal{A}_A$ and therefore to each other, which tells us that $T + D\mathcal{A}$ is complete. $(*)$ is therefore an equivalent definition of model completeness and this definition, which is easy to remember, will be preferred in the sequel.

It is clear that if $T \subseteq \mathcal{L}$ is model complete then so too is every theory that extends it in $\mathcal{L}$. Furthermore, T is then inductive. Indeed, a chain K of T-models is always elementary, by $(*)$. By the chain lemma 4.7 we obtain that $\mathcal{A} \preccurlyeq \bigcup K$ for any $\mathcal{A} \in K$ and so $\bigcup K \models T$ thanks to $\mathcal{A} \models T$, which confirms the claim. Hence, by Theorem 4.9, only an $\forall\exists$-theory can be model complete.

An $\forall\exists$-theory that is not model complete is DO. Let $\mathbb{Q}_a := \{x \in \mathbb{Q} \mid a \leqslant x\}$ for $a \in \mathbb{Q}$. Then $(\mathbb{Q}_1, <) \subseteq (\mathbb{Q}_0, <)$ but $(\mathbb{Q}_1, <) \not\preccurlyeq (\mathbb{Q}_0, <)$ as is easily seen. This choice of models also shows that the complete theory DO_{10} is not be model complete. Another example is SO_{10}, since as noticed on page 150, SO_{10} is not an $\forall\exists$-theory and hence is not model complete. Conversely, a model complete theory need not be complete: A prominent example is ACF which will be treated in Theorem 5.4. Nonetheless, with the following theorem the completeness of a theory can often be obtained more easily than with other methods.

Theorem 5.1. *If T is model complete and has a prime model then T is complete.*

Proof. Suppose $\mathcal{A} \models T$ and let $\mathcal{P} \models T$ be a prime model. Then up to isomorphism $\mathcal{P} \subseteq \mathcal{A}$, and so $\mathcal{P} \preccurlyeq \mathcal{A}$ by $(*)$, in particular $\mathcal{P} \equiv \mathcal{A}$. Hence, all T-models are elementarily equivalent to each other so that T is complete. ❏

The following theorem states additional characterizations of model completeness, of which (ii) is as a rule more easily verifiable than the definition. The implication (ii)⇒(i) carries the name *Robinson's test* for model completeness.

Theorem 5.2. *For any theory T the following items are equivalent:*

(i) *T is model complete,*

(ii) *$\mathcal{A} \subseteq \mathcal{B} \Rightarrow \mathcal{A} \subseteq_{ec} \mathcal{B}$, for all $\mathcal{A}, \mathcal{B} \models T$,*

(iii) *each $\exists$-formula α is equivalent in T to an $\forall$-formula β with $\operatorname{free}\beta \subseteq \operatorname{free}\alpha$,*

(iv) *each formula α is equivalent in T to an $\forall$-formula β with $\operatorname{free}\beta \subseteq \operatorname{free}\alpha$.*

Proof. (i)⇒(ii): evident, since $\mathcal{A} \subseteq \mathcal{B} \Rightarrow \mathcal{A} \preccurlyeq \mathcal{B} \Rightarrow \mathcal{A} \subseteq_{ec} \mathcal{B}$. (ii)⇒(iii): According to Theorem 4.6 it is enough to verify that every ∃-formula $\alpha = \alpha(\vec{x}) \in \mathcal{L}$ is persistent in T. Let $\mathcal{A}, \mathcal{B} \vDash T$, $\mathcal{A} \subseteq \mathcal{B}$, $\vec{a} \in A^n$, and $\mathcal{B} \vDash \alpha(\vec{a})$. Then $\mathcal{A} \vDash \alpha(\vec{a})$, because $\mathcal{A} \subseteq_{ec} \mathcal{B}$ thanks to (ii). (iii)⇒(iv): induction on α. (iii) is used only in the ¬-step: Let $\alpha \equiv \beta$, β some ∀-formula (induction hypothesis). Then $\neg\beta \equiv \gamma$ for some ∀-formula γ, hence $\neg\alpha \equiv \gamma$. (iv)⇒(i): let $\mathcal{A}, \mathcal{B} \vDash T$, $\mathcal{A} \subseteq \mathcal{B}$, and $\mathcal{B} \vDash \alpha(\vec{a})$ with $\vec{a} \in A^n$. Then $\mathcal{A} \vDash \alpha(\vec{a})$ since by (iv), $\alpha(\vec{x}) \equiv_T \beta$ for some ∀-formula β. This shows $\mathcal{A} \preccurlyeq \mathcal{B}$, hence (i). ❑

Remark. If T is countable and has infinite models only, then it is possible to restrict the criterion (ii) to models $\mathcal{A}, \mathcal{B}$ of any chosen infinite cardinal number κ. Then we can prove that an ∃-formula is κ-persistent as defined in the remark on page 148, which by the same remark suffices to prove the claim of Theorem 5.2 and hence (iii). Once we have obtained (iii) we have also (i). This is significant for Lindström's criterion, Theorem 5.7.

A relatively simple example of a model complete theory is $T_{V\mathbb{Q}}$, the theory of (nontrivial) $\mathbb{Q}$-vector spaces $\mathcal{V} = (V, +, 0, \mathbb{Q})$, where 0 denotes the zero vector and each $r \in \mathbb{Q}$ is taken to be a unary operation on the set of vectors V. $T_{V\mathbb{Q}}$ formulates the familiar vector axioms, where e.g. the axiom $r(a+b) = ra + rb$ is reproduced as a schema of sentences, namely $\forall a \forall b\, r(a+b) = ra + rb$ for all $r \in \mathbb{Q}$. Let $\mathcal{V}, \mathcal{V}' \vDash T_{V\mathbb{Q}}$ where $\mathcal{V} \subseteq \mathcal{V}'$. We claim that $\mathcal{V} \subseteq_{ec} \mathcal{V}'$. By Theorem 5.2(iii), $T_{V\mathbb{Q}}$ is then model complete. For the claim let $\mathcal{V}' \vDash \exists\vec{x}\alpha$, with a conjunction α of literals in $x_1, \ldots, x_n$ and constants $a_1, \ldots, a_m, b_1, \ldots, b_k \in V$. Then α is essentially a system of the form

$$\text{(s)} \qquad \begin{cases} r_{11}x_1 + \cdots + r_{1n}x_n = a_1 & s_{11}x_1 + \cdots + s_{1n}x_n \neq b_1 \\ \qquad\vdots & \qquad\vdots \\ r_{m1}x_1 + \cdots + r_{mn}x_n = a_m & s_{k1}x_1 + \cdots + s_{kn}x_n \neq b_k \end{cases}$$

Indeed the only prime formulas are term equations, and every term in $x_1, \ldots, x_n$ is equivalent in $T_{V\mathbb{Q}}$ to some term of the form $r_1x_1 + \cdots + r_nx_n$. Without stepping into details it is plausible by the properties of linear systems that the system (s) has already a solution in $\mathcal{V}$, if it is solvable at all; see for instance [Zi].

For the rest of this section we assume some knowledge of classical algebra where *closure constructions* are frequently undertaken. For instance, a torsion-free abelian group has a *divisible closure*, a field $\mathcal{A}$ has an *algebraic closure* (a minimal a.c. extension of $\mathcal{A}$), and an ordered field has a *real closure*; see Example 2 below. Generally speaking, we start from a theory T and $\mathcal{A} \vDash T^\forall$. By a *closure of* $\mathcal{A}$ *in* T we mean a T-model $\bar{\mathcal{A}} \supseteq \mathcal{A}$ such that $\mathcal{A} \subseteq \mathcal{B} \Rightarrow \bar{\mathcal{A}} \subseteq \mathcal{B}$, for every $\mathcal{B} \vDash T$. More precisely, if $\mathcal{A} \subseteq \mathcal{B}$ then there is an embedding of $\bar{\mathcal{A}}$ into $\mathcal{B}$ leaving A pointwise fixed. In this case we say T *permits a closure operation.* Supposing this, let $\mathcal{A}, \mathcal{B} \vDash T$, $\mathcal{A} \subset \mathcal{B}$, and $b \in B \setminus A$. Then there is a smallest submodel of $\mathcal{B}$ containing $A \cup \{b\}$, the $T^\forall$-model generated in $\mathcal{B}$ by $A \cup \{b\}$, denoted by $\mathcal{A}(b)$. Its closure in T is denoted by $\mathcal{A}^b$. It is called an *immediate extension* of $\mathcal{A}$ in T, because of $\mathcal{A} \subset \mathcal{A}^b \subseteq \mathcal{B}$.

Example 1. Let $T := \mathsf{ACF}$. A $T^{\forall}$-model $\mathcal{A}$ is here an integral domain. T permits a closure operation: $\bar{\mathcal{A}}$ is the so-called *algebraic closure* of the quotient field of $\mathcal{A}$. That there exists an a.c. field $\bar{\mathcal{A}}$ embeddable into every a.c. field $\mathcal{B} \supseteq \mathcal{A}$ is the claim of Steinitz's theorem regarding a.c. fields, [Wae, p. 201]. Whenever $\mathcal{A}, \mathcal{B} \models T$ with $\mathcal{A} \subset \mathcal{B}$ and $b \in B \setminus A$, then b is transcendental over $\mathcal{A}$, since $\mathcal{A}$ is already a.c. Thus $a_0 + a_1 b + \cdots + a_n b^n \neq 0$, for all $a_0, \ldots, a_n \in A$ with $a_n \neq 0$. For this reason $\mathcal{A}(b)$ is isomorphic to the ring $\mathcal{A}(x)$ of polynomials $\sum_{i \leqslant n} a_i x^i$ with the "unknown" x (the image of b). Hence, $\mathcal{A}(b) \simeq \mathcal{A}(x) \simeq \mathcal{A}(c)$ provided $\mathcal{A}, \mathcal{B}, \mathcal{C} \models T$, with $\mathcal{A} \subset \mathcal{B}, \mathcal{C}$ and $b \in B \setminus A$, $c \in C \setminus A$. The isomorphism $\mathcal{A}(b) \simeq \mathcal{A}(c)$ extends in a natural way to the quotient fields of $\mathcal{A}(b), \mathcal{A}(c)$ (represented by the field of rational functions over $\mathcal{A}$) and hence to their closures $\mathcal{A}^b$ and $\mathcal{A}^c$. Thus, a T-model has up to isomorphism only one immediate extension in T. Not so in the next more involved example.

Example 2. A *real closed field* is an ordered field $\mathcal{A}$ (like $\mathbb{R}$) in which every polynomial over $\mathcal{A}$ of odd degree has a zero and every $a \geqslant 0$ is a square in A. These properties will turn out to be equivalent to the continuity scheme CS page 86. Let RCF denote the theory of these fields. Although the order is definable in RCF by $x \leqslant y \leftrightarrow \exists z\, y - x = z^2$, order should here be a basic relation. Let $T := \mathsf{RCF}$. A $T^{\forall}$-model $\mathcal{A}$ is an ordered integral domain that determines the order of its quotient field $\mathcal{Q}$. According to Artin's theorem for real closed fields ([Wae, p. 244]), some $\bar{\mathcal{A}} = \bar{\mathcal{Q}} \models \mathsf{RCF}$ can be constructed, called the *real closure of* $\mathcal{A}$ or $\mathcal{Q}$ in T.
Let $\mathcal{A}, \mathcal{B} \models \mathsf{RCF}$, $\mathcal{A} \subset \mathcal{B}$, and $b \in B \setminus A$. Then b is transcendental over $\mathcal{A}$, because no algebraic extension of $\mathcal{A}$ is orderable (this is another characterization of real closed fields). Here $\mathcal{A}(b)$ is isomorphic to the *ordered* ring $\mathcal{A}(x)$ of polynomials over $\mathcal{A}$. $\mathcal{A}(b)$ determines the isomorphism type of its quotient field $\mathcal{Q}(b)$ (containing the quotients of polynomials $p(b)$ over $\mathcal{A}$) and of $\mathcal{A}^b = \overline{\mathcal{Q}(b)}$. Actually, $<^{\mathcal{A}^b}$ is determined by its restriction to $A \cup \{b\}$, or by the partition $A = \{a \in A \mid a <^{\mathcal{A}^b} b\} \cup \{a \in A \mid b <^{\mathcal{A}^b} a\}$. To see this note that it is provable in RCF that a polynomial $p(x)$ with the zeros $a_1, \ldots, a_n \in A$ decomposes in $\mathcal{A} \models \mathsf{RCF}$ as $c \cdot q(x) \cdot \prod_{i=1}^{n} (x - a_i)$ with $c \in A$, $n \geqslant 0$, and $q(x)$ a product of irreducible polynomials of degree 2 or perhaps =1. In $\mathcal{Q}(b)$ (and $\mathcal{A}^b$) holds $q(b) > 0$. Indeed, each irreducible factor $b^2 + db + e$ of $q(b)$ is > 0 since $b^2 + db + e = (b + \frac{d}{2})^2 + e - \frac{d^2}{4} > 0$ $(d, e \in A)$. Thus we know whether or not $p(b) > 0$ if we know the signs of $b - a_i$ for all zeros a_i of $p(x)$ in A. This suffices to fix the order in $\mathcal{Q}(b)$ as is easily seen, and hence in $\mathcal{A}^b$ by Artin's theorem.

For inductive theories T that permit a closure operation, Robinson's test for model completeness can still be simplified as follows:

Lemma 5.3. *Let T be inductive, and suppose T permits a closure operation. Assume further that $\mathcal{A} \subseteq_{ec} \mathcal{A}'$ for all $\mathcal{A}, \mathcal{A}' \models T$ for the case that $\mathcal{A}'$ is an immediate extension of $\mathcal{A}$ in T. Then T is model complete.*

Proof. Let $\mathcal{A}, \mathcal{B} \models T$, $\mathcal{A} \subseteq \mathcal{B}$. By Theorem 5.2(ii) it suffices to show that $\mathcal{A} \subseteq_{ec} \mathcal{B}$. Let H be the set of all $\mathcal{C} \subseteq \mathcal{B}$ such that $\mathcal{A} \subseteq_{ec} \mathcal{C} \models T$. Trivially $\mathcal{A} \in H$. Since T is inductive, a chain $K \subseteq H$ satisfies $\bigcup K \models T$. One easily verifies $\mathcal{A} \subseteq_{ec} \bigcup K$ as well, so that $\bigcup K \in H$. By Zorn's lemma there is a maximal element $\mathcal{A}_m \in H$. **Claim:** $\mathcal{A}_m = \mathcal{B}$. Assume $\mathcal{A}_m \subset \mathcal{B}$. Then there is an immediate extension $\mathcal{A}'_m \models T$ of $\mathcal{A}_m$ such that $\mathcal{A}_m \subset \mathcal{A}'_m \subseteq \mathcal{B}$. Since $\mathcal{A} \subseteq_{ec} \mathcal{A}_m$, and by hypothesis $\mathcal{A}_m \subseteq_{ec} \mathcal{A}'_m$, we get $\mathcal{A} \subseteq_{ec} \mathcal{A}'_m$. This, however, contradicts the maximality of $\mathcal{A}_m$ in H. Therefore, it must be the case that $\mathcal{A}_m = \mathcal{B}$. Consequently, $\mathcal{A} \subseteq_{ec} \mathcal{B}$. ❏

Theorem 5.4. ACF *is model complete and thus so too* ACF_p*, the theory of a.c. fields of given characteristic* p *(*$= 0$ *or a prime). Moreover* ACF_p *is complete.*

Proof. Let $\mathcal{A}, \mathcal{B} \models \mathsf{ACF}$, $\mathcal{A} \subset \mathcal{B}$, and $b \in B \setminus A$. By Lemma 5.3 it suffices to show that $\mathcal{A} \subseteq_{ec} \mathcal{A}^b$. Here $\mathcal{A}^b$ is an immediate extension of $\mathcal{A}$ in ACF. Let $\alpha := \exists \vec{x} \beta(\vec{x}, \vec{a}) \in \mathcal{L}A$, β quantifier-free, and $\mathcal{A}^b \models \alpha$. We shall prove $\mathcal{A} \models \alpha$ and for this we consider

$$X := \mathsf{ACF} \cup D\mathcal{A} \cup \{p(x) \neq 0 \mid p(x) \text{ a monic polynomial on } A\}.$$

With b for x one sees that $(\mathcal{A}^b, b) \models X$ (b is trancendental over $\mathcal{A}$). Let $(\mathcal{C}, c) \models X$, with c for x. Since $\mathcal{C} \models D\mathcal{A}$, w.l.o.g. $\mathcal{A} \subseteq \mathcal{C}$. By Example 1 $\mathcal{A}^b \simeq \mathcal{A}^c$, and so $\mathcal{A}^c \models \alpha$. $\mathcal{A}^c \subseteq \mathcal{C}$ implies $\mathcal{C} \models \alpha$, for α is an $\exists$-sentence. Since $(\mathcal{C}, c)$ has been chosen arbitrarily we obtain $X \vdash \alpha$, and from this by the finiteness theorem evidently

$$D\mathcal{A}, \textstyle\bigwedge_{i \leqslant k} p_i(x) \neq 0 \vdash_{\mathsf{ACF}} \alpha, \quad \text{for some } k \text{ and monic polynomials } p_0, \ldots, p_k.$$

Particularization and the deduction theorem show $D\mathcal{A} \vdash_{\mathsf{ACF}} \exists x \bigwedge_{i \leqslant k} p_i(x) \neq 0 \to \alpha$. Every a.c. field is infinite (Example 5(c) in **5.2**), and a polynomial has only finitely many zeros in a field. Thus, $D\mathcal{A} \vdash_{\mathsf{ACF}} \exists x \bigwedge_{i \leqslant k} p_i(x) \neq 0$. Hence, $D\mathcal{A} \vdash_{\mathsf{ACF}} \alpha$ and so $\mathcal{A} \models \alpha$. This proves $\mathcal{A} \subseteq_{ec} \mathcal{A}^b$ and in view of Lemma 5.3 the first part of the theorem. The algebraic closure of the prime field of characteristic p is obviously a prime model for ACF_p. Therefore, by Theorem 5.1, ACF_p is complete. ❏

The following significant theorem is won similarly. It was originally proved by Tarski in [Ta2] by means of quantifier elimination. Incidentally, the completeness claim is not obtainable using Vaught's criterion, in contrast to the case of ACF.

Theorem 5.5. *The theory* RCF *of real closed fields is model complete and complete. It is thus identical to the theory of the ordered field of real numbers, and as a complete axiomatizable theory it is also decidable.*

Proof. Let $\mathcal{A} \models \mathsf{RCF}$. It once again suffices to show that $\mathcal{A} \subseteq_{ec} \mathcal{A}^b$ for an immediate extension $\mathcal{A}^b$ of $\mathcal{A}$ in RCF. Let $U := \{a \in A \mid a <^{\mathcal{B}} b\}$, $V := \{a \in A \mid b <^{\mathcal{B}} a\}$, with $\mathcal{B} := \mathcal{A}^b$. Then $U \cup V = A$. Now let $\mathcal{A}^b \models \exists \vec{x} \beta(\vec{x}, \vec{a})$, β quantifier-free, $\vec{a} \in A^m$. The model $(\mathcal{B}, b)$ with b for x then clearly satisfies the set of formulas

$$X := \mathsf{RCF} \cup D\mathcal{A} \cup \{a < x \mid a \in U\} \cup \{x < a \mid a \in V\}.$$

Suppose $(\mathcal{C}, c) \models X$, interpreting x as c. We may assume $\mathcal{A} \subseteq \mathcal{C}$ because $\mathcal{C} \models D\mathcal{A}$. Since $c \notin U \cup V = A$, c is transcendental over $\mathcal{A}$ (see Example 2). Hence, the quotient field $\mathcal{Q}(c)$ of $\mathcal{A}(c)$ is isomorphic to the field of rational functions over $\mathcal{A}$ with the unknown x. The order of $\mathcal{Q}(c)$ is fixed by the partition $A = U \cup V$ coming from $\mathcal{Q}(b)$. Thus, $\mathcal{Q}(b) \simeq \mathcal{Q}(x) \simeq \mathcal{Q}(c)$. The isomorphism $\mathcal{Q}(b) \simeq \mathcal{Q}(c)$ extends to one between the real closures $\mathcal{A}^b$ and $\mathcal{A}^c$. As in Theorem 5.4 we thus obtain $X \vdash \alpha$, and so for some $a_1, \dots, a_k, b_1, \dots, b_l \in A$, where $k, l \geqslant 0$ but $k + l > 0$,

$$D\mathcal{A} \vdash_{\mathsf{RCF}} \exists x(\textstyle\bigwedge_{i=1}^{k} a_i < x \wedge \bigwedge_{i=1}^{l} x < b_i) \to \alpha \qquad (a_i \in U,\ b_i \in V).$$

Now, an ordered field is densely ordered without edge elements, and is infinite. Hence, $\vdash_{\mathsf{RCF}} \exists x(\bigwedge_{i=1}^{k} a_i < x \wedge \bigwedge_{i=1}^{l} x < b_i)$. This results in $D\mathcal{A} \vdash_{\mathsf{RCF}} \alpha$. Therefore $\mathcal{A} \models \alpha$, and $\mathcal{A} \subseteq_{ec} \mathcal{A}^b$ is proved. To verify completeness observe that RCF has a prime model, namely the real closure of $\mathbb{Q}$, the ordered field of the real algebraic numbers. Applying Theorem 5.1 once again confirms the completeness of RCF. ❑

A theory T is called the *model completion* of a theory T_0 of the same language if $T_0 \subseteq T$ and $T + D\mathcal{A}$ is complete for every $\mathcal{A} \models T_0$. Clearly, T is then model complete; moreover, T is model compatible with T_0 ($\mathcal{A} \models T_0$ implies $(\exists \mathcal{C} \in \operatorname{Md} T)\mathcal{A} \subseteq \mathcal{C}$, since $T + D\mathcal{A}$ is consistent). The existence of a model complete extension is necessary for the existence of a model completion of T_0, but not sufficient; see Exercise 1.

A somewhat surprising fact is that a model completion of T is uniquely determined provided it exists. Indeed, let T, T' be model completions of T_0. Both theories are model compatible with T_0, and hence with each other. T, T' are model complete and therefore inductive, so that $T + T'$ is model compatible with T (Exercise 3 in **5.4**). Thus, if $\mathcal{A} \models T$ then there exist some $\mathcal{B} \models T + T'$ with $\mathcal{A} \subseteq \mathcal{B}$, and since T is model complete we obtain $\mathcal{A} \preccurlyeq \mathcal{B}$. This implies $\mathcal{A} \equiv \mathcal{B} \models T'$, and consequently $\mathcal{A} \models T'$. For reasons of symmetry, $\mathcal{A} \models T' \Rightarrow \mathcal{A} \models T$ as well. Therefore $T = T'$.

Example 3. ACF is the model completion of the theory T_J of all integral domains and so a fortiori of the theory T_F of all fields. Indeed, let $\mathcal{A} \models T_J$. By Theorem 5.4, ACF is model complete, hence also $T := \mathsf{ACF} + D\mathcal{A}$ (in $\mathcal{L}A$). Moreover, T is complete, because by Example 1, T has a prime model, namely the closure $\bar{\mathcal{A}}$ of $\mathcal{A}$ in ACF. Using Theorem 5.5, one analogously shows that RCF is the model completion of the theories of ordered commutative rings with unit element, and of ordered fields.

$\mathcal{A} \models T$ is called *existentially closed in* T, or $\exists$*-closed in* T for short, if $\mathcal{A} \subseteq_{ec} \mathcal{B}$ for each $\mathcal{B} \models T$ with $\mathcal{A} \subseteq \mathcal{B}$. For instance, every a.c. field $\mathcal{A}$ is $\exists$-closed in the theory of fields. For let $\mathcal{B} \supseteq \mathcal{A}$ be any field and $\mathcal{C}$ be any a.c. extension of $\mathcal{B}$. Then $\mathcal{A} \preccurlyeq \mathcal{C}$ thanks to the model completeness of ACF. Hence $\mathcal{A} \subseteq_{ec} \mathcal{B}$ by Lemma 4.8(ii). The following lemma generalizes in some sense the fact that every field is embeddable into an a.c. field. Similarly, a group, for instance, is embeddable into a group that is $\exists$-closed in the theory of groups.

Lemma 5.6. *Let T be an $\forall\exists$-theory of some countable language $\mathcal{L}$. Then every infinite model $\mathcal{A}$ of T can be extended to a model $\mathcal{A}^*$ of T such that $|\mathcal{A}^*| = |\mathcal{A}|$, which is $\exists$-closed in T.*

Proof. For the proof we assume, for simplicity, that $\mathcal{A}$ is countable. Then $\mathcal{L}A$ is also countable. Let $\alpha_0, \alpha_1, \ldots$ be an enumeration of the $\exists$-sentences of $\mathcal{L}A$ and $\mathcal{A}_0 = \mathcal{A}_A$.[7] Let $\mathcal{A}_{n+1}$ be an extension of $\mathcal{A}_n$ in $\mathcal{L}A$ such that $\mathcal{A}_{n+1} \models T + \alpha_n$, as long as such an extension exists; otherwise simply put $\mathcal{A}_{n+1} = \mathcal{A}_n$. Since T is inductive, $\mathcal{B}_0 = \bigcup_{n\in\mathbb{N}} \mathcal{A}_n \models T$. If $\alpha = \alpha_n$ is an $\exists$-sentence in $\mathcal{L}A$ valid in some extension $\mathcal{B} \models T$ of $\mathcal{B}_0$, then already $\mathcal{A}_{n+1} \models \alpha$ and thus also $\mathcal{B}_0 \models \alpha$. Now we repeat this construction with $\mathcal{B}_0$ in place of $\mathcal{A}_0$ with respect to an enumeration of all $\exists$-sentences in $\mathcal{L}B_0$ and obtain an $\mathcal{L}B_0$-structure $\mathcal{B}_1 \models T$. Subsequent reiterations produce a sequence $\mathcal{B}_1 \subseteq \mathcal{B}_2 \subseteq \cdots$ of $\mathcal{L}B_n$-structures $\mathcal{B}_{n+1} \models T$. Let $\mathcal{A}^*$ ($\models T$) be the $\mathcal{L}$-reduct of $\bigcup_{n\in\mathbb{N}} \mathcal{B}_n \models T$ and $\mathcal{A}^* \subseteq \mathcal{B} \models T$. Assume $\mathcal{B} \models \exists\vec{x}\beta(\vec{a}, \vec{x})$, $\vec{a} \in (A^*)^n$. Then $\mathcal{B}_m \models \beta(\vec{a}, \vec{b})$ for suitable m. Hence $\bigcup_{n\in\mathbb{N}} \mathcal{B}_n \models \beta(\vec{a}, \vec{b})$ and so $\mathcal{A}^* \models \exists\vec{x}\beta(\vec{a}, \vec{x})$. ❏

With this lemma one readily obtains the following highly applicable criterion for proving the model completeness of certain theories, which, by Vaught's criterion, are always complete at the same time.

Theorem 5.7 (Lindström's criterion). *A countable κ-categorical $\forall\exists$-theory T without finite models is not only complete but also model complete.*

Proof. Since all T-models are infinite, T has a model of cardinality κ, and by Lemma 5.6 also one that is $\exists$-closed in T. But then all T-models of cardinality κ are $\exists$-closed in T, because all these are isomorphic. Thus $\mathcal{A} \subseteq \mathcal{B} \Rightarrow \mathcal{A} \subseteq_{ec} \mathcal{B}$, for all $\mathcal{A}, \mathcal{B} \models T$ of cardinality κ. Therefore, T is model complete according to the remark on page 152. ❏

Examples of applications.

(a) The $\aleph_0$-categorical theory of atomless Boolean algebras.

(b) The $\aleph_1$-categorical theory of nontrivial $\mathbb{Q}$-vector spaces.

(c) The $\aleph_1$-categorical theory of a.c. fields of given characteristic.

A few comments: A Boolean algebra $\mathcal{B}$ is called *atomless* if for each $a \neq 0$ in B there is some $b \neq 0$ in $\mathcal{B}$ with $b < a$ ($<$ is the partial lattice order of $\mathcal{B}$). The proof of (a) is similar to that for densely ordered sets. Also (b) is easily verified. Observe that a $\mathbb{Q}$-vector space of cardinality $\aleph_1$ has a base of cardinality $\aleph_1$. From (c) the model completeness of ACF follows in a new way: If $\mathcal{A}, \mathcal{B} \models \mathsf{ACF}$ and $\mathcal{A} \subseteq \mathcal{B}$ then both fields have the same characteristic p. Since ACF_p is model complete by (c), $\mathcal{A} \preccurlyeq \mathcal{B}$ follows. This obviously implies that ACF is model complete as well.

[7] For uncountable $\mathcal{A}$ we have $|\mathcal{L}A| = |\mathcal{A}|$. In this case one proceeds with an ordinal enumeration of $\mathcal{L}A$ rather than an ordinary one. But the proof is almost the same.

Exercises

1. Prove that of the four theories DO_{ij} only DO_{00} is model complete. Moreover, show that DO has no model completion.

2. Let T be the theory of divisible torsion-free abelian groups. Show that
   (a) T is model complete,
   (b) T is the model completion of the theory T_0 of torsion-free abelian groups.

3. T^* is called the *model companion* of T provided T, T^* are model compatible and T^* is model complete. Show that if T^* exists then T^* is uniquely determined, and $\mathrm{Md}\, T^*$ consists of all models $\exists$-closed in T.

4. Prove that an $\forall\exists$-sentence valid in all finite fields is valid in all a.c. fields. This fact is highly useful in algebraic geometry.

5.6 Quantifier Elimination

Because $\exists x(y < x \wedge x < z) \equiv_{\mathsf{DO}} y < z$, in the theory of densely ordered sets the quantifier in the left-hand formula can be eliminated. In fact, in some theories, including the theory DO_{00} (see **5.2**), the quantifiers can be eliminated from every formula. One says that T ($\subseteq \mathcal{L}^0$) *allows quantifier elimination* if for every $\varphi \in \mathcal{L}$ there exists some open formula $\varphi' \in \mathcal{L}$ such that $\varphi \equiv_T \varphi'$. Quantifier elimination is the oldest method of showing certain theories to be decidable and occasionally also to be complete. Some presentations demand additionally $\mathit{free}\,\varphi' = \mathit{free}\,\varphi$, but this is irrelevant.

A theory T allowing quantifier elimination is model complete by Theorem 5.2(iv), because open formulas are in particular $\forall$-formulas. T is therefore an $\forall\exists$-theory, a remarkable necessary condition for quantifier eliminability.

In order to confirm quantifier elimination for a theory T it suffices to eliminate the prefix $\exists x$ from every formula of the form $\exists x\alpha$, where α is open. Indeed, think of all subformulas of the form $\forall x\alpha$ in a formula φ as being equivalently replaced by $\neg\exists x\neg\alpha$, so that only the $\exists$-quantifier appears in φ. Looking at the farthest-right prefix $\exists x$ in φ one can write $\varphi = \cdots\exists x\alpha\cdots$ with quantifier-free α. Now, if $\exists x\alpha$ is replaceable by an open formula α' then this process can be iterated no matter how long it takes for all $\exists$-quantifiers in φ to disappear.

Thanks to the $\vee$-distributivity of the $\exists$-quantifiers we may moreover assume that the quantifier-free part α of $\exists x\alpha$ from which $\exists x$ has to be eliminated is a conjunction of literals, and that x explicitly occurs in each of these literals: simply convert α

into a disjunctive normal form and distribute $\exists x$ over the disjuncts such that $\exists x$ stands in front of a conjunction of literals only. If x does not appear in any of these literals, $\exists x$ can simply be discarded. Otherwise remove the literals not containing x beyond the scope of $\exists x$, observing that $\exists x(\alpha \wedge \beta) \equiv \exists x\alpha \wedge \beta$ if $x \notin var\,\beta$.

Furthermore it can be supposed that none of the conjuncts is of the form $x = t$ with $x \notin var\,t$. Indeed, since $\exists x(x = t \wedge \alpha) \equiv \alpha\,\frac{t}{x}$, the quantifier has then already been eliminated. We may also assume that x is not $\boldsymbol{v}_0$ (using bound renaming) and that neither $x = x$ nor $x \neq x$ is among the conjuncts. For $x = x$ can equivalently be replaced by $\top$, as can $x \neq x$ by $\bot$. Here one may define $\top$ and $\bot$ as $\boldsymbol{v}_0 = \boldsymbol{v}_0$ and $\boldsymbol{v}_0 \neq \boldsymbol{v}_0$, respectively. Replacement will then introduce $\boldsymbol{v}_0$ as a possible new free variable, but that is harmless. If the language contains a constant c one may replace $\boldsymbol{v}_0$ by c in the above consideration. If not, one may add a constant or even $\bot$ as a new prime formula to the language, similar to what is proposed below for DO.

Call an $\exists$-formula *simple* if it is of the form $\exists x \bigwedge_i \alpha_i$, where every α_i is a literal with $x \in var\,\alpha_i$. Then the above considerations result in the following

Theorem 6.1. *T allows quantifier elimination if every simple $\exists$-formula $\exists x \bigwedge_i \alpha_i$ is equivalent in T to some open formula. Here without loss of generality, none of the literals α_i is $x = x$, $x \neq x$, or of the form $x = t$ with $x \notin var\,t$.*

Example 1. DO_{00} allows quantifier elimination. Because $y \not< z \equiv_T z < y \vee z = y$ and $z \neq y \equiv_T z < y \vee y < z$ and since in general $(\alpha \vee \beta) \wedge \gamma \equiv (\alpha \wedge \gamma) \vee (\beta \wedge \gamma)$, we may suppose that the conjunction of the α_i in Theorem 6.1 does not contain the negation symbol. We are therefore dealing with a formula of the form

$$\exists x(y_1 < x \wedge \cdots \wedge y_m < x \ \wedge\ x < z_1 \wedge \cdots \wedge x < z_k),$$

which is equivalent to $\bot$ if x is one of the variables y_i, z_j. If not, it is equivalent to $\top$ whenever $m = 0$ or $k = 0$, and in the remaining case to $\bigwedge_{i,j=1}^{n} y_i < z_j$. That's it.

DO itself does not allow quantifier elimination. For instance, in $\alpha(y) := \exists x\, x < y$ the quantifier is not eliminable. If $\alpha(y)$ were equivalent in DO to an open formula then $\mathcal{A}, \mathcal{B} \vDash \mathsf{DO}$, $\mathcal{A} \subseteq \mathcal{B}$, $a \in A$, and $\mathcal{B} \vDash \alpha(a)$ would imply $\mathcal{A} \vDash \alpha(a)$. But this is not so for the densely ordered sets $\mathcal{A}, \mathcal{B}$ with $A = \{x \in \mathbb{Q} \mid 1 \leqslant x\}$ and $B = \mathbb{Q}$. Choose $a = 1$. Quantifier elimination does however become possible if the signature $\{<\}$ is expanded by considering the formulas L, R as 0-ary predicate symbols. The fact that $\{\mathsf{L}, \mathsf{R}\}$ forms a Boolean basis for sentences in DO is not yet sufficient for quantifier eliminability. What is needed here is a Boolean basis for the set of all formulas (not only sentences) modulo DO.

Also the theory SO does not allow quantifier elimination in the original language, simply because it is not an $\forall\exists$-theory as was noticed earlier. The same holds for the expansions SO_{ij}.

Example 2. A classic, by no means trivial, result of quantifier elimination by Presburger refers to $Th(\mathbb{N}, 0, 1, +, <)$, with the additional unary predicate symbols $m|$ $(m = 2, 3, \dots)$, explicitly defined by $m|x \leftrightarrow \exists y\, my = x$ where my denotes the m-fold sum $y + \cdots + y$ of y. We shall prove a related result with respect to the group $\mathbb{Z}$ in $\mathcal{L}\{0, 1, +, -, <, 2|, 3|, \dots\}$. Denote the k-fold sum $1 + \cdots + 1$ by k in what follows, and set $(-k)x := -kx$.

Let ZGE be the elementary theory in $\mathcal{L}\{0, 1, +, -, <, 2|, 3|, \dots\}$ whose axioms subsume those for ordered abelian groups, and the axioms

$$\forall x(0 < x \leftrightarrow 1 \leqslant x),\ \forall x(m|x \leftrightarrow \exists y\, my \equiv x) \text{ and } \vartheta_m := \forall x \bigvee_{k<m} m|x + k$$

for $m = 2, 3, \dots$ ZGE-models, more precisely, their reducts to $\mathcal{L} := \mathcal{L}\{0, 1, +, -, <\}$, are called $\mathbb{Z}$-*groups*. These are ordered with smallest positive element 1. The ϑ_m state for a $\mathbb{Z}$-group G that the factor groups G/mG are cyclic of order m. Here $mG := \{mx \mid x \in G\}$. Let ZG denote the reduct theory of ZGE in $\mathcal{L}$ whose models are just the $\mathbb{Z}$-groups. ZGE is a definitorial and hence a conservative extension of ZG (cf. **2.6**). It will turn out that $\mathbb{Z}$-groups are precisely the ordered abelian groups elementarily equivalent to the paradigm structure $(\mathbb{Z}, 0, 1, +, -, <)$. Let us notice that $\vdash_{\mathsf{ZG}} \eta_n$ for each n, where η_n is the formula $0 \leqslant x < n \to \bigvee_{k \leqslant n} x = k$.

We are now going to prove that ZGE allows quantifier elimination. Observe first that since $t \neq s \equiv_{\mathsf{ZGE}} s < t \vee t < s$ and $m \nmid t \equiv_{\mathsf{ZGE}} \bigvee_{i=1}^{m-1} m|t + i$ and $m|t \equiv_{\mathsf{ZGE}} m|-t$ it may be assumed that the kernel of a simple $\exists$-formula is a conjunction of formulas of the form $n_i x = t_i^0$, $n_i' x < t_i^1$, $t_i^2 < n_i'' x$, and $m_i|n_i''' x + t_i^3$ where $x \notin var\, t_i^j$. By multiplying these formulas by a suitable number and using $t < s \equiv_{\mathsf{ZGE}} nt < ns$ and $m|t \equiv_{\mathsf{ZGE}} nm|nt$ for $n \neq 0$, one sees that all the n_i, n_i', n_i'', n_i''' can be made equal to some number $n > 1$. Clearly, in doing so, t_i^j and the "modules" m_i all change. But the problem of elimination is thus reduced to formulas of the following form, where the jth conjunct disappears whenever $k_j = 0$ $(j \leqslant 3)$:

(1) $\exists x\big(\bigwedge_{i=1}^{k_0} nx = t_i^0 \wedge \bigwedge_{i=1}^{k_1} t_i^1 < nx \wedge \bigwedge_{i=1}^{k_2} nx < t_i^2 \wedge \bigwedge_{i=1}^{k_3} m_i|nx + t_i^3\big)$.

With y for nx and $m_0 = n$, (1) is certainly equivalent in ZGE to

(2) $\exists y\big(\bigwedge_{i=1}^{k_0} y = t_i^0 \wedge \bigwedge_{i=1}^{k_1} t_i^1 < y \wedge \bigwedge_{i=1}^{k_2} y < t_i^2 \wedge \bigwedge_{i=1}^{k_3} m_i|y + t_i^3 \ \wedge\ m_0|y\big)$.

According to Theorem 6.1 we can at once assume that $k_0 = 0$, so that the elimination problem, after renaming y back to x, reduces to formulas of the form

(3) $\exists x\big(\bigwedge_{i=1}^{k_1} t_i^1 < x \wedge \bigwedge_{i=1}^{k_2} x < t_i^2 \wedge \bigwedge_{i=0}^{k_3} m_i|x + t_i^3\big)$

where still $x \notin var\, t_i^j$. Let m be the smallest common multiple of $m_0, \dots, m_{k_3}$.

Case 1: $k_1 = k_2 = 0$. Then (3) is equivalent in ZGE to $\bigvee_{j=1}^{m} \bigwedge_{i=0}^{k_3} m_i|j + t_i^3$. Indeed if an x such that $\bigwedge_{i=0}^{k_3} m_i|x + t_i^3$ exists at all, then so does some $x = j \in \{1, \dots, m\}$. For let j be determined by axiom ϑ_m so that $m|x + (m - j)$, i.e., also $m|x - j$ and consequently $m_i|x - j$ for all $i \leqslant k_3$. Then $m_i|x + t_i^3 - (x - j) = j + t_i^3$ also holds for $i = 0, \dots, k_3$ as was claimed.

Case 2: $k_1 \neq 0$ and j as above. Then (3) is equivalent to

(4) $\bigvee_{\mu=1}^{k_1}[\bigwedge_{i=1}^{k_1} t_i^1 \leqslant t_\mu^1 \wedge \bigvee_{j=1}^{m}(\bigwedge_{i=1}^{k_2} t_\mu^1 + j < t_i^2 \wedge \bigwedge_{i=0}^{k_3} m_i | t_\mu^1 + j + t_i^3)]$.

This is a case distinction according to the maximum among the values of the t_i^1. From each disjunct in (4) certainly (3) follows in ZGE (consider $t_i^1 < t_\mu^1 + j$). Now suppose conversely that x is a solution of (3). Then in the case $\bigwedge_{i=1}^{k_1} t_i^1 \leqslant t_\mu^1$ the μth disjunct of (4) is also valid. For this we need only confirm $t_\mu^1 + j < t_i^2$, which comes down to $t_\mu^1 + j \leqslant x$. Were $x < t_\mu^1 + j$, i.e., $0 < x - t_\mu^1 < j$, then $x - t_\mu^1 = k$ follows for some $k < j$ by η_j, that is, $x = t_\mu^1 + k$. Thus, $m_i | t_\mu^1 + j - x = j - k$ for all $i \leqslant k_3$. But this yields the contradiction $m | j - k < m$.

Case 3: $k_1 = 0$ and $k_2 \neq 0$. The argument is analogous to Case 2 but with a distinction according to the smallest term among the $t_i^{k_2}$.

From this remarkable example we obtain the following

Corollary 6.2. ZGE *is model complete.* ZGE *and* ZG *are complete and decidable.*

Proof. Since $\mathbb{Z}$ is obviously a prime model for ZG, completeness follows from model completeness, which in turn follows from quantifier eliminability. Clearly, along with ZGE also its reduct theory ZG is complete. Hence, as complete axiomatizable theories, both these theories are decidable. ❑

Remark 1. Also ZG is model complete; Exercise 1. It is in fact the model completion of the theory of discretely ordered abelian groups because every such group is embeddable into some $\mathbb{Z}$-group (not quite easy to prove). This is a main reason for the interest in ZG. Although model complete, ZG does not allow quantifier elimination.

We now intend to show that theories ACF and RCF of algebraically and real closed fields respectively allow quantifier elimination, even without any expansion of their signatures. We undertake the proof with a model-theoretical criterion for quantifier elimination, Theorem 6.4. In its proof we will use a variant of Theorem 2.3. Call $X \subseteq \mathcal{L}$ a *Boolean basis for* $\mathcal{L}$ in T if *every* $\varphi \in \mathcal{L}$ belongs to $\langle X \rangle$ (page 140). Let $\mathcal{M}, \mathcal{M}'$ be $\mathcal{L}$-models and write $\mathcal{M} \equiv_X \mathcal{M}'$ instead of $(\forall \varphi \in X)(\mathcal{M} \vDash \varphi \Leftrightarrow \mathcal{M}' \vDash \varphi)$, and $\mathcal{M} \equiv \mathcal{M}'$ instead of $(\forall \varphi \in \mathcal{L})(\mathcal{M} \vDash \varphi \Leftrightarrow \mathcal{M}' \vDash \varphi)$.

Theorem 6.3 (Basis theorem for formulas). *Let T be a theory, $X \subseteq \mathcal{L}$, and suppose that $\mathcal{M} \equiv_X \mathcal{M}' \Rightarrow \mathcal{M} \equiv \mathcal{M}'$, for all $\mathcal{M}, \mathcal{M}' \vDash T$. Then X is a Boolean basis for $\mathcal{L}$ in T.*

Proof. Let $\alpha \in \mathcal{L}$ and $Y_\alpha := \{\gamma \in \langle X \rangle \mid \alpha \vdash_T \gamma\}$. One then shows that $Y_\alpha \vdash_T \alpha$ as in the proof of Theorem 2.3 by arguing with a model $\mathcal{M}$ rather than a structure $\mathcal{A}$. The remainder of the proof proceeds along the lines of Theorem 2.3. ❑

A theory T is called *substructure complete* if for all $\mathcal{A}, \mathcal{B}$ where $\mathcal{A} \subseteq \mathcal{B} \vDash T$ the theory $T + \mathcal{D}\mathcal{A}$ is complete. This is basically only a reformulation of T's being the model completion of $T^\forall$. Indeed, let T be substructure complete and $\mathcal{A} \vDash T^\forall$. Then

by Lemma 4.1, $\mathcal{A} \subseteq \mathcal{B}$ for some $\mathcal{B} \models T$, and $T + \mathcal{D}\mathcal{A}$ is hence complete. Conversely, let T be the model completion of $T^\forall$ and $\mathcal{A} \subseteq \mathcal{B} \models T$. Then $\mathcal{A} \models T^\forall$, hence $T + \mathcal{D}\mathcal{A}$ is complete so that T is substructure complete. In view of this fact we need to pick up only one of these properties in the next theorem. There exist yet other criteria, in particular the amalgamability of models of $T^\forall$; see for instance [CK].

Theorem 6.4. *For every theory T in $\mathcal{L}$ the following properties are equivalent:*

(i) *T allows quantifier elimination,* (ii) *T is substructure complete.*

Proof. (i)⇒(ii): Let $\mathcal{A}$ be a substructure of a T-model, $\alpha(\vec{x}) \in \mathcal{L}$, and $\vec{a} \in A^n$ such that $\mathcal{A} \models \alpha\,[\vec{a}]$. Further let $\mathcal{B} \models T, D\mathcal{A}$ so that w.l.o.g. $\mathcal{B} \supseteq \mathcal{A}$. Then also $\mathcal{B} \models \alpha(\vec{a})$, because in view of (i) we may suppose that α contains no quantifiers. Since $\mathcal{B}$ was arbitrary, $D\mathcal{A} \vdash_T \alpha(\vec{a})$. Hence $T + D\mathcal{A}$ is complete.

(ii)⇒(i): Suppose $\mathcal{M} := (\mathcal{A}, w) \models T, \varphi(\vec{x})$ and let X be the set all of literals λ of $\mathcal{L}$. **Claim:** $T_X\mathcal{M} \vdash_T \varphi(\vec{x})$, where $T_X\mathcal{M} := \{\varphi \in X \mid \mathcal{M} \models \varphi\}$ is the set of formulas from X true in $\mathcal{M}$. Let $\mathcal{A}^E$ be the substructure generated from $E := \{a_1, \ldots, a_n\}$ in $\mathcal{A}$, where $a_1 = x_1^w, \ldots, a_n = x_n^w$. By (ii), $T + D\mathcal{A}^E$ is complete and moreover consistent with $\varphi(\vec{a})$ (observe $\mathcal{A}_A \models T + D\mathcal{A}^E + \varphi(\vec{a})$). Hence $D\mathcal{A}^E \vdash_T \varphi(\vec{a})$. Thus, by the finiteness theorem, there are literals $\lambda_0(\vec{x}), \ldots, \lambda_k(\vec{x})$ with $\lambda_i(\vec{a}) \in D\mathcal{A}^E$ and $\bigwedge_{i\leqslant k} \lambda_i(\vec{a}) \vdash_T \varphi(\vec{a})$. Therefore $\bigwedge_{i\leqslant k} \lambda_i(\vec{x}) \vdash_T \varphi(\vec{x})$, because $a_1, \ldots, a_n$ do not appear in T. Certainly $\lambda_i(\vec{x}) \in T_X\mathcal{M}$ for all $i \leqslant k$, hence $T_X\mathcal{M} \vdash_T \varphi(\vec{x})$. This proves the claim. It holds for arbitrary $\varphi(\vec{x}) \in \mathcal{L}$ provided $\mathcal{M} \models \varphi(\vec{x})$, so that $T_X\mathcal{M}$ is clearly maximally consistent. This in turn implies that $\mathcal{M} \equiv_X \mathcal{M}' \Rightarrow \mathcal{M} \equiv \mathcal{M}'$, for all $\mathcal{M}, \mathcal{M}' \models T$ as is easily seen. Thus, according to Theorem 6.3, the literals of $\mathcal{L}$ form a Boolean basis for $\mathcal{L}$ in T, which obviously amounts to saying that T allows quantifier elimination, and (i) is proved. ❑

Corollary 6.5. *An $\forall$-theory T permits quantifier elimination if and only if T is model complete.*

Proof. Due to $\mathcal{A} \subseteq \mathcal{B} \models T \Rightarrow \mathcal{A} \models T$, (ii) in Theorem 6.4 is satisfied provided only $T + D\mathcal{A}$ is complete for all $\mathcal{A} \models T$. But this is granted if T is model complete. ❑

Example 3. Let T be the $\forall$-theory with two unary function symbols f, g whose axioms state that f and g are injective, f and g are mutually inverse ($\forall x\, fgx = x$ and $\forall x\, gfx = x$), and there are no circles (cf. Example 3 in **5.2**). Note that $\forall y \exists x fx = y$ is provable from the axiom $\forall x\, f(gx) = x$. Hence, f and g are bijective. The T-models consist of disjoint countable infinite "threads" which occurred also in the just mentioned example. Hence, T is $\aleph_1$-categorical and thus model complete by Lindström's criterion. By the corollary, T permits the elimination of quantifiers.

Theorem 6.6. ACF *and* RCF *allow quantifier elimination.*

Proof. By Theorem 6.4 it is enough to show that ACF and RCF are substructure complete, or put another way, ACF and RCF are the model completions of $\mathsf{ACF}^\forall$ and $\mathsf{RCF}^\forall$, respectively. Both claims are clear from Example 3 in **5.5**, since $\mathsf{ACF}^\forall$ is identical to the theory of integral domains, and $\mathsf{RCF}^\forall$ is nothing other than the theory of ordered commutative rings with unit element. ❑

This theorem was originally proved by Tarski in [Ta2]. While thanks to a host of model-theoretical methods the above proof is significantly shorter than Tarski's original, the latter is still of import in many algorithmic questions. Decidability and eliminability of quantifiers in RCF have great impact also on other fields of research, in particular on the foundations of geometry which are not treated in this book.

Remark 2. Due to the completeness of RCF, one may also say that the first-order theory of the ordered field $\mathbb{R}$ allows quantifier elimination. Incidentally, the quantifiers in RCF are not eliminable if the order, which is definable in RCF, is not considered as a basic relation. Also the (complete) theory $T := Th(\mathbb{R}, <, 0, 1, +, -, \cdot, \exp)$ with the exponential function exp in the language does not allow quantifier elimination. T is nonetheless model complete as was shown in [Wi]. Because of completeness, the decision problem for T reduces to the still unsolved axiomatization problem, whose solution hinges on the unanswered problem concerning transcendental numbers, *Schanuel's conjecture*, which lies outside the scope of logic (consult the Internet). A particular question related to the conjecture is whether or not e^e is transcendental.

Exercises

1. Show that the theory ZG is model complete in its language, and even in the language $\mathcal{L}\{0, 1, +, -\}$.

2. A structure elementarily equivalent to $(\mathbb{N}, 0, 1, +, <)$ is called an $\mathbb{N}$-*semigroup*. Axiomatize the theory of $\mathbb{N}$-semigroups and show (by tracing back to ZG) that it allows quantifier elimination in $\mathcal{L}\{0, 1, +, <, 1|, 2|, \dots\}$.

3. Let RCF° be the theory of real closed fields without order as a basic notion. Prove that the $\exists y$ is not eliminable in RCF° from $\alpha(x) = \exists y\, y \cdot y = x$.

4. Show that RCF is axiomatized alternatively by the axioms for ordered fields and the continuity scheme CS in **3.3** page 86.

5. Show that the theory T of divisible ordered abelian groups allows quantifier elimination.

5.7 Reduced Products and Ultraproducts

In order to merely indicate the usefulness of the following constructions consider for instance $\mathbb{Z}^n$, a direct power of the additive group $\mathbb{Z}$. By component-wise verification of the axioms it can be shown that $\mathbb{Z}^n$ is itself an abelian group ($n \geqslant 2$). But in this and similar examples we can save ourselves the bother, because by Theorem 7.5 below a Horn sentence valid in all $\mathcal{A}_i$ is also valid in the product $\prod_{i\in I}\mathcal{A}_i$, and the group axioms are Horn sentences in each reasonable signature.

Let $(\mathcal{A}_i)_{i\in I}$ be a family of $\mathcal{L}$-structures and F a proper filter on I ($\neq \emptyset$, cf. **1.5**). We define a relation $\approx_F$ on the domain B of the product $\mathcal{B} := \prod_{i\in I}\mathcal{A}_i$ by

$$a \approx_F b \;\Leftrightarrow\; \{i \in I \mid a_i = b_i\} \in F.$$

This is an equivalence relation on the set B. Indeed, let $I_{a=b} := \{i \in I \mid a_i = b_i\}$. $\approx_F$ is reflexive (since $I_{a=a} = I \in F$) and trivially symmetric, but also transitive, because $I_{a=b}, I_{b=c} \in F \;\Rightarrow\; I_{a=c} \in F$, thanks to $I_{a=b} \cap I_{b=c} \subseteq I_{a=c}$.

Furthermore $\approx_F$ is a *congruence* in the algebraic reduct of $\mathcal{B}$. To see this let f be an n-ary function symbol and $\vec{a} \approx_F \vec{b}$, which for $\vec{a} = (a^1, \ldots, a^n)$, $\vec{b} = (b^1, \ldots, b^n)$ in B^n abbreviates $a^1 \approx_F b^1, \ldots, a^n \approx_F b^n$. Then $I_{\vec{a}=\vec{b}} := \bigcap_{\nu=1}^{n} I_{a^\nu = b^\nu}$ belongs to F. Since certainly $I_{\vec{a}=\vec{b}} \subseteq I_{f\vec{a}=f\vec{b}}$, we get $I_{f\vec{a}=f\vec{b}} \in F$ and hence $f^{\mathcal{B}}\vec{a} \approx_F f^{\mathcal{B}}\vec{b}$.

Now let $C := \{a/F \mid a \in B\}$, where a/F denotes the congruence class of $\approx_F$ to which $a \in B$ belongs. Thus, $a/F = b/F \;\Leftrightarrow\; I_{a=b} \in F$. C becomes the domain of some $\mathcal{L}$-structure $\mathcal{C}$ in that first the operations $f^{\mathcal{C}}$ are defined in a canonical way. With $\vec{a}/F := (a^1/F, \ldots, a^n/F)$ set $f^{\mathcal{C}}(\vec{a}/F) := (f^{\mathcal{B}}\vec{a})/F$. This definition is sound because $\approx_F$ is a congruence. For constant symbols c let of course $c^{\mathcal{C}} := c^{\mathcal{B}}/F$.

Similar to the identity, the relation symbols are interpreted in $\mathcal{C}$ as follows:

$$r^{\mathcal{C}}\vec{a}/F :\Leftrightarrow\; I_{r\vec{a}} \in F \quad \big(I_{r\vec{a}} := \{i \in I \mid r^{\mathcal{A}_i}\vec{a}_i\},\; \vec{a}_i := (a_i^1, \ldots, a_i^n)\big).$$

Also this definition is sound, since $I_{r\vec{a}} \in F$ and $\vec{a} \approx_F \vec{b}$ imply $I_{r\vec{b}} \in F$. Indeed, $\vec{a} \approx_F \vec{b}$ is equivalent to $I_{\vec{a}=\vec{b}} \in F$ and it is readily verified that $I_{r\vec{a}} \cap I_{\vec{a}=\vec{b}} \subseteq I_{r\vec{b}}$.

The $\mathcal{L}$-structure $\mathcal{C}$ so defined is called a *reduced product of the $\mathcal{A}_i$ by the filter F* and is denoted by $\prod_{i\in I}^{F}\mathcal{A}_i$ (some authors denote it by $\prod_{i\in I}\mathcal{A}_i/F$). Imagining a filter F as a system of subsets of I each of which contains "almost all indices," one may think of $\prod_{i\in I}^{F}\mathcal{A}_i$ as arising from $\mathcal{B} = \prod_{i\in I}\mathcal{A}_i$ by identification of those $a, b \in B$ for which the ith projections are the same for almost all indices i.

Let $\mathcal{C} = \prod_{i\in I}^{F}\mathcal{A}_i$. For $w\colon Var \to B$ $(= \prod_{i\in I} A_i)$ the valuation $x \mapsto (x^w)_i$ to A_i is denoted by w_i, so that $x^w = (x^{w_i})_{i\in I}$. Induction on t yields $t^w = (t^{w_i})_{i\in I}$. Define the valuation $w/F \to C$ by $x^{w/F} = x^w/F$. This setting generalizes inductively to

(1) $t^{w/F} = t^w/F$, for all terms t and valuations $w\colon Var \to B$.

To verify (1) consider $(f\vec{t})^{w/F} = f^{\mathcal{C}}(\vec{t}^{\,w/F}) = f^{\mathcal{C}}(\vec{t}^{\,w}/F) = f^{\mathcal{B}}(\vec{t}^{\,w})/F = (f\vec{t})^w/F$. It is easily seen that each $w'\colon Var \to C$ is of the form w/F for suitable $w\colon Var \to B$.

Let $w\colon Var \to B$ and $\alpha \in \mathcal{L}$. Define $I^w_\alpha := \{i \in I \mid \mathcal{A}_i \vDash \alpha\,[w_i]\}$. Then holds

(2) $I^w_{\exists x\beta} \subseteq I^{w'}_\beta$ for some $a \in B$ and $w' = w\frac{a}{x}$.

Indeed, let $i \in I^w_{\exists x\beta}$, i.e., $\mathcal{A}_i \vDash \exists x\beta\,[w_i]$. Choose some $a_i \in A_i$ with $\mathcal{A}_i \vDash \alpha\,[w_i\frac{a_i}{x}]$. For $i \notin I^w_{\exists x\beta}$ pick up *any* $a_i \in A_i$. Then clearly (2) holds with $a = (a_i)_{i\in I}$ and $w' = w\frac{a}{x}$.

The case that F is an ultrafilter on I is of particular interest. By Theorem 7.1, all elementary properties valid in almost all factors carry over to the reduced product, which in this case is called an *ultraproduct*. If $\mathcal{A}_i = \mathcal{A}$ for all $i \in I$ then $\prod^F_{i\in I}\mathcal{A}_i$ is termed an *ultrapower of* $\mathcal{A}$, denoted by $\mathcal{A}^I/F$. The importance of ultrapowers is underlined by Shelah's theorem (not proved here) that $\mathcal{A} \equiv \mathcal{B}$ iff $\mathcal{A}$ and $\mathcal{B}$ have isomorphic ultrapowers. The proof of Theorem 7.1 uses mainly filter properties; the specific ultrafilter property is applied only for confirming $I^w_{\neg\alpha} \in F \Leftrightarrow I^w_\alpha \notin F$.

Theorem 7.1 (Łoś's ultraproduct theorem). *Let $\mathcal{C} = \prod^F_{i\in I}\mathcal{A}_i$ be an ultraproduct of the $\mathcal{L}$-structures $\mathcal{A}_i$. Then for all $\alpha \in \mathcal{L}$ and $w\colon Var \to \prod_{i\in I} A_i$,*

$$(*)\quad \mathcal{C} \vDash \alpha[w/F] \Leftrightarrow I^w_\alpha \in F.$$

Proof by induction on α. $(*)$ is obtained for equations $t_1 = t_2$ as follows:

$$\begin{aligned}
\mathcal{C} \vDash t_1 = t_2\,[w/F] &\;\Leftrightarrow\; t_1^{w/F} = t_2^{w/F} \;\Leftrightarrow\; t_1^w/F = t_2^w/F \quad \big(\text{by (1)}\big)\\
&\;\Leftrightarrow\; \{i \in I \mid t_1^{w_i} = t_2^{w_i}\} \in F \quad \big(t^w = (t^{w_i})_{i\in I}\big)\\
&\;\Leftrightarrow\; \{i \in I \mid \mathcal{A}_i \vDash t_1 = t_2\,[w_i]\} \in F \;\Leftrightarrow\; I^w_{t_1=t_2} \in F.
\end{aligned}$$

One similarly proves $(*)$ for prime formulas of the form $r\vec{t}$. Induction steps:

$$\begin{aligned}
\mathcal{C} \vDash \alpha\wedge\beta\,[w/F] \;\Leftrightarrow\; C \vDash \alpha, \beta\,[w/F] &\;\Leftrightarrow\; I^w_\alpha, I^w_\beta \in F \quad \text{(induction hypothesis)}\\
&\;\Leftrightarrow\; I^w_\alpha \cap I^w_\beta \in F \quad \text{(filter property)}\\
&\;\Leftrightarrow\; I^w_{\alpha\wedge\beta} \in F \quad \text{(since } I^w_{\alpha\wedge\beta} = I^w_\alpha \cap I^w_\beta\text{)}.
\end{aligned}$$

Further, $\mathcal{C} \vDash \neg\alpha\,[w/F] \Leftrightarrow \mathcal{C} \nvDash \alpha\,[w/F] \Leftrightarrow I^w_\alpha \notin F \Leftrightarrow I \setminus I^w_\alpha \in F \Leftrightarrow I^w_{\neg\alpha} \in F$. Now let $I^w_{\forall x\alpha} \in F$, $a \in \prod_{i\in I} A_i$, and $w' := w\frac{a}{x}$. Since $I^w_{\forall x\alpha} \subseteq I^{w'}_\alpha$, also $I^{w'}_\alpha \in F$. Hence, $\mathcal{C} \vDash \alpha\,[w']$ by the induction hypothesis. a was arbitrary, so $\mathcal{C} \vDash \forall x\alpha\,[w/F]$. The converse is with $\beta := \neg\alpha$ equivalent to $I^w_{\exists x\beta} \in F \Rightarrow \mathcal{C} \vDash \exists x\beta\,[w/F]$. This follows from (2) since $(*)$ holds by the induction hypothesis for α, hence also for $\neg\alpha$. ❑

Corollary 7.2. *A sentence α is valid in the ultraproduct $\prod^F_{i\in I}\mathcal{A}_i$ iff α is valid in "almost all" $\mathcal{A}_i$, that is, $\{i \in I \mid \mathcal{A}_i \vDash \alpha\} \in F$. In particular, $\mathcal{A}^I/F \vDash \alpha \Leftrightarrow \mathcal{A} \vDash \alpha$. In other words, an ultrapower of $\mathcal{A}$ is elementarily equivalent to $\mathcal{A}$.*

The last claim is clear since the validity of α in a structure does not depend on the valuation chosen. The ultrapower case can be further strengthened to $\mathcal{A} \preccurlyeq \mathcal{A}^I/F$ (Exercise 2), useful for the construction of special nonstandard models, for instance. From the countless applications of ultraproducts, we present here a very short proof of the compactness theorem for arbitrary first-order languages. The proof is tricky, but undoubtedly the most elegant proof of the compactness theorem.

Theorem 7.3. *Let $X \subseteq \mathcal{L}$ and let I be the set of all finite subsets of X. Assume that every $i \in I$ has a model $(\mathcal{A}_i, w_i)$. Then there exists an ultrafilter F on I such that $\prod_{i\in I}^F \mathcal{A}_i \vDash X\,[w/F]$, where $x^w = (x^{w_i})_{i\in I}$ for $x \in \mathrm{Var}$. In short, if every finite subset of X has a model then the same applies to the whole of X.*

Proof. Let $J_\alpha := \{i \in I \mid \alpha \in i\}$ for $\alpha \in X$. The intersection of finitely many members of $E := \{J_\alpha \mid \alpha \in X\}$ is $\neq \emptyset$; for instance $\{\alpha_0, \dots, \alpha_n\} \in J_{\alpha_0} \cap \dots \cap J_{\alpha_n}$. By the ultrafilter theorem (page 28), there exists an ultrafilter $F \supseteq E$. If $\alpha \in X$ and $i \in J_\alpha$ (that is, $\alpha \in i$) then $\mathcal{A}_i \vDash \alpha\,[w_i]$. Consequently, $J_\alpha \subseteq I_\alpha^w$; hence $I_\alpha^w \in F$. Therefore, $\prod_{i\in I}^F \mathcal{A}_i \vDash \alpha\,[w/F]$ by Theorem 7.1 as claimed. ❏

A noteworthy consequence of these results is the following theorem; by Shelah's theorem mentioned above condition (a) can be converted in a purely algebraic one.

Theorem 7.4. *Let $\boldsymbol{K}_\mathcal{L}$ be the class of all $\mathcal{L}$-structures, and $\boldsymbol{K} \subseteq \boldsymbol{K}_\mathcal{L}$. Then*

(a) *$\boldsymbol{K}$ is Δ-elementary iff $\boldsymbol{K}$ is closed under elementary equivalence and under ultraproducts,*

(b) *$\boldsymbol{K}$ is elementary $\Leftrightarrow$ $\boldsymbol{K}$ is closed under elementary equivalence and both $\boldsymbol{K}$ and $\setminus\boldsymbol{K}$ $(= \boldsymbol{K}_\mathcal{L} \setminus \boldsymbol{K})$ are closed under ultraproducts.*

Proof. (a): A Δ-elementary class is clearly closed under elementary equivalence. The rest of direction $\Rightarrow$ holds by Theorem 7.1. $\Leftarrow$: Suppose $T := \mathit{Th}\,\boldsymbol{K}$ and $\mathcal{A} \vDash T$ and let I be the set of all finite subsets of $\mathit{Th}\,\mathcal{A}$. For each $i = \{\alpha_1, \dots, \alpha_n\} \in I$ there exists some $\mathcal{A}_i \in \boldsymbol{K}$ such that $\mathcal{A}_i \vDash i$, for otherwise $\bigvee_{\nu=1}^n \neg\alpha_\nu \in T$, which contradicts $i \subseteq T$. According to Theorem 7.3 (with $X = \mathit{Th}\,\mathcal{A}$) there exists a $\mathcal{C} := \prod_{i\in I}^F \mathcal{A}_i \vDash \mathit{Th}\,\mathcal{A}$, and if $\mathcal{A}_i \in \boldsymbol{K}$ then so too $\mathcal{C} \in \boldsymbol{K}$. Since $\mathcal{C} \vDash \mathit{Th}\,\mathcal{A}$ we know that $\mathcal{C} \equiv \mathcal{A}$, and therefore $\mathcal{A} \in \boldsymbol{K}$. This shows that $\mathcal{A} \vDash T \Rightarrow \mathcal{A} \in \boldsymbol{K}$. Hence $\mathcal{A} \vDash T \Leftrightarrow \mathcal{A} \in \boldsymbol{K}$, i.e., $\boldsymbol{K}$ is Δ-elementary. (b): $\Rightarrow$ is obvious by (a), because for $\boldsymbol{K} = \mathrm{Md}\,\alpha$ we have $\setminus\boldsymbol{K} = \mathrm{Md}\,\neg\alpha$. $\Leftarrow$: By (a), $\boldsymbol{K} = \mathrm{Md}\,S$ for some $S \subseteq \mathcal{L}^0$. Let I be the set of all nonempty subsets of S. We claim $(*)$: there is some $i = \{\alpha_0, \dots, \alpha_n\} \in I$ such that $\mathrm{Md}\,i \subseteq \boldsymbol{K}$. Otherwise let $\mathcal{A}_i \vDash i$ such that $\mathcal{A}_i \in \setminus\boldsymbol{K}$ for all $i \in I$. Then there exists an ultraproduct $\mathcal{C}$ of the $\mathcal{A}_i$ such that $\mathcal{C} \in \setminus\boldsymbol{K}$ and $\mathcal{C} \vDash i$ for all $i \in I$; hence $\mathcal{C} \vDash S$. This is a contradiction to $\mathrm{Md}\,S \subseteq \boldsymbol{K}$. So $(*)$ holds. Since also $\boldsymbol{K} = \mathrm{Md}\,S \subseteq \mathrm{Md}\,i$, we obtain $\boldsymbol{K} = \mathrm{Md}\,i = \mathrm{Md}\bigwedge_{\nu\leqslant n}\alpha_\nu$. ❏

Application. Let $\boldsymbol{K}$ be the (Δ-elementary) class of all fields of characteristic 0. We show that $\boldsymbol{K}$ is not elementary, and thus in a new way that $\mathit{Th}\,\boldsymbol{K}$ is not finitely axiomatizable. Let $\mathcal{P}_i$ denote the prime field of characteristic p_i (p_0=2, p_1=3, …) and let F be a nontrivial ultrafilter on $\mathbb{N}$. We claim that the field $\prod_{i\in\mathbb{N}}^F \mathcal{P}_i$ has characteristic 0. Indeed, $\{i \in I \mid \mathcal{P}_i \vDash \neg\mathtt{char}_p\}$ is for a given prime p certainly cofinite and belongs to F, so that $\prod_{i\in\mathbb{N}}^F \mathcal{P}_i \vDash \neg\mathtt{char}_p$ for all p. Hence $\setminus\boldsymbol{K}$ is not closed under ultraproducts and so by Theorem 7.4(b), $\boldsymbol{K}$ cannot be elementary.

We now turn to reduced products. Everything said below on them remains valid for direct products; these are the special case with the minimal filter $F = \{I\}$. More precisely, $\prod_{i\in I}^{\{I\}} \mathcal{A}_i \simeq \prod_{i\in I} \mathcal{A}_i$. Filters are always proper in the sequel.

Theorem 7.5. *Let $\mathcal{C} = \prod_{i\in I}^{F} \mathcal{A}_i$ be a reduced product, $w\colon Var \to \prod_{i\in I} A_i$, and α a Horn formula from the corresponding first-order language. Then*

$$(\star) \quad I_\alpha^w \in F \Rightarrow \mathcal{C} \vDash \alpha\,[w/F].$$

In particular, a Horn sentence valid in almost all $\mathcal{A}_i$ is also valid in $\mathcal{C}$.

Proof by induction on the construction of Horn formulas. For prime formulas the converse of $(\star)$ is also valid, because in the proof of $(*)$ from Theorem 7.1 for prime formulas no specific ultrafilter property was used. Moreover, if α is prime then $I_{\neg\alpha}^w \in F \Rightarrow I_\alpha^w \notin F \Rightarrow \mathcal{C} \nvDash \alpha\,[w/F] \Rightarrow \mathcal{C} \vDash \neg\alpha\,[w/F]$. Hence, $(\star)$ is correct for all literals. Now suppose $(\star)$ for a prime formula α and a basic Horn formula β, and let $I_{\alpha\to\beta}^w \in F$. We show that $\mathcal{C} \vDash \alpha \to \beta\,[w/F]$. Let $\mathcal{C} \vDash \alpha\,[w/F]$. Then $I_\alpha^w \in F$ since α is prime. $I_\alpha^w \cap I_{\alpha\to\beta}^w \subseteq I_\beta^w$ leads to $I_\beta^w \in F$; hence $\mathcal{C} \vDash \beta\,[w/F]$ by the induction hypothesis. This shows that $\mathcal{C} \vDash \alpha \to \beta\,[w/F]$ and proves $(\star)$ for all basic Horn formulas. Induction on $\wedge$ and $\forall$ proceeds as in Theorem 7.1 and the $\exists$-step easily follows with the help of (2) above. ❏

According to this theorem the model classes of Horn theories are always closed under reduced products, in particular under direct products. This result strengthens Exercise 1 in **4.1** significantly. We mention finally that also the converse holds: every theory with a model class closed with respect to reduced products is a Horn theory. But the proof of this claim, presented in [CK], is essentially more difficult than that for the similar-sounding Theorem 4.4.

Exercises

1. Show that $\prod_{i\in I}^{F} \mathcal{A}_i$ is isomorphic to $\mathcal{A}_{i_0}$ for some $i_0 \in I$ if F is a trivial ultrafilter. This applies e.g. to ultraproducts on a finite index set (Exercise 3 in **1.5**). Thus, ultraproducts are interesting only if the index set I is infinite.

2. Prove that $\mathcal{A}$ is elementarily embeddable into every ultrapower $\mathcal{A}^I/F$.

3. (Basic in nonclassical logics). Let $\vDash_{\boldsymbol{K}} := \bigcap\{\vDash_{\mathcal{A}} \mid \mathcal{A} \in \boldsymbol{K}\}$ be the consequence relation defined by a class $\boldsymbol{K}$ of L-matrices (page 40). Show that $\vDash_{\boldsymbol{K}}$ is finitary if $\boldsymbol{K}$ is closed under ultraproducts (which is the case, for instance, if $\boldsymbol{K} = \{\mathcal{A}\}$ with finite $\mathcal{A}$). Thus, $\vDash_{\mathcal{A}}$ is finitary for each finite logical matrix.

4. Let $\mathcal{A}, \mathcal{B}$ be Boolean algebras. Prove that $\mathcal{A} \vDash \alpha \Leftrightarrow \mathcal{B} \vDash \alpha$ for all universal Horn sentences α. This holds in particular for identities and quasi-identities. Every sentence of this kind valid in $\mathcal{2}$ is therefore valid in all Boolean algebras.

Chapter 6

Incompleteness and Undecidability

Gödel's fundamental results concerning the incompleteness of formal systems sufficiently rich in content, along with Tarski's on the nondefinability of the notion of truth and Church's on the undecidability of logic, as well as other undecidability results, are all based on essentially the same arguments. A widely known popularization of Gödel's first incompleteness theorem runs as follows:

Consider a formalized axiomatic theory T that describes a given domain of objects $\mathcal{A}$ in a manner that we hope is complete. Moreover, suppose that T is capable of talking in its language $\mathcal{L}$ about its own syntax and proofs from its axioms. This is often possible if T has actually been devised to investigate other things (numbers or sets, say), namely by means of an internal encoding of the syntax of $\mathcal{L}$. Then the sentence γ: "I am unprovable in T" belongs to $\mathcal{L}$, where "I" refers precisely to the sentence γ (clearly, this possibility of self-reference has to be laid down in detail, which was the main work in [Go2]). *Then γ is true in $\mathcal{A}$ but unprovable in T.*

Indeed, if we assume γ is provable, then, like any other provable sentence in T, γ were true in $\mathcal{A}$ and so unprovable, since this is just what γ claims. Thus, our assumption leads to a contradiction. Hence, γ's assertion goes conform with truth; more precisely, γ belongs to the sentences from $\mathcal{L}$ true in $\mathcal{A}$. Put together, our goal of exhaustively capturing all theorems valid in $\mathcal{A}$ by means of the axioms of T has not been achieved and is in fact not achievable as we will see.

Clearly, the above is just a rough simplification of Gödel's Theorem which does not speak at all about a domain of objects, but is rather a *proof-theoretical* assertion the proof of which can be carried out in the framework of Hilbert's finitistic metamathematics. This in turn means about the same as being formalizable and provable in Peano arithmetic PA, introduced in **3.3**.

This result was a decisive point for a well founded criticism of *Hilbert's program*, which aimed to justify infinitistic methods by means of a finitistic understanding

of metamathematics. For a detailed description of what Hilbert was aiming at, see [Kl2] or consult [HB, Vol. 1]. The paradigm of a domain of objects in the above sense is, for a variety of reasons, the structure $\mathcal{N} = (\mathbb{N}, 0, \mathsf{S}, +, \cdot)$. Gödel's theorem states that even for $\mathcal{N}$ a complete axiomatic characterization in its language is impossible, a result with far-reaching consequences. In particular, PA, which aims at telling us as much as possible about $\mathcal{N}$, is shown to be incomplete.

PA is the center point of Chapter **7**. It is of special importance because classical number theory and large parts of discrete mathematics can be developed in it; all interesting combinatorial functions are definable. In addition, known methods for investigating mathematical foundations can be formalized and proved in PA. These methods have stood firm against all kinds of criticism, leaving aside some objections concerning the unrestricted use of two-valued logic, not discussed here.

Some of the steps in Gödel's proof require only modest suppositions regarding T, namely the numeralwise representability of relevant syntactical predicates and functions in T in the sense of **6.3**. It was one of Gödel's decisive discoveries that *all* the predicates required in γ's construction above are primitive recursive[1] and that all predicates and functions of this type are indeed representable in T. As remarked by Tarski and Mostowski, the latter works even in certain finitely axiomatizable, highly incomplete theories T and, in addition, covers all recursive functions. This yields not only the recursive undecidability of T and all its subtheories (in particular the theory $\mathit{Taut}_{\mathcal{L}}$), but also of all consistent extensions of T in its language $\mathcal{L}$.

From this it follows that the first incompleteness theorem as well as Church's and Tarski's results can all be obtained in one go, making essential use of the *fixed-point lemma* in **6.5**, also called the *diagonalization lemma* because it is shown by some kind of diagonalization on the primitive recursive substitution function. Its basic idea can even be recognized in the ancient liar paradox, and is also used in the foregoing popularization of the first incompleteness theorem.

In **6.1** we develop the theory of recursive and primitive recursive functions to the required extent. **6.2** deals with the arithmetization of syntax and of formal proofs. **6.3** and **6.4** treat the representability of recursive functions in axiomatic theories. In **6.5** all the aforementioned results are proved, while the deeper-lying second incompleteness theorem is dealt with in Chapter **7**. Section **6.6** concerns the transferability of decidability and undecidability by interpretation, and **6.7** describes the first-order arithmetical hierarchy, which vividly illustrates the close relationship between logic and recursion theory.

[1] All these predicates are also elementary in the recursion-theoretical sense, see e.g. [Mo], although it requires much more effort to verify this. Roughly speaking, the elementary functions are the "not too rapidly growing" primitive recursive functions. The exponential function $(m, n) \mapsto m^n$ is still elementary, however the hyperexponential function defined on page 186 is not.

6.1 Recursive and Primitive Recursive Functions

In this chapter, along with $i, \ldots, n$ we take $a, \ldots, e$ to denote natural numbers, unless stated otherwise. The set of all n-ary functions with arguments and values in $\mathbb{N}$ is denoted by $\mathbf{F}_n$. For $f \in \mathbf{F}_m$ and $g_1, \ldots, g_m \in \mathbf{F}_n$, we call $h : \vec{a} \mapsto h(g_1\vec{a}, \ldots, g_m\vec{a})$ the (canonical) *composition* of f and the g_i and write $h = f[g_1, \ldots, g_m]$. The arity of h is n. Analogously, let $P[g_1, \ldots, g_m]$ for $P \subseteq \mathbb{N}^m$ and $m > 0$ denote the n-ary predicate $\{\vec{a} \in \mathbb{N}^n \mid P(g_1\vec{a}, \ldots, g_m\vec{a})\}$.

In an intuitive sense $f \in \mathbf{F}_n$ is computable if there is an algorithm for computing $f\vec{a}$ for every $\vec{a}$ in finitely many steps. Sum and product are simple examples. There are uncountably many unary functions on $\mathbb{N}$, and because of the finiteness of every set of computation instructions, only countably many of these can be computable. Thus, there are noncomputable functions. This existence proof brings to mind the one for transcendental real numbers, based on the countability of the set of algebraic numbers. Coming up with concrete examples is, in both cases, less simple.

The computable functions in the intuitive sense have the following properties:

Oc: If $h \in \mathbf{F}_m$ and $g_1, \ldots, g_m \in \mathbf{F}_n$ are computable, so too is $f = h[g_1, \ldots, g_m]$.

Op: If $g \in \mathbf{F}_n$ and $h \in \mathbf{F}_{n+2}$ are computable then so is $f \in \mathbf{F}_{n+1}$, determined by

$$f(\vec{a}, 0) = g\vec{a}; \quad f(\vec{a}, \mathtt{S}b) = h(\vec{a}, b, f(\vec{a}, b)).$$

This are the so-called *recursion equations* for f. The function f is said *to result from g, h by primitive recursion*, or $f = \boldsymbol{Op}(g, h)$ for short.

Oμ: Let $g \in \mathbf{F}_{n+1}$ such that $\forall\vec{a}\,\exists b\, g(\vec{a}, b) = 0$. If g is computable then so is f, given by $f\vec{a} = \mu b[g(\vec{a}, b) = 0]$. Here the right-hand term denotes the smallest b with $g(\vec{a}, b) = 0$. f is said to result from g by the *μ-operation*.

Considering ***Oc***, ***Op***, and ***Oμ*** as generating operations for obtaining new functions from already-constructed ones, we state the following definition due to Kleene:

Definition. The set of p.r. (*primitive recursive*) functions consists of all functions on $\mathbb{N}$ that can be obtained by finitely many applications of ***Oc*** and ***Op*** starting with the following *initial functions*: the constant 0, the successor function $\mathtt{S}$, and the projection functions $\mathrm{I}^n_\nu : \vec{a} \mapsto a_\nu$ $(1 \leqslant \nu \leqslant n,\ n = 1, 2, \ldots)$.
With the additional generating schema ***Oμ*** one obtains the set of all *recursive* or *μ-recursive* functions. A predicate $P \subseteq \mathbb{N}^n$ is called p.r. or *recursive* (also *decidable*) provided the *characteristic function* χ_P of P has the respective property, defined by

$$\chi_P\vec{a} = \begin{cases} 1 & \text{in case } P\vec{a}, \\ 0 & \text{in case } \neg P\vec{a}. \end{cases}$$

Remark 1. By Dedekind's recursion theorem (see e.g. [Ra2]), ***Op*** defines exactly one function $f \in \mathbf{F}_n$ in the sense of set theory (cf. **3.4**). Note that for $n = 0$ the recursion equations reduce to $f0 = c$ and $f\mathtt{S}b = h(b, fb)$, where $c \in \mathbf{F}_0$ and $h \in \mathbf{F}_2$. If the condition $\forall \vec{a}\, \exists b\, g(\vec{a}, b) = 0$ in ***Oμ*** is omitted, then f is regarded as undefined for those $\vec{a}$ for which there is no b with $g(\vec{a}, b) = 0$. In this way the so-called partially recursive functions are defined, which, however, we will not require.

The following examples make it clear that by means of the I^n_ν our stipulations concerning arity in ***Oc*** and ***Op*** can be extensively relaxed. In the examples, however, we will still adjoin the normed notation each time in parentheses.

Examples. Let $\mathtt{S}^0 = \mathrm{I}^1_1$ and $\mathtt{S}^{k+1} = \mathtt{S}[\mathtt{S}^k]$, so that clearly $\mathtt{S}^k : a \mapsto a+k$. By ***Oc*** these functions are all p.r. The n-ary *constant functions* $\mathrm{K}^n_c : \vec{a} \mapsto c$ can be seen to be p.r. as follows: $\mathrm{K}^0_c = \mathtt{S}^c[0]$ $(c \geqslant 0)$, while $\mathrm{K}^1_c 0 = c \ \big(= \mathrm{K}^0_c\big)$ and $\mathrm{K}^1_c \mathtt{S}b = c \ \big(= \mathrm{I}^2_2(b, \mathrm{K}^1_c b)\big)$. For $n > 1$ we have simply $\mathrm{K}^n_c = \mathrm{K}^1_c[\mathrm{I}^n_1]$. Further, the recursion equations

$$a + 0 = a \ \big(= \mathrm{I}^1_1(a)\big); \quad a + \mathtt{S}b = \mathtt{S}(a + b) \ \big(= \mathtt{S}\mathrm{I}^3_3(a, b, a + b)\big)$$

show addition to be a p.r. function. Since $a \cdot 0 = 0 \ \big(= \mathrm{K}^1_0 a\big)$ and $a \cdot \mathtt{S}b = a \cdot b + a \ \big(= \mathrm{I}^3_3(a, b, a \cdot b) + \mathrm{I}^3_1(a, b, a \cdot b)\big)$, it follows that $\cdot$ is p.r. and entirely analogously so is $(a, b) \mapsto a^b$. Also the predecessor function $\mathtt{Pd}$ is p.r. because

$$\mathtt{Pd}\, 0 = 0 \ ; \quad \mathtt{Pd}(\mathtt{S}b) = b \ \big(= \mathrm{I}^2_1(b, \mathtt{Pd}\, b)\big).$$

"Cut-off subtraction" $\dot{-}$, given by $a \dot{-} b = a - b$ for $a \geqslant b$ and $a \dot{-} b = 0$ otherwise, is p.r. since $a \dot{-} 0 = a \ \big(= \mathrm{I}^1_1(a)\big)$ and $a \dot{-} \mathtt{S}b = \mathtt{Pd}(a \dot{-} b) \ \big(= \mathtt{Pd}\, \mathrm{I}^3_3(a, b, a \dot{-} b)\big)$. The absolute difference $|a - b|$ is p.r. because of $|a - b| = (a \dot{-} b) + (b \dot{-} a)$.

One sees easily that if f is p.r. (resp. recursive) then so too is every function that results from swapping, equating, or adjoining fictional arguments. For example, let $f \in \mathbf{F}_2$. For $f_1 := f[\mathrm{I}^2_2, \mathrm{I}^2_1]$ then $f_1(a, b) = f(b, a)$; for $f_2 := f[\mathrm{I}^1_1, \mathrm{I}^1_1]$ clearly $f_2 a = f(a, a)$, and for $f_3 := f[\mathrm{I}^3_1, \mathrm{I}^3_2]$ finally $f_3(a, b, c) = f(a, b)$, for all a, b, c.

From now on we will be more relaxed in writing down applications of ***Oc*** or ***Op***, and the I^n_ν will no longer explicitly appear. If $f \in \mathbf{F}_{n+1}$ is p.r. then so is the function $(\vec{a}, b) \mapsto \prod_{k<b} f(\vec{a}, k)$, since $\prod_{k<0} f(\vec{a}, k) = 1$, $\prod_{k<\mathtt{S}b} f(\vec{a}, k) = \prod_{k<b} f(\vec{a}, k) \cdot f(\vec{a}, b)$. The same holds for $(\vec{a}, b) \mapsto \sum_{k<b} f(\vec{a}, k)$, which is defined by $\sum_{k<0} f(\vec{a}, k) = 0$ and $\prod_{k<\mathtt{S}b} f(\vec{a}, k) = \sum_{k<b} f(\vec{a}, k) + f(\vec{a}, b)$. The *δ-function*, the characteristic function of the singleton $\{0\}$, is defined by $\delta 0 = 1$, $\delta \mathtt{S}n = 0$ and hence is p.r. With δ we easily obtain the characteristic function of the identity relation: $\chi_=(a, b) = \delta|a - b|$. This in turn implies that every finite subset $E = \{a_1, \ldots, a_n\}$ of $\mathbb{N}$ is p.r. because

$$\chi_E(a) \ = \ \chi_=(a, a_1) + \cdots + \chi_=(a, a_n) \quad (= 0 \text{ for } n = 0, \text{ i.e., } E = \emptyset).$$

$\neq$ is p.r. because $\chi_\neq(a, b) = \sigma|a - b|$ with the *signum function* σ, defined by $\sigma 0 = 0$, $\sigma \mathtt{S}n = 1$. Also $\leqslant$ is p.r. because $\chi_\leqslant(a, b) = \sigma(\mathtt{S}b \dot{-} a)$ as is easily verified.

Very important is the closure of the set of p.r. functions with respect to *definition by p.r.* (resp. *recursive*) *case distinction*: If P, g, h are p.r. (resp. recursive) then so

is f, defined by $f\vec{a} = g\vec{a} \cdot \chi_P\vec{a} + h\vec{a} \cdot \delta\chi_P\vec{a}$. Written in the familiar form,

$$f\vec{a} = \begin{cases} g\vec{a} & \text{in case } P\vec{a}, \\ h\vec{a} & \text{in case } \neg P\vec{a}. \end{cases}$$

A simple example is $(a, b) \mapsto \max(a, b)$, defined by $\max(a, b) = b$ if $a \leqslant b$ and $\max(a, b) = a$ otherwise. Almost all functions considered in number theory are p.r., in particular the *prime enumeration* $n \mapsto p_n$ (with $p_0 = 2,\ p_1 = 3, \dots$). The same is true for standard predicates like $\mid$ (divides) and prim (to be a prime number). This will all be verified after some general remarks.

Of fundamental importance is the hypothesis that recursive functions exhaust all the computable functions over $\mathbb{N}$. This hypothesis is called *Church's thesis*; all undecidability results are based on it. Though it is not at all obvious from looking at the definition of the recursive functions, all the variously defined computability concepts turn out to be equivalent, providing evidence in favor of the thesis. One of these concepts is computability by means of a *Turing machine* ([Tu]), a particularly simple abstract model of automated information processing. Also programming languages may be used to define computability, for instance PROLOG; see **4.4**.

Below we compile a list of the easily provable basic facts about p.r. and recursive predicates needed in the following. Further insights, above all concerning the form of their defining formulas, will emerge in **6.3** and thereafter. P, Q, R now denote exclusively predicates of $\mathbb{N}$. In order to simplify the notation of properties of such predicates, we use as metatheoretical abbreviations the prefixes $(\exists a{<}b)$, $(\exists a{\leqslant}b)$, $(\forall a{<}b)$, and $(\forall a{\leqslant}b)$ as in (B) below. Their meaning is self-explanatory.

(A) The set of p.r. (resp. recursive) predicates is closed under forming the complement, union, and intersection of predicates of the same arity, as well as under insertion of p.r. (resp. recursive) functions, and finally under swapping, equating, and adjoining fictional arguments. This is proved as follows: for $P \subseteq \mathbb{N}^n$, $\delta[\chi_P]$ is exactly the characteristic function of $\neg P := \mathbb{N}^n \setminus P$; furthermore $\chi_{P\cap Q} = \chi_P \cdot \chi_Q$ and $\chi_{P\cup Q} = \mathrm{sg}[\chi_P + \chi_Q]$ as well as $\chi_{P[g_1,\dots,g_m]} = \chi_P[g_1, \dots, g_m]$. Since $\chi_{\mathrm{graph}\, f}(\vec{a}, b)$ is the same as $\chi_{=}(f\vec{a}, b)$, graph f is p.r. if f is (though the converse need not hold, see the end of this section). All other mentioned closure properties are simply obtained from the corresponding properties of the characteristic functions.

(B) Let $P, Q, \dots \subseteq \mathbb{N}^{n+1}$. If $Q(\vec{a}, b) \Leftrightarrow (\forall k{<}b)P(\vec{a}, k)$, $R(\vec{a}, b) \Leftrightarrow (\exists k{<}b)P(\vec{a}, k)$, $Q'(\vec{a}, b) \Leftrightarrow (\forall k{\leqslant}b)P(\vec{a}, k)$, and $R'(\vec{a}, b) \Leftrightarrow (\exists k{\leqslant}b)P(\vec{a}, k)$ we say that Q, R, Q', R' result from P by *bounded quantification*. If P is p.r. so too are all these predicates, because $\chi_Q(\vec{a}, b) = \prod_{k<b} \chi_P(\vec{a}, k)$ and $\chi_R(\vec{a}, b) = \mathrm{sg}(\sum_{k<b} \chi_P(\vec{a}, k))$, and similarly if Q, R are replaced by Q', R'. The proofs of these equations are so simple that we pass over them. Briefly, the set of p.r. (resp. recursive) predicates is closed under bounded quantification. For instance, since $a \mid b \Leftrightarrow (\exists k{\leqslant}b)[a \cdot k = b]$, also $\mid$ is p.r.

So too is the predicate prim, because $\mathsf{prim}\,p \Leftrightarrow p \neq 0,1 \;\&\; (\forall a{<}p)[a \mid p \Rightarrow a{=}1]$. Note that $a \mid p \Rightarrow a = 1$ is equivalent (at the metatheoretical level) to $a \not\mid p \vee a = 1$ and is therefore the union of p.r. predicates. Hence, this predicate is indeed p.r.

(C) Suppose $P \subseteq \mathbb{N}^{n+1}$ satisfies $\forall \vec{a}\, \exists b\, P(\vec{a}, b)$ and let $f(\vec{a}) = \mu k[P(\vec{a}, k)]$ be the smallest k such that $P(\vec{a}, k)$. Then by $\boldsymbol{O\mu}$, if P is recursive so too is f, because $f\vec{a} = \mu k[\delta\chi_P(\vec{a}, k) = 0]$; however, in general f is no longer p.r. provided P is p.r. This does hold, though, for the *bounded μ-operation*: if $P \subseteq \mathbb{N}^{n+1}$ is p.r. so too is the function f defined by $f(\vec{a}, m) = \mu k{\leqslant}m[P(\vec{a}, k)]$. Here let

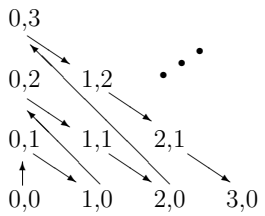
$$\mu k{\leqslant}m[P(\vec{a}, k)] = \begin{cases} \text{the smallest } k \leqslant m \text{ such that } P(\vec{a}, k), \text{ if such a } k \text{ exists,} \\ m \quad \text{otherwise.} \end{cases}$$

Clearly $f(\vec{a}, 0) = 0$, and $f(\vec{a}, \mathtt{S}m) = f(\vec{a}, m)$ if $(\exists k{\leqslant}m)P(\vec{a}, k)$, and $f(\vec{a}, \mathtt{S}m) = \mathtt{S}m$ otherwise. To convert this into a normed recursion we define a p.r. function g by

$$g(\vec{a}, m, b) = \begin{cases} b \text{ if } (\exists k{\leqslant}m)P(\vec{a}, k), \\ \mathtt{S}m \quad \text{otherwise.} \end{cases}$$

Then $f(\vec{a}, \mathtt{S}m) = g(\vec{a}, m, f(\vec{a}, m))$ is easily confirmed. Therefore, f is indeed p.r.

Let $h \in \mathbf{F}_n$ be p.r. and define $\mu k{\leqslant}h\vec{a}[P(\vec{a}, k)] := \mu k{\leqslant}m[P(\vec{a}, k) \;\&\; m = h\vec{a}]$. Then also $\vec{a} \mapsto \mu k{\leqslant}h\vec{a}[P(\vec{a}, k)]$ is p.r. A useful application is the *pairing function* $\wp$, a bijective mapping from $\mathbb{N}^2$ to $\mathbb{N}$, defined by $\wp(a, b) = \sum_{i \leqslant a+b} i \; + a$. It enumerates the pairs (a, b) as in the figure (Exercise 2). One can see in yet another way that $\wp$ is p.r. By a well-known arithmetical formula, $\wp(a, b) = \frac{1}{2}(a + b)(a + b + 1) + a$. Using the bounded μ-operation we get the following equation: $\wp(a, b) = \mu k{\leqslant}(3a + b + 1)^2\,[2k = (a + b)^2 + 3a + b]$.



Here another application of the bounded μ-operation: let $\mathrm{lcm}\{a_\nu \mid \nu{\leqslant}n\}$ denote the least common multiple of $a_0, \ldots, a_n$. Then $n \mapsto \mathrm{lcm}\{f\nu \mid \nu{\leqslant}n\}$ is p.r. provided f is, simply because of the equation $\mathrm{lcm}\{f\nu \mid \nu{\leqslant}n\} = \mu k{\leqslant}\prod_{\nu \leqslant n} f\nu\,[(\forall \nu{\leqslant}n) f\nu \mid k]$.

Still another application of the bounded μ-operation is a rigorous proof that the prime number enumeration is p.r. If p is prime than $p!{+}1$ is certainly not divisible by a prime $q \leqslant p$, for $q \mid p!{+}1$ and $q \mid p!$ yield $q \mid p!{+}1{-}p! = 1$ and hence the contradiction $q \mid 1$. Thus, a prime divisor of $p! + 1$ is a new prime. What is important here is that the smallest prime following p is $\leqslant p!{+}1$. Therefore, the function $n \mapsto p_n$ is uniquely characterized by the equations

$$(*) \quad p_0 = 2\,; \quad p_{n+1} = \mu q{\leqslant}p_n!{+}1[q\,\mathsf{prim} \;\&\; q > p_n].$$

Also $(*)$ is an application of $\boldsymbol{Op}$, because with $f : (a, b) \mapsto \mu q{\leqslant}b[q\,\mathsf{prim} \;\&\; q > a]$, $g : (a, b) \mapsto f(a, b! + 1)$ is p.r. as well, and the second equation in $(*)$ can be written $p_{n+1} = g(n, p_n)$ as is easily verified. Hence, $n \mapsto p_n$ is indeed p.r.

Remark 2. Unlike the set of p.r. functions, the set of μ-recursive functions can no longer be effectively enumerated; indeed, not even all unary ones: if $(f_n)_{n\in\mathbb{N}}$ were such an effective enumeration then $f : n \mapsto f_n(n)+1$ would be computable and hence recursive by Church's thesis. Thus, $f = f_m$ for some m, so that $f_m(m) = f(m) = f_m(m) + 1$, a contradiction. While this seemingly speaks against the thesis, it can in fact be eliminated from the argument using some basic recursion theory. (C) clarifies the distinction between p.r. and recursive functions to some extent. The former can be computed with an effort that can in principle be estimated in advance, whereas the existence condition in the unbounded μ-operation may be nonconstructive, so that even crude estimations of the effort required for computation are impossible. It is wrong to think that non-p.r. computable functions are "growing too fast." There are examples of such functions taking values from $\{0, 1\}$ only. On the other hand, it is simply impossible to compute the digits of $f6$ for the p.r. function $f : n \mapsto \underbrace{2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}}_{n}$. While $f5$ has "only" 19 729 digits, the number $f6$ is already astronomical.

The following considerations are required in **6.2**. They concern the encoding of finite sequences of numbers of arbitrary length. There are basically several possibilities for doing this. One of these is to use the pairing function $\wp$ (or a similar one, cf. [Shoe]) repeatedly. Here we choose the particularly intuitive encoding from [Go2], based on the prime enumeration $n \mapsto p_n$ and the unique prime factorization.

Definition. $\langle a_0, \ldots, a_n\rangle := p_0^{a_0+1} \cdots p_n^{a_n+1}$ $(= \prod_{i\leqslant n} p_i^{a_i+1})$ is called the *Gödel number* of the sequence $(a_0, \ldots, a_n)$. The empty sequence has the Gödel number 1, also denoted by $\langle\rangle$. Let *GN* denote the set of all Gödel numbers.

Clearly, $\langle a_0, \ldots, a_n\rangle = \langle b_0, \ldots, b_m\rangle \;\Rightarrow\; m = n \;\&\; a_i = b_i$ for $i = 1, \ldots, n$. Also, $(a_0, \ldots, a_n) \mapsto \langle a_0, \ldots, a_n\rangle$ is certainly p.r. and by (A), (B) above, so is *GN*, since

$$a \in GN \;\Leftrightarrow\; a \neq 0 \;\&\; (\forall p\leqslant a)(\forall q\leqslant p)[\mathsf{prim}\, p, q \;\&\; p \mid a \;\Rightarrow\; q \mid a].$$

We now create a small provision of p.r. functions useful for the encoding of syntax in **6.2**. Using (C) we define a p.r. function $a \mapsto \ell a$ as follows:

$$\ell a = \mu k\leqslant a[p_k \nmid a].$$

We call ℓa for a Gödel number a the "length" of a, since clearly $\ell 1 = 0$, and for $a = \langle a_0, \ldots, a_n\rangle = \prod_{i\leqslant n} p_i^{a_i+1}$ we have $\ell a = n + 1$, because $k = n + 1$ is the smallest index such that $p_k \nmid a$. Note that $k \leqslant a$ is satisfied since $p_a \nmid a$ in view of $p_a > a$. Also the binary operation $(a, i) \mapsto (a)_i$ is p.r. where the term $(a)_i$ is defined by

$$(a)_i = \mu k\leqslant a[p_i^{k+2} \nmid a].$$

This is the "component-recognition function." $p_i^{k+1} \mid a$ and $p_i^{k+2} \nmid a$ imply $k = (a)_i$, hence $(\langle a_0, \ldots, a_n\rangle)_i = a_i$ for all $i \leqslant n$. This function, printed bold in order to catch the eye, always begins counting the components of a Gödel number with $i = 0$. Therefore, $(a)_{last} := (a)_{\ell a \dot{-} 1}$ is the last component of a Gödel number $a \neq 1$. Which

values $(a)_i$ and ℓ have if their arguments are not Gödel numbers is not important; some authors redefine them so that their value is 0 in this case.

From the above definitions it follows that $a = \prod_{i<\ell a} p_i^{(a)_i+1}$ for Gödel numbers a including $a = 1$. Next we define the *arithmetical concatenation* $*$ by

$$a * b = a \cdot \prod_{i<\ell b} p_{\ell a+i}^{(b)_i+1} \text{ for } a, b \in \mathit{GN} \text{ and } a * b = 0 \text{ otherwise.}$$

Obviously, $\langle a_1, \ldots, a_n\rangle * \langle b_1, \ldots, b_m\rangle = \langle a_1, \ldots, a_n, b_1, \ldots, b_m\rangle$, so that GN is closed under $*$. Moreover, $a, b \leqslant a * b$ whenever $a, b \in \mathit{GN}$ as immediately follows from the definition of $*$. Note also that $a * b \in \mathit{GN} \Rightarrow a, b \in \mathit{GN}$, for arbitrary a, b. Clearly, $*$ is p.r. This function is useful for, among other things, a powerful generalization of ***Op***, the course-of-values recursion explained below.

To every $f \in \mathbf{F}_{n+1}$ corresponds a function $\bar{f} \in \mathbf{F}_{n+1}$ given by

$$\bar{f}(\vec{a}, 0) = \langle\rangle \ \big(= 1\big); \quad \bar{f}(\vec{a}, b) = \langle f(\vec{a}, 0), \ldots, , f(\vec{a}, b-1)\rangle \text{ for } b > 0.$$

$\bar{f}$ encodes the course of values of f. Now let F be a given function in $\mathbf{F}_{n+2}$. Then just as for ***Op*** there is exactly one $f \in \mathbf{F}_{n+1}$ satisfying the functional equation

$$\boldsymbol{Oq}: \quad f(\vec{a}, b) = F(\vec{a}, b, \bar{f}(\vec{a}, b)).$$

Namely, it holds that $f(\vec{a}, 0) = F(\vec{a}, 0, \langle\rangle) = F(\vec{a}, 0, 1)$, $f(\vec{a}, 1) = F(\vec{a}, 1, \langle f(\vec{a}, 0)\rangle)$, $f(\vec{a}, 2) = F(\vec{a}, 2, \langle f(\vec{a}, 0), f(\vec{a}, 1)\rangle)$, etc. In ***Oq***, $f(\vec{a}, b)$ in general depends for $b > 0$ on all values $f(\vec{a}, 0), \ldots, f(\vec{a}, b-1)$, not just on $f(\vec{a}, b-1)$ as in ***Op***. Therefore ***Oq*** is called the schema of *course-of-values recursion*. A simple example is the Fibonacci sequence $(fn)_{n\in\mathbb{N}}$, defined by $f0 = 0$, $f1 = 1$ and $fn = f(n-1) + f(n-2)$ for $n \geqslant 2$. The F in "normal form" ***Oq*** is given here by $F(b, c) = b$ for $b \leqslant 1$ and $F(b, c) = (c)_{b-1} + (c)_{b-2}$ otherwise. Indeed, $f0 = 0 = F(0, \bar{f}0)$, $f1 = 1 = F(1, \bar{f}1)$, and $fn = f(n-1) + f(n-2) = (\bar{f}n)_{n-1} + (\bar{f}n)_{n-2} = F(n, \bar{f}n)$ whenever $n \geqslant 2$.

Op is a special case of ***Oq***. If $f = \boldsymbol{Op}(g, h)$ and F is defined by the equations $F(\vec{a}, 0, c) = g(\vec{a})$ and $F(\vec{a}, \mathtt{S}b, c) = h(\vec{a}, b, (c)_b)$, then f also satisfies ***Oq*** with this F as may straightforwardly be checked while observing that $f(\vec{a}, b) = \big(\bar{f}(\vec{a}, \mathtt{S}b)\big)_b$.

Theorem 1.1. *Let f satisfy **Oq**. If F is primitive recursive then so too is f.*

Proof. Since $\langle c_0, \ldots, c_b\rangle = \langle c_0, \ldots, c_{b-1}\rangle * \langle c_b\rangle$ for $b > 0$, the function $\bar{f}$ satisfies

$$\bar{f}(\vec{a}, 0) = 1; \quad \bar{f}(\vec{a}, \mathtt{S}b) = \bar{f}(\vec{a}, b) * \langle f(\vec{a}, b)\rangle = \bar{f}(\vec{a}, b) * \langle F(\vec{a}, b, \bar{f}(\vec{a}, b))\rangle.$$

The second equation can be written $\bar{f}(\vec{a}, \mathtt{S}b) = h(\vec{a}, b, \bar{f}(\vec{a}, b))$, where h defined by $h(\vec{a}, b, c) = c * \langle F(\vec{a}, b, c)\rangle$. With F also the function h is p.r. Hence, by ***Op***, $\bar{f}$ is p.r. But then also f, because in view of ***Oq***, f is a composition of p.r. functions. ❑

We now make precise the intuitive notion of recursive (or effective) enumerability. $M \subseteq \mathbb{N}$ is called r.e. (*recursively enumerable*) if there is some recursive $R \subseteq \mathbb{N}^2$ such that $M = \{b \in \mathbb{N} \mid (\exists a \in \mathbb{N}) Rab\}$. In short, M is the *range* of some recursive relation.

Since $a \in M \Leftrightarrow (\exists b{\in}\mathbb{N})R'ab$ where $R'ab \Leftrightarrow Rba$, M is at the same time the *domain* of some recursive relation.

It is readily shown that $M \neq \emptyset$ is r.e. if and only if $M = ran\, f$ for some recursive $f \in \mathbf{F}_1$; Exercise 4. This characterization corresponds perfectly to our intuition: stepwise computation of $f0, f1, \ldots$ provides an effective enumeration of M in the intuitive sense. This enumeration can be carried out by a computer that puts out $f0, f1, \ldots$ successively and does not stop its execution by itself.

The empty set is r.e. because it is the domain of the empty binary relation, which is recursive, and even p.r. since its characteristic function is the constant function K_0^2. In view of the above characterization of r.e. sets $M \neq \emptyset$, one could have defined these from the outset as the ranges of unary recursive functions. But the first definition has the advantage of immediately expanding to the n-dimensional case given below, and it avoids a case distinction as to whether or not M is empty.

More generally, a predicate $P \subseteq \mathbb{N}^n$ is called r.e. provided $P\vec{a} \Leftrightarrow (\exists x{\in}\mathbb{N})Q(x, \vec{a})$ for some $(n+1)$-ary recursive predicate Q. Note that a recursive predicate P is r.e. Indeed, $P\vec{a} \Leftrightarrow (\exists b{\in}\mathbb{N})P'(b, \vec{a})$; here $P'(b, \vec{a}) \Leftrightarrow P\vec{a}$ (adjoining a fictional variable). It is not quite easy to present an ad hoc example of an r.e. predicate that is not recursive. But such examples arise in a natural way in **6.5**, where we will show the undecidability of several axiomatic theories.

It is easily seen that a function $f \in \mathbf{F}_n$ is recursive provided graph f is, simply because $f\vec{a} = \mu b[\mathrm{graph}\, f(\vec{a}, b)]$ (or in strict terms of $\boldsymbol{O\mu}$, $f\vec{a} = \mu b[\delta\chi_{\mathrm{graph}\, f}(\vec{a}, b) = 0]$), that is, f can immediately be isolated from graph f with the μ-operator. Conversely, if f is recursive then so is graph f, because $\chi_{\mathrm{graph}\, f}(\vec{a}, b) = \chi_{=}(f\vec{a}, b)$. This equation also shows that graph f is p.r. whenever f is p.r. On the other hand, it is highly interesting to notice that there is a function $f \in \mathbf{F}_1$ (and not only one) whose graph is p.r. although f itself is not p.r. Much preparation is needed for getting such an example, namely the f constructed in Exercise 4 in **6.5**.

Exercises

1. Let $a \leqslant fa$ for all a. Prove that if f is p.r. (resp. recursive) then so is $ran\, f$. Show the same for $f \in \mathbf{F}_n$ whenever $a_1, \ldots, a_n \leqslant f\vec{a}$ for all $\vec{a} \in \mathbb{N}^n$.

2. Prove in detail that the pairing function $\wp\colon \mathbb{N}^2 \to \mathbb{N}$ is bijective and that its diagram in the figure on page 172 is correct.

3. Since $\wp\colon \mathbb{N}^2 \to \mathbb{N}$ is bijective, there are functions $\varkappa_1, \varkappa_2 \in \mathbf{F}_1$ such that $\wp(\varkappa_1 n, \varkappa_2 n) = n$, for all n. Prove that $\varkappa_1, \varkappa_2$ are p.r. (one need not exhibit explicit terms for these functions).

4. Let $\emptyset \neq M \subseteq \mathbb{N}$. Show M is r.e. iff $M = ran\, f$ for some recursive $f \in \mathbf{F}_1$.

6.2 Arithmetization

Roughly put, arithmetization (or Gödelization) is the description of the syntax of a formal language $\mathcal{L}$ and of formal proofs from an axiom system by means of arithmetical operations and relations on natural numbers. It presupposes the encoding of strings from the alphabet of $\mathcal{L}$ by natural numbers. Syntactical functions and predicates correspond in this way to well-defined functions and predicates on $\mathbb{N}$.

Thus many goals at once become attainable. First of all, the intuitive idea of a computable word function can be made more precise using the notion of recursive functions. Second, syntactical predicates like for instance '$x \in var\alpha$' can be replaced by corresponding predicates of $\mathbb{N}$. Third, using encoding, statements about syntactical functions, predicates and formal proofs can be formulated in theories $T \subseteq \mathcal{L}$ able to speak about arithmetic, and perhaps be proved in T.

We demonstrate the arithmetization of syntax using as an example the language $\mathcal{L} = \mathcal{L}_{ar}$ whose extralogical symbols are $0, \mathtt{S}, +, \cdot$. This is the language of Peano arithmetic PA. However, the same procedure can be carried out analogously for other formal languages, as will be apparent in the course of our considerations.

The first step is to assign uniquely to every basic symbol ζ of $\mathcal{L}$ a number $\sharp\zeta$, its *symbol code*. The following table provides an example for $\mathcal{L} = \mathcal{L}_{ar}$:

ζ	$=$	$\neg$	$\wedge$	$\forall$	(	)	0	S	+	$\cdot$	$\boldsymbol{v}_0$	$\boldsymbol{v}_1$	...
$\sharp\zeta$	1	3	5	7	9	11	13	15	17	19	21	23	

Next we encode the string $\xi = \zeta_0 \cdots \zeta_n$ by its *Gödel number*, which is the number $\langle \sharp\zeta_0, \ldots, \sharp\zeta_n \rangle = p_0^{1+\sharp\zeta_0} \cdots p_n^{1+\sharp\zeta_n}$. The empty string gets the Gödel number 1.

Example. The term 0 and the prime formula $0 = 0$ have the still comparatively small Gödel numbers $2^{1+\sharp 0} = 2^{14}$ and $2^{14} \cdot 3^2 \cdot 5^{14}$, respectively. The term $\underline{1}$ has the Gödel number $2^{16} \cdot 3^{14}$. This encoding is not particularly economical, but that need not concern us here. Nor is it a problem that the symbol code of $=$ is the same as the Gödel number of the empty string. For note that $==$, considered as a string of length 1, has the Gödel number $2^2 = 4$.

In the following, ξ, η, ϑ denote strings (or words) of the basic symbols of $\mathcal{L}$; the set of these strings is denoted by $\mathfrak{S}_{\mathcal{L}}$. Let $\dot{\xi}$ be the Gödel number of ξ, and $\dot{t}$, $\dot{\alpha}$ therefore that of the the term t and the formula α, respectively. If we write $\xi\eta$ for the concatenation of $\xi, \eta \in \mathfrak{S}_{\mathcal{L}}$, then obviously $(\xi\eta)^{\cdot} = \dot{\xi} * \dot{\eta}$, where $*$ is the arithmetical concatenation from **6.1**. $\dot{\mathfrak{S}}_{\mathcal{L}} = \{\dot{\xi} \mid \xi \in \mathfrak{S}_{\mathcal{L}}\}$ is a p.r. subset of the set of all Gödel numbers. Indeed, since $\mathcal{L}$-symbols are encoded by odd numbers,

$$n \in \dot{\mathfrak{S}}_{\mathcal{L}} \;\Leftrightarrow\; n \in GN \,\&\, (\forall k{<}\ell n)\, 2 \nmid (n)_k.$$

At least for the time being, it is necessary to distinguish between the *symbol* ζ and the *string* ζ, which actually means the single-element sequence (ζ). The Gödel

number of the *string* ζ is $2^{1+\sharp\zeta}$. For example, the prime term 0 (which is a one-letter string) has the Gödel number $\dot{0} = 2^{1+\sharp 0}$, while the symbol 0 has the symbol code 13. Similarly, we must distinguish between $\boldsymbol{v}_i$ as a term and $\boldsymbol{v}_i$ as a symbol. The term $\boldsymbol{v}_i$ and the symbol $\boldsymbol{v}_i$ are equally denoted only for faster readability.

Remark 1. One could, right from the beginning, identify symbols with their codes and strings with their Gödel numbers, so that $\dot{\varphi} = \varphi$ and $\dot{t} = t$ for formulas φ and terms t, and syntactical predicates are arithmetical from the outset. We postpone this until we have convinced ourselves that syntax can indeed adequately be encoded in arithmetic. Further, the alphabet of $\mathcal{L}_{ar}$ could easily be replaced by a finite one, consisting, say, of the symbols $=, \neg, \ldots, \cdot, \boldsymbol{v}$, in that $\boldsymbol{v}_0$ is replaced by the string $\boldsymbol{v}0$, $\boldsymbol{v}_1$ by $\boldsymbol{v}\mathsf{S}0$, and so on. Other encodings found in the literature arise from the identification of the letters in such alphabets with the digits of a suitable number base.

In the following, let $\dot{W} = \{\dot{\xi} \mid \xi \in W\}$ for sets $W \subseteq \mathfrak{S}_{\mathcal{L}}$ of words. A corresponding notation will be used for many-place word predicates P. We call P p.r. or recursive whenever $\dot{P}$ is p.r. or recursive, respectively. So, for example, if we talk about a recursive axiom system $X \subseteq \mathcal{L}$, it is always understood that $\dot{X}$ is recursive. Other properties, such as recursively enumerable or representable, can be transfered to word predicates by means of the above or a similar arithmetization.

All these remarks refer not just to $\mathcal{L} = \mathcal{L}_{ar}$, but to an arbitrary *arithmetizable* (or gödelizable) language $\mathcal{L}$, by which we simply mean that $\mathcal{L}$ possesses finitely or countably many specified basic symbols, so that each string can be assigned a number code in a computable way. In this way, the concepts of an axiomatizable or decidable theory, already used in **3.3**, obtain an absolutely precise meaning.

Of course, one must distinguish between the axioms and theorems of an axiomatic theory; the axiom systems of familiar theories like PA and ZFC are readily seen to be p.r., while these theories considered as sets of theorems are shown in **6.5** to be undecidable and cannot even be extended in any way to decidable theories.

The main goal now is the arithmetization of the formal proof method. We use $\vdash$ from now on to denote the Hilbert calculus of **3.6** consisting of the axiom system Λ with the axiom schemas $\Lambda 1$–$\Lambda 10$ given there and MP as the only rule of inference, based on some fixed arithmetizable language $\mathcal{L}$.

Just as for strings, for a finite sequence $\Phi = (\varphi_0, \ldots, \varphi_n)$ of $\mathcal{L}$-formulas we call $\dot{\Phi} := \langle \dot{\varphi}_0, \ldots, \dot{\varphi}_n \rangle$ its *Gödel number*. This includes in particular the case that Φ is a proof from X ($\subseteq \mathcal{L}$) in the sense of **3.6**, which in the general case also contains formulas from Λ. Note that $\dot{\Phi} \neq \dot{\xi}$ for all $\xi \in \mathfrak{S}_{\mathcal{L}}$, because $(\dot{\Phi})_0 = \dot{\varphi}_0$ is even, so that $2 \mid (\dot{\Phi})_0$, whereas $2 \nmid (\dot{\xi})_0$ because the symbol codes are odd. This is the case in our example language $\mathcal{L}_{ar}$ and may actually be presupposed throughout. Thus, we can comfortably distinguish the Gödel numbers of formulas and terms from the Gödel numbers of finite sequences of formulas.

Now let T ($\subseteq \mathcal{L}^0$) be a theory axiomatized by some fixed axiom system $X \subseteq T$. Examples are PA or ZFC. The language $\mathcal{L}_\in$ is obviously simpler than $\mathcal{L}_{ar}$, which of course simplifies encoding. A proof $\Phi = (\varphi_0, \ldots, \varphi_n)$ from X is also called *a proof in* T. Here and elsewhere X is tacitly understood to be an essential part of T. First define the p.r. functions $\tilde{\neg}$, $\tilde{\wedge}$, $\tilde{\to}$ as follows: $\tilde{\neg}a := \dot{\neg} * a$, $a \,\tilde{\wedge}\, b := \dot{(} * a * \dot{\wedge} * b * \dot{)}$ and $a \,\tilde{\to}\, b := \tilde{\neg}(a \,\tilde{\wedge}\, \tilde{\neg}b)$ (argument parentheses in the last expression should not be mixed up with parentheses belonging to the alphabet of $\mathcal{L}$).

Let proof_T denote the unary arithmetical predicate that corresponds to the syntactical predicate 'Φ is a proof in T from X'. We denote the arithmetical predicates corresponding to 'Φ is a proof for φ' (the last component of Φ) and to 'there is a proof for φ in T' by bew_T and bwb_T, respectively (coming from beweis=proof and beweisbar=provable). The precise definitions of these predicates look as follows:

(1) $\mathit{proof}_T(b) \;\Leftrightarrow\; b \in \mathit{GN} \;\&\; b \neq 1$
$\qquad\qquad \&\;\; (\forall k{<}\ell b)[(b)_k \in \dot{X} \cup \dot{\Lambda} \vee (\exists i, j < k)(b)_i = (b)_j \,\tilde{\to}\, (b)_k]$,

(2) $\mathit{bew}_T(b, a) \;\Leftrightarrow\; \mathit{proof}_T(b) \;\&\; a = (b)_{\mathit{last}}$, (3) $\mathit{bwb}_T\, a \;\Leftrightarrow\; \exists b\; \mathit{bew}_T(b, a)$.

Since bwb_T is a unary predicate, we may omit the argument parentheses in writing $\mathit{bwb}_T\, a$. Easily obtained from (1), (2), and (3) are

(4) $\vdash_T \alpha \;\Leftrightarrow\; \mathit{bew}_T(n, \dot{\alpha})$ for some $n \;\Leftrightarrow\; \mathit{bwb}_T\, \dot{\alpha}$,

(5) $\mathit{bew}_T(c, a) \;\&\; \mathit{bew}_T(d, a \,\tilde{\to}\, b) \Rightarrow \mathit{bew}_T(c * d * \langle b\rangle, b)$, for all a, b, c, d,

(6) $\mathit{bwb}_T\, a \;\&\; \mathit{bwb}_T(a \,\tilde{\to}\, b) \Rightarrow \mathit{bwb}_T\, b$, for all a, b,

(7) $\mathit{bwb}_T\, \dot{\alpha} \;\&\; \mathit{bwb}_T(\alpha \to \beta)^{\cdot} \Rightarrow \mathit{bwb}_T\, \dot{\beta}$, for all $\alpha, \beta \in \mathcal{L}$.

(4) is clear, for $\vdash_T \alpha$ iff there is a proof Φ for α iff $\exists n\, \mathit{bew}_T(n, \dot{\alpha})$ (choose $n = \dot{\Phi}$). (5) tells us in arithmetical language the familiar story that joining together proofs for $\alpha, \alpha \to \beta$ and tacking on β yields a proof for β. (5) immediately yields (6) by particularization, and (6) implies (7) since $(\alpha \to \beta)^{\cdot} = \dot{\alpha} \,\tilde{\to}\, \dot{\beta}$.

Remark 2. We will not need (5)–(7) until **7.1**. But it is instructive for our later transfer of proofs to PA to verify (5) first naively. This is simple when we use the following facts: for all $a, b \in \mathit{GN}$, $\ell(a * b) = \ell a + \ell b$, $(a * b)_i = a_i$ for $i < \ell a$, $(a * b)_{\ell a + i} = b_i$ for $i < \ell b$, and $\ell\langle c\rangle = 1$, $(\langle c\rangle)_0 = c$ for all $c \in \mathbb{N}$. Since it would impede the proof of (5), we did not add $(\forall k{<}b)(b)_k \in \dot{\mathcal{L}}$ to the right-hand side of (1). This is in fact not necessary, since induction on the length of the proof code b readily shows that $\mathit{proof}_T(b)$ implies $(\forall k{<}\ell b)(b)_k \in \dot{\mathcal{L}}$. Here we need $a, a \,\tilde{\to}\, b \in \dot{\mathcal{L}} \Rightarrow b \in \dot{\mathcal{L}}$, for all $a, b \in \mathbb{N}$; Exercise 2.

Now we really get down to work and show that the syntactic basic notions up to the predicate bew_T are p.r. In **6.5** basically only their recursiveness is important; not until Chapter **7** do we make essential use of their p.r. character. We return to our example $\mathcal{L} = \mathcal{L}_{ar}$, because the proofs of the following lemmas are not entirely

independent of the language's syntax and the selected encoding, though they can be proved for other arithmetizable languages in nearly the same way.

In addition to the already-defined $\tilde{\neg}$, $\tilde{\wedge}$, and $\tilde{\rightarrow}$, we define $n \mathrel{\tilde{=}} m := n * \dot{=} * m$ $(= n * 2^2 * m)$ and $\tilde{\forall}(i,n) := \dot{\forall} * i * n$. $\tilde{\exists}$ is defined similarly. Finally, for $\mathtt{S}, +, \cdot$ let $\tilde{\mathtt{S}}n = \dot{\mathtt{S}} * n$, $n \tilde{+} m = \dot{(} * n * \dot{+} * m * \dot{)}$, and similarly for $\cdot$. Then $(s = t)^{\cdot} = \dot{s} \mathrel{\tilde{=}} \dot{t}$ and $(\mathtt{S}t)^{\cdot} = \tilde{\mathtt{S}}\dot{t}$ hold, for example, as does $(\forall x\alpha)^{\cdot} = \tilde{\forall}\dot{x}\dot{\alpha}$ $(= \tilde{\forall}(\dot{x}, \dot{\alpha}))$. All these functions are obviously primitive recursive.

The set $\mathcal{V}$ of variable terms is p.r. because $n \in \dot{\mathcal{V}} \Leftrightarrow (\exists k \leqslant n)\, n = 2^{22+2k}$. Thus $\mathcal{T}_{prim} := \mathcal{V} \cup \{0\}$, the set of all prime terms of $\mathcal{L}$, is p.r. as well. For arbitrary strings ξ, η let $\xi \leqslant \eta$ mean $\dot{\xi} \leqslant \dot{\eta}$. For example, $\xi \leqslant \eta$ holds if ξ is a substring of η, in particular if ξ denotes a subformula of the formula η. This follows immediately from the property $a, b \leqslant a * b$ for Gödel numbers a, b, which was noted on page 174.

Lemma 2.1. *The set $\mathcal{T}$ of all terms is primitive recursive.*

Proof. By the recursive definition of $\mathcal{T}$, $t \in \mathcal{T}$ if and only if

$$t \in \mathcal{T}_{prim} \vee (\exists\, t_1, t_2 < t)[t_1, t_2 \in \mathcal{T} \;\&\; (t = \mathtt{S}t_1 \vee t = (t_1 + t_2) \vee t = (t_1 \cdot t_2))].$$

Therefore the corresponding arithmetical equivalence holds as well, namely

$$(*) \qquad n \in \dot{\mathcal{T}} \;\Leftrightarrow n \in \dot{\mathcal{T}}_{prim} \vee (\exists\, i, k < n)[i, k \in \dot{\mathcal{T}} \;\&\; Q(n, i, k)]$$

where $Q(n,i,k) \Leftrightarrow (n = \tilde{\mathtt{S}}i \vee n = i \tilde{+} k \vee n = i \tilde{\cdot} k)$. We now show how to convert this "informal definition" of $\dot{\mathcal{T}}$, which on the right-hand side makes use of elements of $\dot{\mathcal{T}}$ smaller than n only, into a course-of-values recursion for $\chi_{\dot{\mathcal{T}}}$, whence $\chi_{\dot{\mathcal{T}}}$, and so $\mathcal{T}$ would turn out to be p.r. Consider the p.r. predicate P defined by

$$P(a,n) \;\Leftrightarrow\; n \in \dot{\mathcal{T}}_{prim} \vee (\exists\, i, k < n)[(a)_i = (a)_k = 1 \;\&\; Q(n,i,k)].$$

We claim that the characteristic function $f := \chi_{\dot{\mathcal{T}}}$ satisfies

$$\textbf{\textit{Oq}}\text{:} \quad fn = \chi_P(\bar{f}n, n) \qquad \big(\bar{f}n = \langle f(0), \ldots, f(n-1)\rangle\big)$$

and hence is p.r. by Theorem 1.1. Indeed, since $fi = fk = 1 \Leftrightarrow i, k \in \dot{\mathcal{T}}$, we have

$$\begin{aligned} n \in \dot{\mathcal{T}} \;&\Leftrightarrow\; n \in \dot{\mathcal{T}}_{prim} \vee (\exists\, i, k < n)[fi = fk = 1 \;\&\; Q(n,i,k)] \qquad \big(\text{by } (*)\big) \\ &\Leftrightarrow\; P(\bar{f}n, n) \qquad \big(\text{because } (\bar{f}n)_i = fi \text{ and } (\bar{f}n)_k = fk\big). \end{aligned}$$

From this equivalence it clearly follows that $fn = 1 \Leftrightarrow \chi_P(\bar{f}n, n) = 1$, which in turn implies ***Oq*** since both f and χ_P take values from $\{0,1\}$ only. ❑

Lemma 2.2. *The set $\mathcal{L}$ $(= \mathcal{L}_{ar})$ of all formulas is primitive recursive.*

Proof. $\mathcal{L}_{prim}$ is p.r. because $n \in \dot{\mathcal{L}}_{prim} \;\Leftrightarrow\; (\exists\, i, k < n)[i, k \in \dot{\mathcal{T}} \;\&\; n = i \mathrel{\tilde{=}} k]$. If we consider $\dot{x} < \dot{\xi}$ for every $\xi \in \mathfrak{S}_{\mathcal{L}}$ and $x \in var\xi$ (because then $\xi = \eta x \theta$ for some strings $\eta, \theta \in \mathfrak{S}_{\mathcal{L}}$), then the predicate '$\varphi \in \mathcal{L}$' clearly satisfies the condition

$$\varphi \in \mathcal{L}_{prim} \vee (\exists\, \alpha, \beta, x < \varphi)[\alpha, \beta \in \mathcal{L} \;\&\; x \in \mathcal{V} \;\&\; (\varphi = \neg\alpha \vee \varphi = (\alpha \wedge \beta) \vee \varphi = \forall x\alpha)].$$

This "informal definition" can then be transformed just as in Lemma 2.1 into a course-of-values recursion of the characteristic function of $\dot{\mathcal{L}}$ using the characteristic function of the certainly p.r. predicate P given by

$$P(a,n) \;\Leftrightarrow\; n \in \dot{\mathcal{L}}_{prim} \vee (\exists\, i,k,j<n)[(a)_i = (a)_k = 1 \;\&\; j \in \dot{\mathcal{V}} \;\&\; (n = \tilde{\neg} i \vee n = i \;\tilde{\wedge}\; k \vee n = \tilde{\forall} jk)]. \quad \square$$

Beginning with the substitution $\xi \mapsto \xi\,\frac{t}{x}$, which is interesting both for formulas and terms, we may now define a ternary p.r. function $(m,i,k) \mapsto [m]_i^k$ so that

$$(*) \qquad [\dot{\xi}]_{\dot{x}}^{\dot{t}} = (\xi\,\tfrac{t}{x})^{\cdot} \text{ for all } \xi \in \mathcal{L} \cup \mathcal{T}.$$

For this we first translate the equations of the recursive definition for $\xi\,\frac{t}{x}$ into corresponding requirements for $[m]_i^k$. For all $m \in \dot{\mathcal{L}} \cup \dot{\mathcal{T}}$, $i \in \dot{\mathcal{V}}$ and $k \in \mathbb{N}$ let

$[m]_i^k = k$ if $i = m \in \dot{\mathcal{T}}_{prim}$, $[m]_i^k = m$ if $i \neq m \in \dot{\mathcal{T}}_{prim}$, $[\tilde{\neg} m]_i^k = \tilde{\neg}[m]_i^k$,
$[\tilde{\mathsf{S}} m]_i^k = \tilde{\mathsf{S}}[m]_i^k$, $[m \tilde{+} n]_i^k = [m]_i^k \tilde{+} [n]_i^k$ and similarly for $\cdot$, $\wedge$ and $=$,
$[\tilde{\forall} jm]_i^k = \tilde{\forall}(j,m)$ for $j = i$, $[\tilde{\forall} jm]_i^k = \tilde{\forall}(j,[m]_i^k)$ for $j \neq i$.

For all remaining triples m,i,k let $[m]_i^k = 0$. It is left to the reader to construct (using p.r. case distinction) a course-of-values recursion for the determination of $[m]_i^k$ such that the given conditions and hence $(*)$ are satisfied.

As was already noticed the predicate 'x occurs in ξ', or '$x \in var\,\xi$' for short, is p.r. since $x \in var\,\xi \;\Leftrightarrow\; x \in \mathcal{V}$ & $(\exists\,\eta,\vartheta \leqslant \xi)(\xi = \eta x \vartheta)$. Replacing here $\eta x \vartheta$ by $\eta \forall x \vartheta$ makes it clear that '$x \in bnd\,\alpha$' is p.r. as well. The binary predicate '$x \in free\,\alpha$' is also p.r. because $x \in free\,\alpha \;\Leftrightarrow\; x \in \mathcal{V}$ & $\alpha\,\frac{0}{x} \neq \alpha$ ($\Leftrightarrow x \in \mathcal{V}$ & $[\dot{\alpha}]_{\dot{x}}^{\dot{0}} \neq \dot{\alpha}$). Consequently $\mathcal{L}^0$ is p.r. With these preparations we now prove

Lemma 2.3. *The set Λ of logical axioms is primitive recursive.*

Proof. $\Lambda 1$ is p.r. because $\varphi \in \Lambda 1$ if and only if

$$(\exists\,\alpha,\beta,\gamma<\varphi)[\alpha,\beta,\gamma \in \mathcal{L} \;\&\; \varphi = (\alpha \to \beta \to \gamma) \to (\alpha \to \beta) \to (\alpha \to \gamma)].$$

To characterize the corresponding arithmetical predicate we use the p.r. function $\tilde{\to}$. One reasons similarly for $\Lambda 2$–$\Lambda 4$. For a p.r. characterization of $\Lambda 5$ use the fact that the ternary predicate '$\alpha, \frac{t}{x}$ collision-free' is p.r. For '$\alpha, \frac{t}{x}$ collision-free' holds iff $(\forall y{<}\alpha)(y \in bnd\,\alpha \;\&\; y \in var\,t \Rightarrow y = x)$. Further, the predicate '$\varphi = \forall x\alpha \to \alpha\,\frac{t}{x}$' which depends on φ,α,x,t, is p.r., as can be seen by applying $(m,i,k) \mapsto [m]_i^k$. Hence, $\Lambda 5$ is p.r. as well, because $\varphi \in \Lambda 5$ if and only if

$$(\exists\,\alpha,x,t<\varphi)(\alpha \in \mathcal{L} \;\&\; x \in \mathcal{V} \;\&\; t \in \mathcal{T} \;\&\; \varphi = \forall x\alpha \to \alpha\,\tfrac{t}{x} \;\;\&\;\; \alpha, \tfrac{t}{x} \text{ collision-free}).$$

Similarly it is shown that $\Lambda 6$–$\Lambda 10$ are p.r. Thus, each of the schemas Λi is p.r. and therefore so is $\Lambda_0 := \Lambda 1 \cup \cdots \cup \Lambda 10$. But then the same holds for Λ itself, because $k \mapsto \sharp \boldsymbol{v}_k$ is surely p.r. and every $\alpha \in \Lambda$ can be written $\alpha = \forall \vec{x}\alpha_0$ with some (possibly empty) prefix $\forall \vec{x}$ and for some $\alpha_0 \in \Lambda_0$, and then it must hold that

$$n \in \dot{\Lambda} \;\Leftrightarrow\; n \in \dot{\mathcal{L}} \;\&\; (\exists\, m, k < n)(n = m * k \;\&\; 2 \dot{|} \ell m \;\&\; k \in \dot{\Lambda}_0$$
$$\&\; (\forall i {<} \ell m)[2 \dot{|} i \;\&\; (m)_i = \sharp\forall \;\vee\; 2 \dot{\nmid} i \;\&\; (\exists k \leqslant n)(m)_i = \sharp \boldsymbol{v}_k].$$

The second line of this formula tells us that m is the Gödel number of a prefix $\forall x_1 \cdots \forall x_l$. This is a string of length $m = 2l$. ❑

All of the above holds completely analogously for every arithmetizable language. Hence, given a p.r. or recursive axiom system X, $X \cup \Lambda$ is p.r. (resp. recursive) as well. This applies in particular to the axiom systems of PA and ZFC. These are p.r. like every other common axiom system, despite the difference in their strengths. The proof is carried out in a manner fairly similar to that of Lemma 2.3.

The main result of this section that now follows, is completely independent of the strength of an axiomatic theory T. The strength of a theory T first comes into the picture when we want to prove something about bew_T and bwb_T *within* T *itself.*

Theorem 2.4. *Let X be a p.r. axiom system for a theory T of an arithmetizable language. Then the predicate bew_T is p.r. The same holds if we substitute here "recursive" for "primitive recursive." T is in either case recursively enumerable.*

Proof. Definition (2) on page 178 shows that bew_T is p.r. Because of (3) on the same page, $\dot{T} = \{a \in \dot{\mathcal{L}}^0 \mid bwb_T\, a\}$ is the range of a (primitive) recursive relation and thus is r.e. Clearly, the last part of the theorem is proved in the same manner. ❑

Theorem 2.4 can be strengthened only in particular circumstances, for example, if T is complete. Although bew_T is a (primitive) recursive predicate for each axiomatic arithmetizable theory T, bwb_T need not be recursive as, for example, in the case $T = \mathsf{Q}$. This is a famous finitely axiomatizable theory presented in the next section whose particular role for applied recursion theory was revealed in [TMR].

Exercises

1. Prove that if a theory T has a recursively enumerable axiom system X, then T also possesses a recursive axiom system (W. Craig).

2. Let $\mathcal{L} = \mathcal{L}_{ar}$. Prove (a) $a * b \in \dot{\mathfrak{S}}_{\mathcal{L}} \Leftrightarrow a, b \in \dot{\mathfrak{S}}_{\mathcal{L}}$, (b) $\tilde{\neg} a \in \dot{\mathcal{L}} \Leftrightarrow a \in \dot{\mathcal{L}}$, $a \,\tilde{\wedge}\, b \in \dot{\mathcal{L}} \Leftrightarrow a, b \in \dot{\mathcal{L}}$, and (c) $a \,\tilde{\rightarrow}\, b \in \dot{\mathcal{L}} \Leftrightarrow a, b \in \dot{\mathcal{L}}$, for all $a, b \in \mathbb{N}$.

3. Let T ($\subseteq \mathcal{L}^0_{ar}$) be axiomatizable and $\alpha \in \mathcal{L}^0_{ar}$. (a) Define a binary p.r. function f such that $bew_{T+\alpha}(\dot{\Phi}, \dot{\varphi}) \;\Rightarrow\; bew_T(f(\dot{\Phi}, \dot{\alpha}), (\alpha \rightarrow \varphi)^{\cdot})$ (arithmetization of the deduction theorem). (b) Show that $bwb_{T+\alpha}\, \dot{\varphi} \;\Leftrightarrow\; bwb_T(\alpha \rightarrow \varphi)^{\cdot}$.

4. Show that the set of quantifier-free sentences of $\mathcal{L}_{ar}$ true in $\mathcal{N}$ is p.r. That the corresponding does not hold for the set all sentences of $\mathcal{L}_{ar}$ will be shown in Section **6.5**.

6.3 Representability of Arithmetical Predicates

First of all we consider the finitely axiomatized theory Q with the axioms

Q1: $\forall x\, \mathtt{S}x \neq 0$, Q2: $\forall x \forall y(\mathtt{S}x = \mathtt{S}y \to x = y)$, Q3: $(\forall x \neq 0)\exists y\, x = \mathtt{S}y$,

Q4: $\forall x\, x + 0 = x$, Q5: $\forall x \forall y\, x + \mathtt{S}y = \mathtt{S}(x + y)$,

Q6: $\forall x\, x \cdot 0 = 0$, Q7: $\forall x \forall y\, x \cdot \mathtt{S}y = x \cdot y + x$.

These axioms characterize Q as a modest subtheory of Peano arithmetic PA. Both theories are formalized in $\mathcal{L}_{ar}$, the first-order language in $0, \mathtt{S}, +, \cdot$, and are subtheories of $Th\mathcal{N}$, where $\mathcal{N}$ as always denotes the standard model $(\mathbb{N}, 0, \mathtt{S}, +, \cdot)$. In Q, PA and related theories in $\mathcal{L}_{ar}$, let $\leqslant$ and $<$ be defined by $x \leqslant y \leftrightarrow \exists z\, z + x = y$ and $x < y \leftrightarrow x \leqslant y \wedge x \neq y$, respectively. As in **3.3**, the term $\mathtt{S}^n 0$ is denoted by $\underline{n}$.

From the results of this and the next section, not only will the recursive undecidability of Q be derived, but also that of every subtheory and every consistent extension of Q, see **6.5**. If we were interested only in undecidability results, we could simplify the proof of the representation theorem 4.2 by noting that all recursive functions can already be obtained with ***Oc*** and ***Oμ*** from the somewhat larger set of initial functions $0, \mathtt{S}, I^n_\nu, +, \cdot, \dot{-}$. But even ignoring the considerable effort required to prove the eliminability of the schema ***Op*** at the price of additional initial functions, such an approach would blur the distinction between primitive recursive and μ-recursive functions, relevant for some details in Chapter **7**.

$\forall x\, x \neq \mathtt{S}x$ is easily provable in PA by induction, but Q is too weak to allow a proof of this sentence. Its unprovability follows from the fact that $(\mathbb{N} \cup \{\infty\}, 0, \mathtt{S}, +, \cdot)$ satisfies all axioms of Q, but not $\forall x\, x \neq \mathtt{S}x$. Here ∞ is a new object and the operations $\mathtt{S}, +, \cdot$ are extended to $\mathbb{N} \cup \{\infty\}$ by putting $\mathtt{S}\infty = \infty$, $\infty \cdot 0 = 0$, and

$$\infty + n = n + \infty = \infty + \infty = n \cdot \infty = \infty \cdot m = \infty, \text{ for all } n \text{ and all } m \neq 0.$$

This model shows the unprovability in Q of many familiar laws of arithmetic, which tell us that $\mathcal{N}$ is an ordered commutative semiring with smallest element 0 and unit element $1 := \mathtt{S}0$, with the order defined as in Q above. These laws are collected in the following axiom system defining a still finitely axiomatizable theory $\mathsf{N} \subseteq \mathcal{L}_{ar}$:

N0: $x + 0 = x$	N1: $x + y = y + x$	N2: $(x + y) + z = x + (y + z)$
N3: $x \cdot 1 = x$	N4: $x \cdot y = y \cdot x$	N5: $(x \cdot y) \cdot z = x \cdot (y \cdot z)$
N6: $x \cdot (y + z) = x \cdot y + x \cdot z$	N7: $x \leqslant y \vee y \leqslant x$	N8: $x + y = x + z \to y = z$
N9: $x < y \leftrightarrow \mathtt{S}x \leqslant y$	N10: $x \leqslant 0 \to x = 0$	N11: $0 \neq 1$

$\forall$-quantifiers in the axioms are omitted. N is, like Q, a subtheory of PA, but with stricter axioms. These are all provable in PA, see Exercise 2 in **3.3**. The axioms of Q are derivable in N. For instance, $\vdash_{\mathsf{N}} \mathtt{S}x \neq 0$, since $\mathtt{S}x = 0$ implies $x < 0$ by N9; hence $x = 0$ by N10, but $\mathtt{S}0 = 0$ contradicts N11. Thus, $\mathsf{Q} \subseteq \mathsf{N} \subseteq \mathsf{PA}$.

In this section we simply write $\vdash \alpha$ for $\vdash_{\mathsf{Q}} \alpha$ and $\alpha \vdash \beta$ for $\alpha \vdash_{\mathsf{Q}} \beta$ etc. We also write occasionally $\alpha \vdash \beta \vdash \gamma$ for $\alpha \vdash \beta \;\&\; \beta \vdash \gamma$, $\vdash t_1 = t_2 = t_3$ for $\vdash t_1 = t_2 \wedge t_2 = t_3$, and $\vdash \alpha \equiv \beta$ instead of $\vdash \alpha \;\&\; \alpha \equiv \beta$, just for brevity. The use of $\vdash$ in the subtle derivations carried out below helps one see what is going on and makes the metainduction used there more vivid. Some of the proofs can be seen as "transplanting inductions from PA into the metatheory." For instance, consider $\forall x\, x \neq \mathsf{S}x$ which is provable in PA but unprovable in Q. Nontheless, we still can prove by metainduction on n that $\underline{n} \neq \mathsf{S}\underline{n}$ is provable in Q, for all n. $\vdash 0 \neq \mathsf{S}0$ is clear by Q1. The induction step $\vdash \underline{n} \neq \mathsf{S}\underline{n} \Rightarrow \vdash \mathsf{S}\underline{n} \neq \mathsf{SS}\underline{n}$ follows from $\underline{n} \neq \mathsf{S}\underline{n} \vdash \mathsf{S}\underline{n} \neq \mathsf{SS}\underline{n}$. This in turn follows from $\mathsf{S}\underline{n} = \mathsf{SS}\underline{n} \vdash \underline{n} = \mathsf{S}\underline{n}$, an application of Q2. We now shall prove

C0: $\vdash \mathsf{S}x + \underline{n} = x + \mathsf{S}\underline{n}$,

C1: $\vdash \underline{m} + \underline{n} = \underline{m+n}$, $\underline{m} \cdot \underline{n} = \underline{m \cdot n}$, C2: $\vdash \underline{n} \neq \underline{m}$ for $n \neq m$,

C3: $\vdash \underline{m} \leqslant \underline{n}$ for $m \leqslant n$, C4: $\vdash \underline{m} \not\leqslant \underline{n}$ for $m \not\leqslant n$,

C5: $x \leqslant \underline{n} \vdash x = \underline{0} \vee \cdots \vee x = \underline{n}$, C6: $\vdash x \leqslant \underline{n} \vee \underline{n} \leqslant x$.

From C5 follows $x < \underline{n} \vdash x = \underline{0} \vee \cdots \vee x = \underline{n-1}$, or $x < \underline{n} \vdash \bigvee_{i<n} x = \underline{i}$ for short, which is $\bot$ for $n = 0$. The proofs of C0–C6 will be carried out by induction (more precisely, metainduction) on n. Always remember that $0 = \underline{0}$ and $\mathsf{S}\underline{n} = \underline{\mathsf{S}n}$.

C0: Clear for $n = 0$, because $\vdash \mathsf{S}x + 0 = \mathsf{S}x = \mathsf{S}(x + 0) = x + \mathsf{S}0$ by Q4 and Q5. Our induction hypothesis is $\vdash \mathsf{S}x + \underline{n} = x + \mathsf{S}\underline{n}$. This yields, in view of axiom Q5, the induction claim $\vdash \mathsf{S}x + \mathsf{S}\underline{n} = \mathsf{S}(\mathsf{S}x + \underline{n}) = \mathsf{S}(x + \mathsf{S}\underline{n}) = x + \mathsf{SS}\underline{n}$.

C1: By Q4, $\vdash \underline{m} + 0 = \underline{m}$, and since $\underline{m} = \underline{m+0}$ we get $\vdash \underline{m} + \underline{0} = \underline{m+0}$. The induction hypothesis $\vdash \underline{m} + \underline{n} = \underline{m+n}$ yields $\vdash \underline{m} + \mathsf{S}\underline{n} = \mathsf{S}(\underline{m} + \underline{n}) = \mathsf{S}\,\underline{m+n}$, by Q5, and the last term is the same as $\underline{m + \mathsf{S}n}$. This proves the induction step. Analogously we derive $\vdash \underline{m} \cdot \underline{n} = \underline{m \cdot n}$ with Q6, Q7 and what was shown already.

C2: Clear for $n = 0$, for then $m = \mathsf{S}k$ for some k, and so $\vdash 0 \neq \underline{m}$ by Q1. Assume that $\mathsf{S}n \neq m$. By Q1, $\vdash \underline{\mathsf{S}n} \neq \underline{m}$ in case $m = 0$. Otherwise $m = \mathsf{S}k$ for some k, so that $n \neq k$, hence $\vdash \underline{n} \neq \underline{k}$ by the induction hypothesis. Thus, $\vdash \underline{\mathsf{S}n} \neq \underline{m}$ by Q2.

C3: $m \leqslant n$ implies $k + m = n$ for some k, hence $\underline{k+m} = \underline{n}$. Thus, $\vdash \underline{k} + \underline{m} = \underline{n}$ by C1. Therefore $\vdash \exists z\, z + \underline{m} = \underline{n}$, which just means $\vdash \underline{m} \leqslant \underline{n}$.

C4: $m \not\leqslant n \Rightarrow m \neq 0$, hence $m = \mathsf{S}k$, some k. Let $m \not\leqslant 0$. Then $\vdash \underline{m} \not\leqslant 0$ because $\underline{m} \leqslant 0 \vdash \mathsf{S}\underline{k} \leqslant 0 \vdash \exists v\, v + \mathsf{S}\underline{k} = 0 \vdash \exists v\, \mathsf{S}(v + \underline{k}) = 0 \vdash \bot$ by Q1. Now let $m \not\leqslant \mathsf{S}n$. Then $k \not\leqslant n$ and so $\vdash \underline{k} \not\leqslant \underline{n}$ by the induction hypothesis, which yields $\vdash \underline{m} \not\leqslant \mathsf{S}\underline{n}$ by Q2.

C5: Clear for $n = 0$, because $x \neq 0, x \leqslant 0 \vdash \exists v \mathsf{S}v = 0 \vdash \bot$ by Q3, Q5, Q1. The induction claim is equivalent to $x \neq 0, x \leqslant \underline{\mathsf{S}n} \vdash \bigvee_{i=1}^{n+1} y = \underline{i}$. It is derived as follows:

$$\begin{aligned} x \neq 0, x \leqslant \underline{\mathsf{S}n} \;&\vdash \exists y (x = \mathsf{S}y \wedge y \leqslant \underline{n}) && \text{(Q3, Q5, and Q2)} \\ &\vdash \exists y (x = \mathsf{S}y \wedge \textstyle\bigvee_{i \leqslant n} y = \underline{i}) && \text{(induction hypothesis)} \\ &\vdash \exists y (x = \mathsf{S}y \wedge \textstyle\bigvee_{i=1}^{n+1} \mathsf{S}y = \underline{i}) \equiv_{\mathsf{Q}} \textstyle\bigvee_{i=1}^{n+1} x = \underline{i}. \end{aligned}$$

C6: Clear for $n = 0$ since $\vdash 0 \leqslant x$. Further, $\underline{n} < x \vdash \exists y \mathtt{S}y + \underline{n} = x \vdash \exists y y + \mathtt{S}\underline{n} = x$, by Q3 and C0, provided one has first shown $\vdash 0 + \underline{n} = \underline{n}$ by induction on n. Thus, $\underline{n} < x \vdash \underline{\mathtt{S}n} \leqslant x$. C5, C3 leads to $x \leqslant \underline{n} \vdash x \leqslant \underline{\mathtt{S}n}$. This and the former yield the inductive step, because $x \leqslant \underline{n} \vee \underline{n} \leqslant x \ \vdash \ x \leqslant \underline{n} \vee \underline{n} < x \ \vdash \ x \leqslant \underline{\mathtt{S}n} \vee \underline{\mathtt{S}n} \leqslant x$.

With these preparations we now give the following crucial definition, in which $T \supseteq \mathsf{Q}$ is supposed for simplicity's sake. This will cover all our applications.

Definition. $P \subseteq \mathbb{N}^n$ is called *numeralwise representable*[2] or simply *representable* in $T \supseteq \mathsf{Q}$ if there is some $\alpha = \alpha(\vec{x})$ (a *representing formula*) such that

$$\mathrm{R}^+\colon \ P\vec{a} \ \Rightarrow \vdash_T \alpha(\underline{\vec{a}}) \ ; \quad \mathrm{R}^-\colon \ \neg P\vec{a} \ \Rightarrow \vdash_T \neg\alpha(\underline{\vec{a}}).$$

Examples. The identity relation $\{(a, a) \mid a \in \mathbb{N}\}$ is represented by $x = y$, because $\vdash \underline{a} = \underline{b}$ is trivial if $a = b$, and $\vdash \underline{a} \neq \underline{b}$ is derivable for $a \neq b$ by C2. By C3 and C4 the formula $x \leqslant y$ represents the $\leqslant$-predicate ("in Q" is often omitted). $x \neq x$ represents the empty set, represented as well by each sentence α with $\neg\alpha \in \mathsf{Q}$.

For consistent $T \supseteq \mathsf{Q}$, whenever R^+, R^- are valid then so too are their converses, so that in fact $P\vec{a} \ \Leftrightarrow \ \vdash_T \alpha(\underline{\vec{a}})$ and $\neg P\vec{a} \ \Leftrightarrow \ \vdash_T \neg\alpha(\underline{\vec{a}})$. Note that a $P \subseteq \mathbb{N}^n$, represented by $\alpha(\vec{x})$, is recursive by Church's thesis: simply turn on the enumeration machine for Q and wait until $\alpha(\vec{a})$ or $\neg\alpha(\vec{a})$ appears. The set of n-ary representable predicates is closed under union, intersection, and complement, as well as swapping, equating, and adjoining fictional arguments. If P, Q are represented respectively by $\alpha(\vec{x}), \beta(\vec{x})$, then so too are $P \cap Q$ by $\alpha(\vec{x}) \wedge \beta(\vec{x})$ and $\neg P$ by $\neg\alpha(\vec{x})$, etc.

A predicate P represented in Q by α is clearly representable by the same α in any consistent extension of Q, in particular in $Th\mathcal{N}$. But this just means definability of P in $\mathcal{N}$ by α in the sense of **2.3**, because $\mathcal{N} \vDash \alpha\ [\vec{a}]$ is equivalent to $\mathcal{N} \vDash \alpha(\underline{\vec{a}})$. In short, definability of P in $\mathcal{N}$ and representability of P in $Th\mathcal{N}$ coincide. In the main, however, we consider representability in Q to obtain some strong results needed in **6.5**. We always have to look carefully at the representing formulas.

One could define $f \in \mathbf{F}_n$ to be representable if graph f is representable. However, it turns out that this definition is equivalent to a stronger notion of representability for functions that will be introduced after some additional preparation.

Predicates and functions definable in $\mathcal{N}$, that is, by $0, \mathtt{S}.+, \cdot$, are called *arithmetical* after [Go2]. From now on this word will always have this meaning. The arithmetical predicates encompass the representable ones. In order to discover more about these objects we consider their defining formulas more closely. Prime formulas in $\mathcal{L}_{ar}$ are equations, also called *Diophantine equations*. If $\delta(\vec{x}, \vec{y})$ is such an equation and $P\vec{a} \Leftrightarrow \mathcal{N} \vDash \exists\vec{y}\delta(\underline{\vec{a}}, \vec{y})$, then P is called *Diophantine*. A simple example is $\leqslant$, because

[2] In [Go2] representable predicates are called *entscheidungsdefinit*, in [HB] *vertretbar*, in [Kl1] *numeralwise expressible*, in [TMR] *definable*, in [Hej] *decidable*, and in [En] *representable*.

$a \leqslant b \Leftrightarrow \exists y\, y + a = b$ (this notation is an informal and faster legible substitute for the lengthy $a \leqslant b \Leftrightarrow \mathcal{N} \vDash \exists y\, y + \underline{a} = \underline{b}$). In fact, all predicates definable in $\mathcal{N}$ by $\exists$-formulas $\exists \vec{y} \varphi$ from $\mathcal{L}_{ar}$ with kernel φ are Diophantine. The proof is not difficult: Think of φ as being constructed from literals by means of $\wedge, \vee$, and use the following equivalences in an inductive proof on φ of what has been claimed:

$$\begin{aligned} s \neq t &\equiv_{\mathcal{N}} \exists z(\mathsf{S}z + s = t \vee \mathsf{S}z + t = s), \\ s_1 = t_1 \vee s_2 = t_2 &\equiv_{\mathcal{N}} s_1 s_2 + t_1 t_2 = s_1 t_2 + s_2 t_1, \\ s_1 = t_1 \wedge s_2 = t_2 &\equiv_{\mathcal{N}} s_1^2 + t_1^2 + s_2^2 + t_2^2 = \underline{2}(s_1 t_1 + s_2 t_2). \end{aligned}$$

A classification of arithmetical formulas and predicates helpful not only for the sake of representability is given by the following definition, to be generalized in **6.7**:

Definition. A formula is called Δ_0 or a Δ_0-*formula* if it is generated from prime formulas of $\mathcal{L}_{ar}$ by $\wedge, \neg$, and *bounded quantification*, i.e., if α is a Δ_0-formula then so is $(\forall x \leqslant t)\alpha$ $\big(:= \forall x(x \leqslant t \to \alpha)\big)$; here t is any $\mathcal{L}_{ar}$-term with $x \notin var\, t$. Let φ be Δ_0. Then every formula of the form $\exists \vec{x} \varphi$ is called a Σ_1-*formula* while $\forall \vec{x} \varphi$ is said to be a Π_1-*formula*. Further: $P \subseteq \mathbb{N}^n$ is said to be Δ_0, Σ_1, or Π_1 whenever P is defined in $\mathcal{N}$ by a Δ_0-formula, Σ_1-formula, or Π_1-formula, respectively. Δ_0, Σ_1, and Π_1 denote the sets of Δ_0-, Σ_1- and Π_1-predicates. In addition, $\Delta_1 := \Sigma_1 \cap \Pi_1$.

We will call a formula Δ_0, Σ_1 or Π_1 also if it is equivalent to one of the above. In this sense, for instance, if α is Δ_0 then so too are $(\exists x \leqslant t)\, \alpha$ $\big(\equiv \neg(\forall x \leqslant t) \neg \alpha\big)$ and $(\forall x < t)\alpha$ $\big(\equiv (\forall x \leqslant t)(x = t \vee \alpha)\big)$. Note that Δ_1 consists of the predicates, that are both Σ_1- and Π_1-definable, with possibly distinct formulas. Obviously, Π_1 consists of the complements of the $P \in \Sigma_1$. There are no Δ_1-formulas; there is no meaningful definition of such formulas as we will see. By Exercise 3 in **2.4**, Σ_1 and Π_1 are closed under union and intersection of predicates of the same arity, and Δ_1 moreover under complements, as is Δ_0. Note that if $P \in \mathbb{N}^m$ and $g_1, \ldots, g_m \in \mathbf{F}_n$ are Σ_1 so too is $Q = P[g_1, \ldots, g_m]$, because $Q\vec{a} \Leftrightarrow \exists \vec{y}(\bigwedge_{i=1}^n y_i = g_i \vec{a}\ \ \&\ P\vec{y})$.

Examples. Diophantine equations are the simplest Δ_0-formulas. To these belong the formulas $y = t(\vec{x})$ with $y \notin var\, t$, which define the term functions $\vec{a} \mapsto t^{\mathcal{N}}(\vec{a})$. Since $a \,|\, b \Leftrightarrow (\exists c \leqslant b)(a \cdot c = b)$, divisibility and thus also the predicate prim are Δ_0. Because $\wp(a, b) = c \Leftrightarrow 2c = (a + b)^2 + 3a + b$, the graph of the pairing function $\wp$ is Δ_0. The same holds for the relation of two numbers being *coprime*, denoted by $\perp$ and defined by $a \perp b :\Leftrightarrow (\forall c \leqslant a + b)(c \,|\, a, b \Rightarrow c = 1)$, that is, a, b have no common prime factor. Diophantine predicates are trivially Σ_1. Surprisingly, by Theorem 5.6 the converse holds as well, although it had been conjectured for some time that the set $P_2 := \{a \in \mathbb{N} \mid (\forall p \leqslant a)(\mathsf{prim}\, p\ \&\ p \,|\, a \Rightarrow p = 2)\}$ of all powers of 2 was not Diophantine. P_2 is obviously Δ_0. Note that this does not yet mean that the graph of $n \mapsto 2^n$ is Δ_0, although the latter is in fact the case; see Remark 1.

Remark 1. More generally, the predicate '$a^b = c$' is Δ_0, though it is difficult to prove this fact. Indeed, even the proof in **6.4** that this predicate is arithmetical requires effort. Earlier results from Bennet, Paris, Pudlak, among others, are generalized in [BD] as follows: if $f \in \mathbf{F}_{n+1}$ (more precisely, graph f) is Δ_0 then so is $g\colon (\vec{a}, n) \mapsto \prod_{i\leqslant n} f(\vec{a}, i)$, and the recursion equation $g(\vec{x}, \mathsf{S}y) = g(\vec{x}, y)\cdot f(\vec{x}, y)$ is provable in $I\Delta_0$. This theory is an important weakening of PA. It results from N by adjoining the induction schema restricted to Δ_0-formulas. $I\Delta_0$ plays a role in various questions, e.g., in complexity theory ([Kr]). Induction on the Δ_0-formulas readily shows that all Δ_0-predicates are p.r. The converse does not hold; an example is the graph of the very rapidly growing *hyperexponentiation*, defined by $\mathrm{hex}(a, 0) = 1$ and $\mathrm{hex}(a, Sb) = a^{\mathrm{hex}(a,b)}$. Stated more suggestively, $\mathrm{hex}(a, n) = \underbrace{a^{a^{\cdot^{\cdot^{a}}}}}_{n}$.

A model-theoretical glance at Q facilitates a quick proof of the following interesting theorem. It claims that even the seemingly weak theory Q is Σ_1-*complete*. This result is significantly strengthened for $T = \mathsf{PA}$ in **7.1**, where it is shown that the Σ_1-completeness of PA is *provable within* PA. If stated as "$\vdash_{\mathsf{Q}} \alpha$ or $\vdash_{\mathsf{Q}} \neg\alpha$, for Δ_0-sentences α" Theorem 3.1 could also be shown with proof-theoretical means. The reader may try to prove this on his own, to compare the proof-theoretic approach and the model-theoretic approach chosen here. C1 and C2 guarantee that $n \mapsto \underline{n}^{\mathcal{A}}$ provides an embedding of $\mathcal{N}$ in any model $\mathcal{A}$ of Q. Thus, $\mathcal{N}$ is a prime model of Q in the sense of **5.1**, so that w.l.o.g. $\mathcal{N} \subseteq \mathcal{A}$. Moreover, by C5, $\mathcal{A}$ is an *end extension* of $\mathcal{N}$, which is to mean that the elements of $A\setminus\mathbb{N}$ are located "at the end" of A; more precisely, $a \leqslant^{\mathcal{A}} b$ and $b \in \mathbb{N}$ imply $a \in \mathbb{N}$, for all $a \in A$.

Theorem 3.1 (on the Σ_1-completeness of Q). *Every Σ_1-sentence true in $\mathcal{N}$ is already provable in Q and hence in each extension $T \supseteq \mathsf{Q}$.*

Proof. It is enough to show for an arbitrary $\mathcal{A} \vDash \mathsf{Q}$ with $\mathcal{N} \subseteq \mathcal{A}$,

$$(*) \quad \mathcal{N} \vDash \alpha \;\Leftrightarrow\; \mathcal{A} \vDash \alpha, \text{ for all } \Delta_0\text{-sentences } \alpha.$$

Indeed, let $\mathcal{N} \vDash \exists\vec{x}\varphi(\vec{x})$ where $\varphi(\vec{x})$ is Δ_0 and $\mathcal{N} \vDash \alpha := \varphi(\vec{\underline{a}})$, say. Then, by $(*)$, $\mathcal{A} \vDash \alpha$ for each $\mathcal{A} \vDash \mathsf{Q}$. Thus, $\vdash_{\mathsf{Q}} \alpha$ and hence $\vdash_{\mathsf{Q}} \exists\vec{x}\varphi(\vec{x})$. Clearly, $(*)$ holds for all prime sentences α. The induction steps for $\wedge, \neg$ are obvious. It remains to verify the step for bounded quantification. Let $\mathcal{N} \vDash (\forall x\leqslant t)\beta(x) \in \mathcal{L}^0_{ar}$ where $\beta(x)$ is Δ_0 and (necessarily) $var\, t = \emptyset$, so that $\binom{*}{*}\colon\ a \leqslant^{\mathcal{N}} t^{\mathcal{N}} \Rightarrow \mathcal{N} \vDash \beta(\underline{a})$, for all $a \in \mathbb{N}$. To prove $\mathcal{A} \vDash (\forall x\leqslant t)\beta(x)$, let $w\colon Var \to A$, $a := x^w$, and $a \leqslant^{\mathcal{A}} t^{\mathcal{A}}$. Clearly $t^{\mathcal{A}} = t^{\mathcal{N}} \in \mathbb{N}$. Since $\mathcal{A}$ is an end extension of $\mathcal{N}$, we get $a \in \mathbb{N}$. Hence, $\mathcal{N} \vDash \beta(\underline{a})$ by $\binom{*}{*}$, and so $\mathcal{A} \vDash \beta(\underline{a})$ by the induction hypothesis. This proves $\mathcal{A} \vDash (\forall x\leqslant t)\beta$ and hence the direction $\Rightarrow$ of our induction step. The converse is obvious since $\mathcal{N} \subseteq \mathcal{A}$. ❑

If $\varphi(\vec{x})$ is Δ_0 then $\mathcal{N} \vDash \varphi(\vec{\underline{a}}) \;\Rightarrow \vdash_{\mathsf{Q}} \varphi(\vec{\underline{a}})$ and $\mathcal{N} \vDash \neg\varphi(\vec{\underline{a}}) \;\Rightarrow \vdash_{\mathsf{Q}} \neg\varphi(\vec{\underline{a}})$ by the theorem, because both $\varphi(\vec{\underline{a}})$ and $\neg\varphi(\vec{\underline{a}})$ are trivially Σ_1. Thus, we obtain

Corollary 3.2. *A Δ_0-formula represents in Q the predicate that it defines in $\mathcal{N}$.*

Lemma 3.3. *Let $P \subseteq \mathbb{N}^{n+1}$ be represented by $\alpha(\vec{x}, y)$. Then both $(\exists z{<}y)\alpha(\vec{x}, z)$ and $(\forall z{<}y)\alpha(\vec{x}, z)$ represent the predicates Q and R, where*

$$Q(\vec{a}, b) :\Leftrightarrow (\exists c{<}b)P(\vec{a}, c) \text{ and } R(\vec{a}, b) :\Leftrightarrow (\forall c{<}b)P(\vec{a}, c).$$

Proof. R^+: Suppose $Q(\vec{a}, b)$, hence $P(\vec{a}, c)$ for some $c < b$. Then $\vdash \underline{c} < \underline{b} \wedge \alpha(\underline{\vec{a}}, \underline{c})$. Consequently, $\vdash (\exists z{<}\underline{b})\alpha(\underline{\vec{a}}, z)$. To prove R^- suppose $\neg Q(\vec{a}, b)$, hence $\neg P(\vec{a}, i)$ for all $i < b$. Thus, $\bigvee_{i<b} z = \underline{i} \vdash \neg\alpha(\underline{\vec{a}}, z)$. By C5 we have $z < \underline{b} \vdash \bigvee_{i<b} z = \underline{i}$ and so $z < \underline{b} \vdash \neg\alpha(\underline{a}, z)$. Therefore, $\vdash (\forall z{<}\underline{b})\neg\alpha(\underline{\vec{a}}, z) \equiv \neg(\exists z{<}\underline{b})\alpha(\underline{\vec{a}}, z)$. This proves R^-. For the predicate R it is enough to notice that $R(\vec{a}, b) \Leftrightarrow \neg(\exists c{<}b)\neg P(\vec{a}, c)$. ❏

Since $(\exists z{\leqslant}y)\alpha \equiv (\exists z{<}y)\alpha \vee \alpha\frac{z}{y}$, the lemma shows that, for representable P, the predicates defined by $(\exists c{\leqslant}b)P(\vec{a}, c)$ and $(\forall c{\leqslant}b)P(\vec{a}, c)$ are representable as well.

Following [Go2] and [TMR], we now define the notion of a representable function. Although representability of f is much stronger a notion than representability of graph f, Lemma 3.4(b) will show that both properties coincide.

Definition. $f \in \mathbf{F}_n$ is *representable* in T (if "in T" is omitted we always mean $T = \mathsf{Q}$ and write $\vdash$ for $\vdash_{\mathsf{Q}}$) if there is a formula $\varphi(\vec{x}, y)$ such that for all $\vec{a} \in \mathbb{N}^n$,

$$\mathrm{R}^+: \quad \vdash_T \varphi(\underline{\vec{a}}, \underline{f\vec{a}}), \qquad \mathrm{R}^=: \quad \varphi(\underline{\vec{a}}, y) \vdash_T y = \underline{f\vec{a}}.$$

If φ is Δ_0 (resp. Σ_1 or Π_1) then f is said to be Δ_0- (resp. Σ_1- or Π_1-) *representable.* A similar phrase is used for predicates. In particular, $P \subseteq \mathbb{N}^n$ is Δ_1-representable if P is both Σ_1- *and* Π_1-representable.

Since $\mathrm{R}^=$ is equivalent to $\vdash_T \varphi(\underline{\vec{a}}, y) \to y = \underline{f\vec{a}}$, it is easily seen that R^+ and $\mathrm{R}^=$ together are replaceable by the single condition $y = \underline{f\vec{a}} \equiv_T \varphi(\underline{\vec{a}}, y)$ for all $\vec{a}$. If f is represented by $\varphi(\vec{x}, y)$ then graph f is represented by the same formula, because if $b \neq f\vec{a}$ and so $\vdash \underline{b} \neq \underline{f\vec{a}}$ by C2, then $\vdash \neg\varphi(\underline{\vec{a}}, \underline{b})$ by $\mathrm{R}^=$, so that R^- holds.

Lemma 3.4. (a) *Let $P \subseteq \mathbb{N}^{n+1}$ be represented by $\alpha(\vec{x}, y)$ and suppose $\forall\vec{a}\,\exists b P(\vec{a}, b)$. Then $\varphi(\vec{x}, y) := \alpha(\vec{x}, y) \wedge (\forall z{<}y)\neg\alpha(\vec{x}, z)$ represents $f : \vec{a} \mapsto \mu b[P(\vec{a}, b)]$. If P is Δ_0-representable (that is, represented by some Δ_0-formula) then so is f. If P is Δ_1-representable then f is Σ_1-representable.* (b) *f is representable provided* graph f *is representable.* (c) *If f is Σ_1-representable then f is Π_1-representable as well.* (d) *If χ_P is Σ_1-representable then P is Δ_1-representable.*

Proof. By Lemma 3.3, $\varphi(\vec{x}, y)$ represents the predicate defined by $\varphi(\vec{x}, y)$ and this is clearly graph f. Hence, R^+ holds. To verify $\mathrm{R}^=$ it has to be shown that

$$(*) \quad \alpha(\underline{\vec{a}}, y) \wedge (\forall z{<}y)\neg\alpha(\underline{\vec{a}}, z) \vdash y = \underline{f\vec{a}}.$$

Suppose $b := f\vec{a}$. Then $\underline{b} < y \vdash (\exists z{<}y)\alpha(\underline{\vec{a}}, z)$, because $\vdash \alpha(\underline{\vec{a}}, \underline{b})$. Contraposition yields $(\forall z{<}y)\neg\alpha(\underline{\vec{a}}, z) \vdash \underline{b} \not< y$. By C5 and R^- we have $y < \underline{b} \vdash \bigvee_{i<b} y = \underline{i} \vdash \neg\alpha(\underline{\vec{a}}, y)$. Hence $\alpha(\underline{\vec{a}}, y) \vdash y \not< \underline{b}$. So $\alpha(\underline{\vec{a}}, y) \wedge (\forall z{<}y)\neg\alpha(\underline{\vec{a}}, y) \vdash y \not< \underline{b} \wedge \underline{b} \not< y \vdash y = \underline{b}$ by C6.

This proves (∗). Clearly, φ in (a) is Δ_0 if α is Δ_0. Let P be represented at the same time by the Π_1-formula β. Repeating the above with $\alpha(\vec{x},y)\wedge(\forall v{<}y)\neg\beta(\vec{x},v)$ (a Σ_1-formula by Exercise 2) in place of φ, yields the additional claim. (b) follows from applying (a) to $P = \operatorname{graph} f$ while noting that $f\vec{a} = \mu b[P(\vec{a},b)]$. (c): Let the Σ_1-formula $\varphi(\vec{x},y)$ represent f and $z \notin \mathit{var}\,\varphi$. Then $\varphi'(\vec{x},y) := \forall z(\varphi(\vec{x},z)\to z{=}y)$ is a Π_1-formula that represents f as well: Application of $\mathrm{R}^{=}$ results in $\vdash \varphi'(\underline{\vec{a}},\underline{f\vec{a}})$ which confirms R^{+} for φ', and because of $\vdash \varphi(\underline{\vec{a}},\underline{f\vec{a}})$, we obtain $\mathrm{R}^{=}$ for φ' from

$$\varphi'(\underline{\vec{a}},y) = \forall z(\varphi(\underline{\vec{a}},z)\to y{=}z) \;\vdash\; \varphi(\underline{\vec{a}},\underline{f\vec{a}})\to y{=}\underline{f\vec{a}} \;\vdash\; y{=}\underline{f\vec{a}}.$$

(d): Let χ_P be Σ_1-represented by $\varphi(\vec{x},y)$. Then P is clearly Σ_1-represented by $\varphi(\vec{x},\underline{1})$ and Π_1-represented by $\neg\varphi(\vec{x},0)$. ❏

Remark 2. graph $\wp$ is represented in Q by $\alpha(x,y,z) = z{\cdot}\underline{2}{=}(x{+}y){\cdot}\mathsf{S}(x{+}y){+}x{\cdot}\underline{2}$. Thus, by Lemma 3.4(a), $\wp$ is represented by the Δ_0-formula $\alpha(x,y,z)\wedge(\forall u{<}z)\neg\alpha(x,y,u)$. We mention that in PA (but not in Q) even the quantifier-free α represents the function $\wp$.

Lemma 3.5. (a) *Let $P\subseteq\mathbb{N}^k$ be represented by $\alpha(\vec{y})$, and $g_i\in\mathbf{F}_n$ represented by γ_i for $i=1,\dots,k$. Then $\beta(\vec{x}) := \exists\vec{y}\,[\bigwedge_i\gamma_i(\vec{x},y_i)\wedge\alpha(\vec{y})]$ represents the predicate $Q := P[g_1,\dots,g_k]$. If the γ_i are Σ_1 and P is Δ_1-representable then so is Q.* (b) *If $h\in\mathbf{F}_m$ and $g_1,\dots,g_m\in\mathbf{F}_n$ are representable then so is $f = h[g_1,\dots,g_m]$.*

Proof. Let $b_i := g_i\vec{a}$, so that $\vdash\gamma_i(\underline{\vec{a}},\underline{b_i})$ for $i=1,\dots,k$ and let $\vec{b}=(b_1,\dots,b_k)$. If $Q\vec{a}$ holds, hence $P\vec{b}$, then $\vdash\alpha(\underline{\vec{b}})$, whence $\vdash\bigwedge_i\gamma_i(\underline{\vec{a}},\underline{b_i})\wedge\alpha(\underline{\vec{b}})$, and so $\vdash\beta(\underline{\vec{a}})$. But if $\neg Q\vec{a}$ and thus $\neg P\vec{b}$, then clearly $\vdash\neg\alpha(\underline{\vec{b}})$. Using $\mathrm{R}^{=}$ for the γ_i, this then yields $\bigwedge_i\gamma_i(\underline{\vec{a}},y_i)\vdash\bigwedge_i\; y_i{=}\underline{b_i}\vdash\neg\alpha(\vec{y})$. Hence $\vdash\forall\vec{y}\,[\bigwedge_i\gamma_i(\underline{\vec{a}},y_i)\to\neg\alpha(\vec{y})]\equiv\neg\beta(\underline{\vec{a}})$. If the γ_i and also α are Σ_1, then so too is β. If P is represented by the Π_1-formula $\alpha'(\vec{x})$ at the same time, then Q is represented by the Π_1-formula $\forall\vec{y}\,[\bigwedge_i\gamma_i(\vec{x},y_i)\to\alpha'(\vec{y})]$, as is easily seen. (b) results without difficulty from (a) applied to graph h. ❏

Exercises

1. Suppose α is a Δ_0-formula so that $\exists\vec{x}\alpha$ is Σ_1 and $\forall\vec{x}\alpha$ is Π_1. Construct Δ_0-formulas β and γ such that $\exists\vec{x}\alpha\equiv_{\mathcal{N}}\exists x\beta$ and $\forall\vec{x}\alpha\equiv_{\mathcal{N}}\forall x\gamma$ (*quantifier compression*). Each Δ_0-predicate is p.r. (Remark 1). Hence, each Σ_1-predicate P is r.e. and w.l.o.g. of the form $(\exists b{\in}\mathbb{N})Q(\vec{a},b)$ with $Q\in\Delta_0$.

2. Show that Σ_1 is closed under bounded quantification, that is, if $\alpha=\alpha(\vec{x},y)$ defines some Σ_1-predicate, then so do $(\forall z{<}y)\alpha\,{}^{z}\!/_{y}$ and $(\exists z{<}y)\alpha\,{}^{z}\!/_{y}$. The analogue holds for Π_1 and hence also for Δ_1.

3. Prove that $\alpha(\vec{x})\wedge y{=}\underline{1}\vee\neg\alpha(\vec{x})\wedge y{=}\underline{0}$ represents χ_P provided α represents P.

4. Show that every Δ_0-formula is equivalent to a formula constructed from literals by means of $\wedge,\vee$, and the bounded quantifiers $(\forall x{\leqslant}t)$ and $(\exists x{\leqslant}t)$.

6.4 The Representability Theorem

For the representability of all recursive or just all p.r. functions, it is helpful to have a representable function $g \in \mathbf{F}_2$ that satisfies the following: for every n and every sequence $c_0, \ldots, c_n$ there exists a number c such that $(*)\colon\ g(c,i) = c_i$ for all $i \leqslant n$. In short, c can be chosen such that the values $g(c,0), g(c,1), \ldots, g(c,n)$ are the given ones. Now, there are many p.r. functions g that can do this. For instance, if $g\colon (c,i) \mapsto (c)_i$ then $(*)$ holds with $c = p_0^{1+c_0} \cdots p_n^{1+c_n}$. Initially there is no obvious way to show the representability of such a function g in Q or in some extension of Q within the frame of the language $\mathcal{L}_{ar}$. Therefore, K. Gödel, who around 1930 was working on this and related problems, in the words of A. Mostowski "phoned with God." Although nowadays several possibilities are known, we follow the original, which has not lost any of its attraction.

Let $\boldsymbol{\alpha}(a,b,i) := \mathrm{rem}(a : (1 + (1+i)b))$, where $\mathrm{rem}(c : d)$ denotes the remainder of c divided by d $(\neq 0)$ and $\mathrm{rem}(c : 0) := 0$. Note that $\mathrm{rem}(c : d)$ is well defined since for c, d $(\neq 0)$ there are unique $q, r \in \mathbb{N}$ with $r < d$ such that $c = qd + r$ (this can readily be shown by induction on c). Clearly, graph $\boldsymbol{\alpha}$ has the Δ_0-definition

$$\boldsymbol{\alpha}(a,b,i) = k \ \Leftrightarrow\ (\exists c \leqslant a)[a = c(1 + (1+i)b) + k \ \&\ k < 1 + (1+i)b].$$

Hence, the function $\boldsymbol{\alpha}$ is Δ_0-representable by Lemma 3.4(a). The same holds for the pairing function $\wp$. Because $\wp$ is bijective there are unary functions $\varkappa_1, \varkappa_2$ such that $\wp(\varkappa_1 k, \varkappa_2 k) = k$ for all k. Their explicit form is insignificant; we just require the obvious property $\varkappa_1 k, \varkappa_2 k \leqslant k$. The function $\boldsymbol{\beta}\colon (c,i) \mapsto \boldsymbol{\alpha}(\varkappa_1 c, \varkappa_2 c, i)$ is called the *$\boldsymbol{\beta}$-function*. Since $\boldsymbol{\beta}(c,i) = k \ \Leftrightarrow\ (\exists a \leqslant c)(\exists b \leqslant c)[\wp(a,b) = c \ \&\ \boldsymbol{\alpha}(a,b,i) = k]$, graph $\boldsymbol{\beta}$ is Δ_0. Hence, by Lemma 3.4, $\boldsymbol{\beta}$ is represented by a Δ_0-formula, which is denoted by `beta`. Omitting the argument parentheses in `beta`, this means that

(1) $\vdash_{\mathsf{Q}} \mathtt{beta}\,\underline{c}\,\underline{i}\,y \leftrightarrow y = \underline{\boldsymbol{\beta}(c,i)}$, for all $c, i \in \mathbb{N}$.

The following simple number-theoretical facts known for ages will be applied in proving the property of the $\boldsymbol{\beta}$-function stated in Lemma **4.1** below.

Euclid's lemma. *Let a, b be positive and coprime $(a \perp b)$. Then there exist $x, y \in \mathbb{N}$ such that $xa + 1 = yb$.* (The converse is obvious: $c \mid a, b \Rightarrow c \mid yb - xa = 1 \Rightarrow c = 1$.)

Proof by $<$-induction on $s = a + b$. Trivial for $s \leqslant 2$, i.e., $a = b = 1$. Let $s > 2$. Then $a \neq b$, say $a > b$, and $a - b \perp b$ as well $(p \mid a - b, b \ \Rightarrow\ p \mid a - b + b = a)$. Since $(a-b)+b < s$, there are $x, y \in \mathbb{N}$ with $x(a-b)+1 = yb$ by the induction hypothesis. Hence, $xa + 1 = y'b$ with $y' = x + y$. The case $a < b$ is treated similarly. ❑

Chinese remainder theorem. *Let $c_i < d_i$ for $i = 0, \ldots, n$ and let $d_0, \ldots, d_n$ be pairwise coprime. Then there exists some $a \in \mathbb{N}$ such that* $\mathrm{rem}(a : d_i) = c_i$ *for $i = 0, \ldots, n$.*

Proof by induction on n. For $n = 0$ this is clear putting $a = c_0$. Now suppose the assumptions hold for $n > 0$. By the induction hypothesis, $\mathrm{rem}(a : d_i) = c_i$ for some a and all $i < n$. Further, $k := \mathrm{lcm}\{d_\nu \mid \nu{<}n\}$ and d_n are coprime (Exercise 1). Thus, by Euclid's lemma, there are numbers $x, y \in \mathbb{N}$ such that $xk + 1 = yd_n$. Multiplying both sides by $c_n(k-1)+a$ gives $x'k+c_n(k-1)+a = y'd_n$ with new values $x', y' \in \mathbb{N}$. Let $a' := (x' + c_n)k + a = y'd_n + c_n$. Then $\mathrm{rem}(a' : d_i) = \mathrm{rem}(a : d_i) = c_i$ for all $i < n$ (because $d_i | k$). But also $\mathrm{rem}(a' : d_n) = c_n$, since $c_n < d_n$. ❑

Unlike those in most textbooks of number theory, the proof above is constructive and easily transferable to PA as will be shown in **7.1**. In logic it is occasionally not just important what you prove, but how you prove it.

Lemma 4.1 (on the $\boldsymbol{\beta}$-function). *For every n and every sequence $c_0, \ldots, c_n$ there exists some c such that $\boldsymbol{\beta}(c, i) = c_i$ for $i = 0, \ldots, n$.*

Proof. It suffices to provide numbers a and b such that $\boldsymbol{\alpha}(a, b, i) = c_i$ for all $i \leqslant n$. Because of $\boldsymbol{\beta}(\wp(a, b), i) = \boldsymbol{\alpha}(a, b, i)$ the claim is then satisfied with $c = \wp(a, b)$. Let $m = \max\{n, c_0, \ldots, c_n\}$ and $b := \mathrm{lcm}\{i + 1 \mid i \leqslant m\}$.[3] We claim that the numbers $d_i := 1 + (1 + i) \cdot b > c_i$ $(i \leqslant n)$ are pairwise coprime. Indeed, let p be a prime factor of d_i. Then $p > m$, for otherwise $p | b | d_i - 1$, contradicting $p | d_i$. If $p | d_i, d_j$ for $i < j \leqslant n$, then $p | d_j - d_i = (j - i)b$. But since $p \nmid b$ in view of $p > m$, it follows that $p | j - i \leqslant n \leqslant m < p$. Thus $j - i = 0$. Hence, $d_0, \ldots, d_n$ are pairwise coprime. By the Chinese remainder theorem there is an a such that $\mathrm{rem}(a : d_i) = c_i$, that is, $\boldsymbol{\alpha}(a, b, i) = c_i$ for $i = 0, \ldots, n$. ❑

Remark 1. Already at this stage we gain the interesting insight that the exponential function $(a, b) \mapsto a^b$ is explicitly definable in $\mathcal{N}$, namely by the Σ_1-formula

$$\delta_{exp}(x, y, z) := \exists u[\boldsymbol{\beta}(u, 0) = \mathsf{S}0 \wedge (\forall v{<}y)\, \boldsymbol{\beta}(u, \mathsf{S}v) = \boldsymbol{\beta}(u, v) \cdot x \wedge \boldsymbol{\beta}(u, y) = z].$$

More precisely, δ_{exp} is the *description* of a Σ_1-formula arising after the elimination of the occurring $\boldsymbol{\beta}$-terms by means of (1) and the use of further $\exists$-quantifiers. By induction on b one sees that $\mathcal{N} \vDash \delta_{exp}(\underline{a}, \underline{b}, \underline{c})$ implies $a^b = c$. Suppose conversely that $a^b = c$. Then Lemma 4.1 guarantees a sought-for u such that $\mathcal{N} \vDash \delta_{exp}(\underline{a}, \underline{b}, \underline{c})$: simply choose u such that $\boldsymbol{\beta}(u, i) = a^i$ for all $i \leqslant b$. This argument is generalized in Theorem 4.2 below.

For simplicity, we assume $T \supseteq \mathsf{Q}$ in Theorem 4.2 below, though it holds as well if Q is merely interpretable in T in the sense of **6.6**, for instance in ZFC. For the derivation of undecidability results or a simplified version of the first incompleteness theorem, the "Moreover" part of the theorem is not needed.

Theorem 4.2 (Representability theorem). *Each recursive function f—and hence every recursive predicate—is representable in an arbitrary consistent axiomatic extension $T \supseteq \mathsf{Q}$. Moreover, f is Σ_1-representable.*

[3] Here Gödel chooses $b = m!$, but our choice later alleviates the proof of this lemma in PA.

Proof. It suffices to construct a Σ_1-formula that represents f in Q. For the initial functions 0, $\mathtt{S}$, I^n_ν we may choose the formulas $\boldsymbol{v}_0 = 0$, $\boldsymbol{v}_1 = \mathtt{S}\boldsymbol{v}_0$ and $\boldsymbol{v}_n = \boldsymbol{v}_\nu$. Now let $f = h[g_1, \ldots, g_m]$ and suppose $\beta(\vec{y}, z)$ and $\gamma_i(\vec{x}, y_i)$ are Σ_1-formulas which represent h and the g_i. Then $\varphi(\vec{x}, z) := \exists \vec{y}\,[\bigwedge_i \gamma_i(\vec{x}, y_i) \wedge \beta(\vec{y}, z)]$ is such a formula for f (Lemma 3.5). Next let $f = \boldsymbol{Op}(g, h)$ and f, g both be Σ_1-representable. Then the predicate P, defined by $P(\vec{a}, b, c) \Leftrightarrow \boldsymbol{\beta}(c, 0) = g\vec{a} \wedge (\forall v{<}b)\boldsymbol{\beta}(c, \mathtt{S}v) = h(\vec{x}, v, \boldsymbol{\beta}(c, v))$, clearly results from a Δ_0- and hence Δ_1-definable predicate by the insertion of Σ_1-representable functions, and hence is Δ_1-representable according to Lemma 3.5(a). Obviously $P(\vec{a}, b, c)$ is equivalent to

$$(*) \qquad \boldsymbol{\beta}(c, i) = f(\vec{a}, i) \text{ for all } i \leqslant b.$$

By Lemma 4.1, for given $\vec{a}, b$ there is some c satisfying $(*)$, hence $\forall \vec{a}, b\, \exists c P(\vec{a}, b, c)$. Thus, $\tilde{f}\colon \vec{a} \mapsto \mu c P(\vec{a}, b, c)$ is Σ_1-representable by Lemma 3.4. Since $P(\vec{a}, b, \tilde{f}(\vec{a}, b))$, $(*)$ holds with $c = \tilde{f}(\vec{a}, b)$. This, for $i = b$, yields $f(\vec{a}, b) = \boldsymbol{\beta}(\tilde{f}(\vec{a}, b), b)$. Thus, as a composition of Σ_1-representable functions, f is Σ_1-representable. Finally, let f result from g by $\boldsymbol{O\mu}$, that is, $f\vec{a} = \mu b[P(\vec{a}, b)]$, where $P(\vec{a}, b) \Leftrightarrow g(\vec{a}, b) = 0$ and g is Σ_1-representable. By Lemma 3.4(c), g is Π_1-representable, too. This clearly implies that P is Δ_1-representable. Hence, f is Σ_1-representable by Lemma 3.4(a). ❏

Let $T \supseteq \mathsf{Q}$ be a theory in $\mathcal{L}_{ar}$. To $\varphi \in \mathcal{L}_{ar}$ corresponds *within* T the term $\underline{n}$ with $n := \dot{\varphi}$, which will be denoted by $\ulcorner\varphi\urcorner$ (or $\underline{\dot{\varphi}}$) and called the *Gödel term* of φ. For example, $\ulcorner \boldsymbol{v}_0 = 0\urcorner$ is $\underline{\dot{\boldsymbol{v}}_0 \doteq \dot{0}}$ $(= \underline{2^{22} \cdot 3^2 \cdot 5^{14}})$. Analogously $\ulcorner t\urcorner$ is defined for terms t. For instance, $\ulcorner \underline{1}\urcorner = \ulcorner \mathtt{S}0\urcorner = \underline{2^{16} \cdot 3^{14}}$. If T is axiomatized, also $\ulcorner\Phi\urcorner = \underline{\dot{\Phi}}$ for proofs Φ in T is well defined. For instance, $(\boldsymbol{v}_0 = \boldsymbol{v}_0)$ is for such a T a trivial proof of length 1 by axiom $\Lambda 9$ in **3.6**. Its Gödel term is $\underline{2^{\dot{v}_0 \doteq \dot{v}_0 + \dot{1}}}$. The predicate bew_T is p.r. (Theorem 2.4), hence Σ_1-representable (Theorem 4.2), by the formula $\mathtt{bew}_T(y, x)$, say. Define $\mathtt{bwb}_T(x) := \exists y\, \mathtt{bew}_T(y, x)$. Then Theorem 4.2 and (4) from page 178 obviously yield the following important

Corollary 4.3. *Let $T \supseteq \mathsf{Q}$ be axiomatizable. Then $\vdash_T \varphi \Rightarrow \vdash_T \mathtt{bew}_T(\underline{n}, \ulcorner\varphi\urcorner)$ for some n (hence $\vdash_T \varphi \Rightarrow \vdash_T \mathtt{bwb}_T(\ulcorner\varphi\urcorner)$), and $\nvdash_T \varphi \Rightarrow \vdash_T \neg\, \mathtt{bew}_T(\underline{n}, \ulcorner\varphi\urcorner)$ for all n.*

Theorem 4.2 has several other important consequences, for example Theorem 4.5 below. Before stating it we will acquaint ourselves with a method of eliminating Church's thesis from certain intuitively clear arguments that demand justification when "decidable" is identified with "recursive." Clearly, such an elimination must in principle always be possible if the thesis is to retain its legitimacy. For instance, Church's thesis was essentially used in the proof of Theorem 3.5.2. We reformulate it together with a rigorous proof.

Theorem 4.4. *A complete axiomatizable theory T is recursive.*

Proof. Because of completeness, $f\colon a \mapsto \mu b[a \in \dot{\mathcal{L}}^0 \;\Rightarrow\; bew_T(b, a) \vee bew_T(b, \dot{\neg} a)]$ is a well defined function. To see this, denote the recursive predicate in square brackets

by $P(a,b)$; then $\forall a \exists b P(a,b)$ (note that $P(a,0)$ in case $a \notin \dot{\mathcal{L}}^0$). By $\boldsymbol{O\mu}$, then, f is recursive. Note that $(*)$: $a \in \dot{T} \Leftrightarrow a \in \dot{\mathcal{L}}^0$ & $bew_T(fa,a)$ immediately implies the recursiveness of T. In order to prove $(*)$ let $a \in \dot{T}$, so certainly $a \in \dot{\mathcal{L}}^0$. Then for $b = fa$, the smallest b such that $bew_T(b,a) \vee bew_T(b,\dot{\neg}a)$, the first disjunct must hold, because due to the consistency of T, no $c \in \mathbb{N}$ with $bew_T(c,\dot{\neg}a)$ can exist at all. Hence, $bew_T(fa,a)$. The $\Leftarrow$-direction in $(*)$ is obvious. ❑

This proof illustrates sufficiently well the distinction between a primitive recursive and a recursive decision procedure. Even when X and thus the predicate P in the proof above are primitive recursive, the defined recursive function f need not be so, because the completeness of T may have been established in a nonconstructive way. The use of Church's thesis in the proofs of (i)⇒(ii) and (iii)⇒(ii) of the following theorem can be eliminated in almost exactly the same manner as above, although then the proof would lose much of its transparency.

Theorem 4.5. *For a predicate $P \subseteq \mathbb{N}^n$ and any consistent axiomatizable theory $T \supseteq \mathsf{Q}$ the following are equivalent:*

(i) *P is representable in T,* (ii) *P is recursive,* (iii) *P is Δ_1.*

Proof. (i)⇒(ii): Suppose P is represented in T by $\alpha(\vec{x})$. Given $\vec{a}$ we set going the enumeration machine of T and wait until $\alpha(\vec{a})$ or $\neg\alpha(\vec{a})$ appears. Thus, P is decidable and hence recursive by Church's thesis. (ii)⇒(i),(iii): By Theorem 4.2, χ_P is representable in T by a Σ_1-formula, hence P is Δ_1-representable by Lemma 3.4(d) and of course by the corresponding formulas also defined in $\mathcal{N}$. Thus, $P \in \Delta_1$. (iii)⇒(ii): Let P be defined by the Σ_1-formula $\alpha(\vec{x})$ and the Π_1-formula $\beta(\vec{x})$. Given $\vec{a}$ we kick start the enumeration machine for Q and wait until one of the Σ_1-sentences $\alpha(\underline{\vec{a}})$ or $\neg\beta(\underline{\vec{a}})$ appears. In the first case $P\vec{a}$ holds; in the second it does not. The procedure terminates because Q is Σ_1-complete by Theorem 3.1. ❑

This theorem tells us that in all consistent axiomatic extensions of Q exactly the same predicates are representable, namely the recursive ones. Moreover, Δ_1 contains precisely the recursive predicates, from which it easily follows that Σ_1 consists just of the r.e. predicates (observe Exercise 2 in **6.3**). Theorem 4.5 clarifies fairly well the close relationship between logic and recursion theory. It is independent of Church's thesis. Even if the thesis for some theoretical or practical reason had to be revised, the distinguished role of the μ-recursive functions would not be affected.

Remark 2. The above results allows us to define recursive or decidable predicates directly as follows: $P \subseteq N^n$ is recursive iff there is some finitely axiomatizable theory in which P is representable. We need only to notice that a predicate representable in any finitely axiomatizable theory in which representability makes sense, is recursive by Church's thesis. In this and the previous section we met several formulas or classes of those that represent predicates in Q and hence are recursive. It would of course be nice to provide a somewhat

more surveyable system of formulas that represent the recursive predicates, or at least that define them in $\mathcal{N}$. Unfortunately, such a system of formulas cannot be recursively enumerated. Indeed, suppose there is such an enumeration. Let $\alpha_0, \alpha_1, \ldots$ be the resulting sub-enumeration of its members in $\mathcal{L}^1_{ar}$. These define in $\mathcal{N}$ the recursive sets. Then also $\{n \in \mathbb{N} \mid n \notin \alpha_n^{\mathcal{N}}\}$ is recursive, hence is defined in $\mathcal{N}$ by α_m, say, so that $n \in \alpha_m^{\mathcal{N}} \Leftrightarrow n \notin \alpha_n^{\mathcal{N}}$. However, this equivalence yields for $n = m$ the contradiction $m \in \alpha_m^{\mathcal{N}} \Leftrightarrow m \notin \alpha_m^{\mathcal{N}}$.

In **6.5** we need a p.r. "substitution" function and in **7.1** a generalization of it. Let $\mathrm{cf}\, n := (\underline{n})^{\cdot}$ denote the Gödel number of the "cipher term" $\underline{n}$ $(= \mathtt{S}^n 0)$. Then $n \mapsto \mathrm{cf}\, n$ is p.r. since $\mathrm{cf}\, 0 = \dot{0}$ and $\mathrm{cf}\, \mathtt{S}n = \dot{\mathtt{S}} * \mathrm{cf}\, n$. Let $\mathrm{sb}_x(m, n) = [m]^{\mathrm{cf}\, n}_{\dot{x}}$ and define the p.r. function $\mathrm{sb}_{\vec{x}} \in \mathbf{F}_{n+1}$ as follows: $\mathrm{sb}_{\vec{x}}(m, \vec{a}) = \mathrm{sb}_{x_n}(\mathrm{sb}_{x_1,\ldots,x_{n-1}}(m, a_1, \ldots, a_{n-1}), a_n)$, $n > 1$. Here the x_i denote arbitrary but distinct variables. The function $\mathrm{sb}_{\vec{x}}$ may also be denoted by $\mathrm{sb}_{x_1 \cdots x_n}$. In order to have it defined for all sequences $\vec{x}$ including the empty one, set $\mathrm{sb}_{\emptyset}(m) = m$. Let $\dot{\alpha}_{\vec{x}}(\vec{\underline{a}})$ denote the Gödel number of the formula $\alpha_{\vec{x}}(\vec{\underline{a}})$ that arises from α by stepwise substituting $\underline{a_i}$ at the free occurrences of x_i in α for $i = 1, \ldots, n$ (see also page 48). Then we obtain

Theorem 4.6. $\mathrm{sb}_{\vec{x}}(\dot{\alpha}, \vec{a}) = \dot{\alpha}_{\vec{x}}(\vec{\underline{a}})$, *for arbitrary* $\alpha \in \mathcal{L}$ *and all* $\vec{a} \in \mathbb{N}^n$.

Proof. Since $\alpha_{\vec{x}}(\vec{\underline{a}})$ results from applying simple substitutions stepwise, we need only show that $\mathrm{sb}_x(\dot{\alpha}, a) = \dot{\alpha}_x(\underline{a})$ for all $\alpha \in \mathcal{L}$, $x \in \mathit{Var}$, and $a \in \mathbb{N}$. This is done by induction on α, starting with the proof of $\mathrm{sb}_x(\dot{t}, a) = \dot{t}_x(\underline{a})$; see Exercise 3. ❑

Example. Let α be $\mathtt{S}x = y$. Then $\mathrm{sb}_x(\dot{\alpha}, a) = (\mathtt{S}\underline{a} = y)^{\cdot}$ for all $a \in \mathbb{N}$. Furthermore, $\mathrm{sb}_{xy}(\dot{\alpha}, a, \mathtt{S}a) = (\mathtt{S}\underline{a} = \underline{\mathtt{S}a})^{\cdot} = \tilde{\mathtt{S}}\,\mathrm{cf}\, a \mathbin{\tilde{=}} \mathrm{cf}\, \mathtt{S}a = \tilde{\mathtt{S}}\,\mathrm{cf}\, a \mathbin{\tilde{=}} \tilde{\mathtt{S}}\,\mathrm{cf}\, a$, because $\mathrm{cf}\, \mathtt{S}a = \tilde{\mathtt{S}}\,\mathrm{cf}\, a$. But $\mathrm{sb}_x((\alpha\, {}^{\mathtt{S}x}_{y})^{\cdot}, a) = \tilde{\mathtt{S}}\,\mathrm{cf}\, a \mathbin{\tilde{=}} \tilde{\mathtt{S}}\,\mathrm{cf}\, a$ as well. Hence $\mathrm{sb}_{xy}(\dot{\alpha}, a, \mathtt{S}a) = \mathrm{sb}_x((\alpha\, {}^{\mathtt{S}x}_{y})^{\cdot}, a)$.

The example is generalized in Exercise 3, where we write $\vec{x}$ in place of $\vec{a}$. This simplifies the formulation of item (b) of the exercise. Therein the tuple $\vec{x}\,'$ may of course be empty, in which case (b) reduces to $\mathrm{sb}_{\vec{x}}(\dot{\alpha}, \vec{x}) = \dot{\alpha}$.

Exercises

1. Let $a, b, a_0, \ldots, a_n$ $(n > 0)$ be positive natural numbers and p a prime. Prove (a) $p \mid ab \Rightarrow p \mid a \vee p \mid b$, (b) $p \mid \mathrm{lcm}\{a_\nu \mid \nu \leqslant n\} \Rightarrow p \mid a_\nu$ for some $\nu \leqslant n$, and (c) $\mathrm{lcm}\{a_\nu \mid \nu < n\}$ and a_n are coprime provided $a_0, \ldots, a_n$ are pairwise coprime.

2. Provide a defining Σ_1-formula for the prime enumeration $n \mapsto p_n$.

3. Verify for arbitrary $\alpha, \beta \in \mathcal{L}_{ar}$ the following equations in $\mathbb{N}$:
   (a) $\mathrm{sb}_{\vec{x}}((\alpha \mathbin{\tilde{\wedge}} \beta)^{\cdot}, \vec{x}) = \mathrm{sb}_{\vec{x}}(\dot{\alpha}, \vec{x}) \mathbin{\tilde{\wedge}} \mathrm{sb}_{\vec{x}}(\dot{\beta}, \vec{x})$, and analogously for $\neg$, $\rightarrow$, and $\forall$.
   (b) $\mathrm{sb}_{\vec{x}}(\dot{\alpha}, \vec{x}) = \mathrm{sb}_{\vec{x}\,'}(\dot{\alpha}, \vec{x}\,')$ where $\vec{x}\,'$ covers all $x \in \mathit{free}\,\alpha$ such that $x \in \mathit{var}\,\vec{x}$.
   (c) $\mathrm{sb}_{\vec{x},x}(\dot{\alpha}, \vec{x}, t) = \mathrm{sb}_{\vec{x}}((\alpha\, {}^{t}_{\dot{x}})^{\cdot}, \vec{x})$ for $t \in \{0, y, \mathtt{S}y\}$ in the case $x \notin \mathit{free}\,\alpha$ or $y \in \mathit{var}\,\vec{x}$; otherwise $\mathrm{sb}_{\vec{x},x}(\dot{\alpha}, \vec{x}, t) = \mathrm{sb}_{\vec{x},y}((\alpha\, {}^{t}_{\dot{x}})^{\cdot}, \vec{x}, y)$. Here $y \notin \mathit{bnd}\,\alpha$.

6.5 The Theorems of Gödel, Tarski, Church

Call a theory $T \subseteq \mathcal{L}$ *arithmetizable* if $\mathcal{L}$ is arithmetizable and a sequence $(\underline{n})_{n\in\mathbb{N}}$ of constant terms is available such that $\vdash_T \underline{n} \neq \underline{m}$ for $n \neq m$ and $\mathrm{cf}: n \mapsto (\underline{n})^{\cdot}$ is p.r. This are minimal requirements for that representabilty of arithmetical predicates in T makes sense. They are trivially satisfied for $T \supseteq \mathsf{Q}$, but also for ZFC with respect to ω-terms (page 90). Terms and formulas are coded within T similar as in theories in $\mathcal{L}_{ar}$. In particular, $\ulcorner\alpha\urcorner = \dot{\underline{\alpha}}$ always denotes the Gödel term of a formula α. However, in order to evoke a concrete picture of the following two fairly general lemmas, take $\mathcal{L} = \mathcal{L}_{ar}$ and $T = \mathsf{PA}$ as standard examples.

A sentence γ is called a *fixed point of* $\alpha = \alpha(x)$ *in* T if $\gamma \equiv_T \alpha(\ulcorner\gamma\urcorner)$; equivalently, $\vdash_T \gamma \leftrightarrow \alpha(\ulcorner\gamma\urcorner)$. In intuitive terms, γ then says "α applies to me." The p.r. function sb_x from **6.4** is representable in T under relatively weak assumptions by Theorem 4.2. Hence, the lemmas below have a large spectrum of application.

Fixed-point lemma. *Let T be an arithmetizable theory and suppose that sb_x is representable in T. Then for each $\alpha = \alpha(x) \in \mathcal{L}$ there is some $\gamma \in \mathcal{L}^0$ such that*

(1) $\gamma \equiv_T \alpha(\ulcorner\gamma\urcorner)$.

Proof. Let $x_1, x_2, y \neq x$ and $\mathtt{sb}(x_1, x_2, y)$ be a formula representing sb_x in T. Then $\mathtt{sb}(\ulcorner\varphi\urcorner, \underline{n}, y) \equiv_T y = \ulcorner\varphi(\underline{n})\urcorner$ for all $\varphi = \varphi(x)$ and n. With $\underline{n} = \ulcorner\varphi\urcorner$ we then get

(2) $\mathtt{sb}(\ulcorner\varphi\urcorner, \ulcorner\varphi\urcorner, y) \equiv_T y = \ulcorner\varphi(\ulcorner\varphi\urcorner)\urcorner$.

Let $\beta(x) := \forall y(\mathtt{sb}(x, x, y) \to \alpha\,\frac{y}{x})$. Then $\gamma := \beta(\ulcorner\beta\urcorner)$ yields what we require. Indeed,

$$\begin{array}{lll} \gamma & = & \forall y(\mathtt{sb}(\ulcorner\beta\urcorner, \ulcorner\beta\urcorner, y) \to \alpha\,\frac{y}{x}) \\ & \equiv_T & \forall y(y = \ulcorner\beta(\ulcorner\beta\urcorner)\urcorner \to \alpha\,\frac{y}{x}) \qquad \big((2) \text{ with } \varphi := \beta(x)\big) \\ & = & \forall y(y = \ulcorner\gamma\urcorner \to \alpha\,\frac{y}{x}) \qquad \big(\text{because } \gamma = \beta(\ulcorner\beta\urcorner)\big) \\ & \equiv & \alpha(\ulcorner\gamma\urcorner). \qquad \square \end{array}$$

A fixed point can in the most interesting cases of α fairly easily be constructed, see **7.4**. The following lemma also formulates a frequently appearing argument.

Nonrepresentability lemma. Let T be a theory as in the fixed-point lemma. Then T (more precisely $\dot{T}$) is not representable in T itself.

Proof. Let T be represented by the formula $\tau(x)$. We show that even the weaker assumption (a): $(\forall\alpha{\in}\mathcal{L}^0) \nvdash_T \alpha \Leftrightarrow \vdash_T \neg\tau(\ulcorner\alpha\urcorner)$ leads to a contradiction. Indeed, let γ be a fixed point of $\neg\tau(x)$ according to (1), so that (b): $\vdash_T \gamma \Leftrightarrow \vdash_T \neg\tau(\ulcorner\gamma\urcorner)$. Choosing $\alpha = \gamma$ in (a) clearly yields with (b) the contradiction $\nvdash_T \gamma \Leftrightarrow \vdash_T \gamma$. $\square$

We now formulate Gödel's first incompleteness theorem, giving in fact three versions, of which the second corresponds essentially to the original. For simplicity, let henceforth $\mathcal{L} \supseteq \mathcal{L}_{ar}$ and $T \supseteq \mathsf{Q}$, ensuring the applicability of the two lemmas above.

However, all of the following holds for theories T, such as ZFC, in which Q is just interpretable in the sense of **6.6**.

Theorem 5.1 (the popular version). *Every consistent (recursively) axiomatizable theory $T \supseteq \mathsf{Q}$ is incomplete.*

Proof. If T is complete then it is recursive by Theorem 4.4, hence representable in T by Theorem 4.2, which is impossible by the nonrepresentability lemma. ❑

Unlike the proofs of Theorems 5.1′ and 5.1″, the above proof is nonconstructive, for it does not explicitly provide a sentence α such that $\nvdash_T \alpha$ and $\nvdash_T \neg\alpha$.

Stronger than the consistency of T is the so-called *ω-consistency* of T ($\subseteq \mathcal{L}_{ar}$), i.e., for all $\varphi = \varphi(x)$ such that $\vdash_T \exists x\varphi(x)$ we have $\nvdash_T \neg\varphi(\underline{n})$ for at least one n, or equivalently, if $\vdash_T \neg\varphi(\underline{n})$ for all n, then $\nvdash_T \exists x\varphi(x)$. Clearly, if $\mathcal{N} \models T$ then T is surely ω-consistent, because the supposition $\vdash_T \exists x\alpha$ and $\vdash_T \neg\alpha(\underline{n})$ for all n implies the contradiction $\mathcal{N} \models \exists x\alpha, \forall x\neg\alpha$. Thus, from a semantic perspective the theories Q and PA are certainly ω-consistent. Proof theory tries to eliminate nonfinitistic semantics, and there are famous consistency proofs for PA that presuppose considerably less than the full semantic approach; see for instance [Tak].

Theorem 5.1′ (the original version). *For every ω-consistent theory $T \supseteq \mathsf{Q}$ axiomatized by a p.r. axiom system X, there is a Π_1-sentence α such that neither $\vdash_T \alpha$ nor $\vdash_T \neg\alpha$. In other words, α is independent in T. There exists a primitive recursive function that assigns such an α to a formula representing X.*

Proof. Let bew_T be represented in T by the Σ_1-formula $\mathtt{bew}(y, x)$, see page 191. For $\mathtt{bwb}(x) = \exists y\,\mathtt{bew}(y, x)$ from Corollary 4.3 we obtain (a): $\vdash_T \varphi \Rightarrow \vdash_T \mathtt{bwb}(\ulcorner\varphi\urcorner)$, for all φ. Let γ be a fixed point of $\neg\,\mathtt{bwb}(x)$ by (1), so that (b): $\gamma \equiv_T \neg\,\mathtt{bwb}(\ulcorner\gamma\urcorner)$. The assumption $\vdash_T \gamma$ yields $\vdash_T \mathtt{bwb}(\ulcorner\gamma\urcorner)$ by (a), but $\vdash_T \neg\,\mathtt{bwb}(\ulcorner\gamma\urcorner)$ by (b), contradicting the consistency of T. Thus, $\nvdash_T \gamma$. Now assume $\vdash_T \neg\gamma$, so that $\vdash_T \mathtt{bwb}(\ulcorner\gamma\urcorner)$ by (b), hence (c): $\vdash_T \exists y\,\mathtt{bew}(y, \ulcorner\gamma\urcorner)$). Obviously $\nvdash_T \gamma$ because T is consistent. Applying Corollary 4.3 once again, we infer that $\vdash_T \neg\,\mathtt{bew}(\underline{n}, \ulcorner\gamma\urcorner)$ for all n. However, this and (c) contradict the ω-consistency of T. Consequently $\vdash_T \neg\gamma$ is impossible as well. Thus, γ is independent in T. But then too is the Π_1-sentence $\alpha := \neg\,\mathtt{bwb}(\ulcorner\gamma\urcorner)$ which is equivalent to γ in T. The claim of the p.r. assignment follows evidently from the construction of γ in the proof of (1). ❑

This theorem remains valid without restriction if the axiom system X is just r.e. In this case X can be replaced by some recursive X' (Exercise 1 in **6.2**), so that bew_T is still recursive according to Theorem 2.4.

Theorem 5.1″ (Rosser's strengthening of Theorem 5.1′). *The assumption of ω-consistency in Theorem 5.1′ can be weakened to the consistency of T.*

Proof. Let T be consistent and $\mathtt{prov}(x) := \exists y[\mathtt{bew}(y,x) \wedge (\forall z{<}y)\neg\,\mathtt{bew}(z,\dot{\neg}x)]$. We think here of the p.r. function $\dot{\neg}$ as having been eliminated in the usual way by a formula representing it. Because of the consistency of T, $\mathtt{prov}(x)$ says essentially the same as $\mathtt{bwb}(x)$ and has the following fundamental properties:

$$\text{(a)}\ \vdash_T \alpha \Rightarrow \vdash_T \mathtt{prov}(\ulcorner\alpha\urcorner), \qquad \text{(b)}\ \vdash_T \neg\alpha \Rightarrow \vdash_T \neg\,\mathtt{prov}(\ulcorner\alpha\urcorner).^4$$

Indeed, suppose $\vdash_T \alpha$ so that $\vdash_T \mathtt{bew}(\underline{n},\ulcorner\alpha\urcorner)$ for some n (observe Corollary 4.3). Since $\nvdash_T \neg\alpha$ it follows that $\vdash_T \neg\,\mathtt{bew}(\underline{k},\ulcorner\neg\alpha\urcorner)$ for all k. Therefore, C5 in **6.3** gives $\vdash_T (\forall z{<}\underline{n})\neg\,\mathtt{bew}(z,\ulcorner\neg\alpha\urcorner)$ and so $\vdash_T \mathtt{bew}(\underline{n},\ulcorner\alpha\urcorner) \wedge (\forall z{<}\underline{n})\neg\,\mathtt{bew}(z,\ulcorner\neg\alpha\urcorner)$, whence particularization yields the claim $\vdash_T \mathtt{prov}(\ulcorner\alpha\urcorner)$. Proof of (b): Suppose $\vdash_T \neg\alpha$, say $\vdash_T \mathtt{bew}(\underline{m},\ulcorner\neg\alpha\urcorner)$. Since $\nvdash_T \alpha$, we have $\vdash_T (\forall y{\leqslant}\underline{m})\neg\,\mathtt{bew}(y,\ulcorner\alpha\urcorner)$ by C5. This gives $\mathtt{bew}(y,\ulcorner\alpha\urcorner) \vdash_T y > \underline{m}$ by C6. Since $y > \underline{m} \vdash_T (\exists z{<}y)\,\mathtt{bew}(z,\ulcorner\neg\alpha\urcorner)$ (choose $z = \underline{m}$) it follows that $\vdash_T \forall y[\mathtt{bew}(y,\ulcorner\alpha\urcorner) \to (\exists z{<}y)\,\mathtt{bew}(z,\ulcorner\neg\alpha\urcorner)] \equiv \neg\,\mathtt{prov}(\ulcorner\alpha\urcorner)$. This confirms (b). Now let $\gamma \equiv_T \neg\,\mathtt{prov}(\ulcorner\gamma\urcorner)$ according to (1). The assumption $\vdash_T \neg\gamma$ then yields $\vdash_T \mathtt{prov}(\ulcorner\gamma\urcorner)$, contradicting (b), and the assumption $\vdash_T \gamma$ leads to a contradiction as in Theorem 5.1′. Thus, neither $\vdash_T \gamma$ nor $\vdash_T \neg\gamma$. ❑

$T \subseteq \mathcal{L}^0_{ar}$ is called *ω-incomplete* if there is some $\varphi = \varphi(x)$ such that $\vdash_T \varphi(\underline{n})$ for all n and yet $\nvdash_T \forall x\varphi$. We show that PA is not only incomplete but ω-incomplete. Let $\gamma \equiv_{\mathsf{PA}} \neg\,\mathtt{bwb}_{\mathsf{PA}}(\ulcorner\gamma\urcorner)$ and $\varphi(x) := \neg\,\mathtt{bew}_{\mathsf{PA}}(x,\ulcorner\gamma\urcorner)$. Then, by Theorem 5.1′, $\nvdash_{\mathsf{PA}} \gamma \equiv_{\mathsf{PA}} \neg\,\mathtt{bwb}_{\mathsf{PA}}(\ulcorner\gamma\urcorner) \equiv \forall x\varphi$, that is, $\nvdash_{\mathsf{PA}} \forall x\varphi$. On the other hand, since $\nvdash_{\mathsf{PA}} \gamma$ we know from Corollary 4.3 that $\vdash_{\mathsf{PA}} \varphi(\underline{n})$ $(= \neg\,\mathtt{bew}_{\mathsf{PA}}(\underline{n},\ulcorner\gamma\urcorner))$ for all n. Note that $\varphi(x)$ is even a Π_1-formula which is particularly surprising.

$\alpha \in \mathcal{L}^0$ is said to be *true in* $\mathcal{A}$ if $\mathcal{A} \vDash \alpha$. In particular, $\alpha \in \mathcal{L}^0_{ar}$ is called *true* (more precisely, true in $\mathcal{N}$ or true in reality, as some people like to say) if $\mathcal{N} \vDash \alpha$. If there is some $\tau(x) \in \mathcal{L}$ with a single free variable such that $\mathcal{A} \vDash \alpha \Leftrightarrow \mathcal{A} \vDash \tau(\ulcorner\alpha\urcorner)$, for all $\alpha \in \mathcal{L}^0$, it is said that *truth of $\mathcal{A}$ is definable in $\mathcal{A}$*. Clearly, this is equivalent to the representability of $Th\,\mathcal{A}$ in $Th\,\mathcal{A}$. For $\mathcal{A} = \mathcal{N}$, however, such a possibility is excluded by the nonrepresentability lemma. We therefore obtain

Theorem 5.2 (Tarski's nondefinability theorem). *The notion of truth in $\mathcal{N}$ is not definable in $\mathcal{N}$; in other words, $Th\,\mathcal{N}$ is not arithmetical.*

In this theorem lies the origin of a highly developed theory of definability in $\mathcal{N}$ (see also **6.7**). The theorem holds correspondingly for every domain of objects $\mathcal{A}$ whose language is arithmetizable and in which the function sb_x is representable for some variable x.

We now turn to undecidability results. First of all we prove the claim in Exercise 1 of **3.6** without recourse to Church's thesis.

[4] In particular $\vdash_T \neg\,\mathtt{prov}(\ulcorner\bot\urcorner)$. That the latter is not the case if we write $\mathtt{bwb}$ instead of $\mathtt{prov}$ is the import of Gödel's second incompleteness theorem 7.2.2. Thus, $\mathtt{bwb}$ and $\mathtt{prov}$ behave *within* T very differently, although $\mathtt{bew}_T(y,x) \equiv_{\mathcal{N}} \mathtt{prov}(y,x)$.

Lemma 5.3. *Every finite extension T' of a decidable theory T of one and the same (arithmetizable) language $\mathcal{L}$ is decidable.*

Proof. Suppose T' extends T by $\alpha_0, \ldots, \alpha_n$ and $\alpha := \bigwedge_{i \leqslant n} \alpha_i$, so that $T' = T + \alpha$. Since $\beta \in T' \Leftrightarrow \alpha \to \beta \in T$, we obtain $n \in \dot{T}' \Leftrightarrow n \in \dot{\mathcal{L}}^0 \;\&\; \dot{\alpha}_0 \mathbin{\dot{\to}} n \in \dot{T}$. Now, $\dot{T}$, $\dot{\mathcal{L}}^0$, and $\dot{\to}$ are recursive. Hence the same applies to $\dot{T}'$. ❏

That T' belongs to the same language as T is important here. A decidable theory T axiomatized by $X \subseteq \mathcal{L}^0$, if considered as a theory in $\mathcal{L}' \supset \mathcal{L}$ with the same axiom system X, may well be undecidable, e.g., due to the additional tautologies of $\mathcal{L}'$.

$T_0 \subseteq \mathcal{L}_0$ is called *strongly undecidable* if T_0 is consistent and each theory in $\mathcal{L}$ compatible with T_0 is undecidable. Then each theory T compatible with T_0 in a language $\mathcal{L} \supseteq \mathcal{L}_0$ is also undecidable, for otherwise $T \cap \mathcal{L}_0$ would clearly be decidable. If T_0 is strongly undecidable so is every consistent $T_1 \supseteq T_0$, for if T is compatible with T_1, then it is also compatible with T_0. Moreover, each subtheory of T_0 in $\mathcal{L}_0$ is then also undecidable, or T_0 is *hereditarily undecidable* in the terminology of [TMR]. The weaker a strongly undecidable theory, the wider the scope of applications. This will become plain by means of examples in the next section.

Theorem 5.4. ([TMR]). Q *is strongly undecidable.*

Proof. Assume $T \cup \mathsf{Q}$ is consistent and T decidable. The same holds by Lemma 5.3 for the finite extension $T_1 = T + \mathsf{Q}$. But then, by Theorem 4.2, T_1 is representable in itself, which again is impossible by the nonrepresentability lemma. ❏

Theorem 5.5 (Church's undecidability theorem). *The set $\mathsf{Taut}_{\mathcal{L}}$ of all tautological sentences is undecidable for $\mathcal{L} \supseteq \mathcal{L}_{ar}$.*

Proof. $\mathsf{Taut}_{\mathcal{L}}$ is surely compatible with Q, hence undecidable by Theorem 5.4. ❏

This result readily carries over to the language with a single binary relation, as will be shown in the next section, and hence to all expansions of this language. Indeed, it carries over to all languages with the exception of those containing unary predicate symbols only and at most one unary function symbol. For the tautologies of these languages there exist various decision procedures; see [ML, vol. I].

By Theorem 5.4, in particular $Th\mathcal{N}$ is undecidable; likewise is every subtheory of $Th\mathcal{N}$, for instance Peano arithmetic PA and each of its subtheories, as well as all consistent extensions of PA, because these are all compatible with Q. $Th\mathcal{N}$ is not even axiomatizable, since an axiomatizable complete theory is decidable. Further conclusions concerning undecidable theories will be drawn in **6.6**.

Alongside undecidability results concerning formalized theories, numerous special results can also be obtained in a similar manner; for instance negative solutions to word problems of all kinds, and halting problems (see e.g. [Rog] or [Ba, C2]).

Of these perhaps the most spectacular was the solution to Hilbert's tenth problem: Does an algorithm exist that for every polynomial $p(\vec{x})$ with integer coefficients decides whether the Diophantine equation $p(\vec{x}) = 0$ has a solution in $\mathbb{Z}$? The answer is no, as Matiyasevich proved in 1970.

We briefly sketch the proof. It suffices to show that no algorithm exists for the solvability of all Diophantine equations in $\mathbb{N}$. Indeed, by a well-known theorem from Lagrange, every natural number is the sum of four squares of integers. Consequently $p(\vec{x}) = 0$ is solvable in $\mathbb{N}$ iff $p(u_1^2 + v_1^2 + w_1^2 + z_1^2, \cdots, u_n^2 + v_n^2 + w_n^2 + z_n^2) = 0$ is solvable in $\mathbb{Z}$. Thus, if we could decide the solvability of Diophantine equations in $\mathbb{Z}$, then we could solve as well the corresponding problem in $\mathbb{N}$. For the latter notice first of all that the question of solvability of $p(\vec{x}) = 0$ in natural numbers is equivalent to the solvability of a Diophantine equation of $\mathcal{L}_{ar}$ (i.e., an equation $s(\vec{x}) = t(\vec{x})$), by simply bringing all terms of $p(\vec{x})$ preceded by a minus sign "to the other side." Thus, Hilbert's problem is reduced to the question of a decision procedure for the problem $\mathcal{N} \vDash \exists \vec{x} \delta(\vec{x})$, where $\delta(\vec{x})$ runs through all Diophantine equations in $\mathcal{L}_{ar}$.

The negative solution to the last question follows easily from the much further-reaching Theorem 5.6, which establishes a surprising connection between number and recursion theory; it is proved in detail in [Mat]. This theorem is a paradigm of the experience that certain mathematical questions lead to results whose significance extends way beyond that of an answer to the original question.

Theorem 5.6. *An arithmetical predicate P of any arity is Diophantine if and only if P is recursively enumerable.*

To give at least an indication of the proof, let the Diophantine predicate $P \subseteq \mathbb{N}^n$ be defined by $P\vec{a} \Leftrightarrow \mathcal{N} \vDash \exists \vec{y} \delta^{\mathcal{N}}(\underline{\vec{a}}, \vec{y})$, with the equation $\delta(\vec{x}, \vec{y})$. The defining formula for P is Σ_1 and hence is r.e., because $\delta^{\mathcal{N}}(\underline{\vec{a}}, \vec{y})$ is recursive by Theorem 4.5. This is so to speak the trivial direction of the claim. The converse, that every r.e. predicate is Diophantine, is too large in scope to be given here. Much tricky inventiveness must be used in order to show that numerous arithmetical predicates and functions are Diophantine. Among these is the ternary predicate '$a^b = c$', which for a long time resisted the proof of being Diophantine. This theorem easily yields

Corollary 5.7. (a) *Hilbert's tenth problem has a negative answer.* (b) *For every axiomatizable theory $T \supseteq \mathsf{Q}$, in particular for $T = \mathsf{PA}$, there exists an unsolvable Diophantine equation whose unsolvability is provable in T.*

Proof. bwb_{Q} is by **6.2** r.e. Hence, by Theorem 5.6, there exists a Diophantine equation $\delta(x, \vec{y})$ such that $(*)$: $bwb_{\mathsf{Q}}(n) \Leftrightarrow \mathcal{N} \vDash \exists \vec{y}\, \delta(\underline{n}, \vec{y})$. We claim that even for the set $\{\exists \vec{y}\, \delta(\underline{n}, \vec{y}) \mid n \in \mathbb{N}\}$ of Diophantine sentences there is no decision procedure. Otherwise $\{n \in \mathbb{N} \mid \mathcal{N} \vDash \exists \vec{y}\, \delta(\underline{n}, \vec{y})\}$ would be recursive, and thus by $(*)$ so too bwb_{Q},

contradicting Theorem 5.4. This proves (a). If the unsolvability of every unsolvable Diophantine equation $\delta(\vec{x})$ were provable in T, then either $\vdash_T \neg\exists\vec{x}\delta(\vec{x})$ (provided $\delta(\vec{x})$ is unsolvable) or $\vdash_T \exists\vec{x}\delta(\vec{x})$ otherwise, because of the Σ_1-completeness of T. Since the theorems of T are r.e., one would then have a decision procedure for the solvability of Diophantine equations, which contradicts part (a). ❑

Theorem 5.6 can be yet further strengthened; namely, it can be proved within PA. Thus, one obtains the following theorem, whose name stems from **M**atiyasevich, **R**obinson, **D**avis, and **P**utnam, all of whom made significant contributions to the solution of Hilbert's tenth problem. Because of the lengthy proof, we shall not use the theorem, though in fact many things would thereby be simplified.

MRDP theorem. *For every Σ_1-formula α there exists an $\exists$-formula φ in $\mathcal{L}_{ar}$ such that $\alpha \equiv_{\mathsf{PA}} \varphi$. Here φ is without loss of generality of the form $\exists\vec{x}\, s = t$.*

Fermat's meanwhile proved conjecture

$$(*) \qquad (\forall x\, y\, z \neq 0)(\forall n{>}2)\, x^n + y^n \neq z^n$$

is a Π_1-sentence, because by Theorem 4.5 $(x, y) \mapsto x^y$ is Δ_1 and a fortiori explicitly definable by $0, \mathsf{S}, +, \cdot$ in $\mathcal{N}$. This was noticed already in Remark 1 in **6.4**. Hence, the conjecture $(*)$ is a candidate for a sentence which may be independent in PA.

Remark. It would be interesting to discover whether the conjecture's proof at the end of the last century, or any modification can be carried out in PA. A demonstration that this is not the case would be hardly less sensational than the solution of the problem itself. However, it seems that the proof can be carried out in a suitable conservative extension of PA (communication by letter to the author from G. Kreisel). Note also the following: Since PA is ω-incomplete already for Π_1-formulas (page 196), it may even be the case that $\vdash_{\mathsf{PA}} (\forall x \forall y \forall z \neq 0)\, x^{\underline{n}} + y^{\underline{n}} \neq z^{\underline{n}}$ for every single $n > 2$, although $(*)$ is not provable in PA.

Exercises

1. Show that an ω-incomplete theory in $\mathcal{L}_{ar}$ has a consistent but ω-inconsistent extension.

2. Suppose T is complete; prove the equivalence of

   (i) T is strongly undecidable, (ii) T is hereditarily undecidable.

3. Let Δ be a finite list containing explicit definitions of new symbols in terms of those occurring in $\mathcal{L}$. Show that if T is decidable then so is $T+\Delta$ (independent of whether all definitions in Δ are legitimate in T).

4. Construct a primitive recursive function $f : \mathbb{N} \to \mathbb{N}$ such that $ran\, f$ is not recursive. *Hint*: Note that the set of all proofs in Q is p.r.

6.6 Transfer by Interpretation

Interpretability is a powerful method to transfer model-theoretical and other properties, such as undecidability, from one theory to another. Roughly speaking, interpreting a theory $T_0 \subseteq \mathcal{L}_0$ into a theory $T_1 \subseteq \mathcal{L}_1$ means to make the basic notions of T_0 understandable in T_1 via explicit definitions. Quantifiers from T_0 run over subdomains of the domains of T_1-models, that is, "for all x" from T_0 is replaced in T_1 by "for all $x \in \mathtt{P}$", where $\mathtt{P}$ is a unary predicate symbol for the domains of T_0-models. We consider the most important concepts, interpretability from Tarski (also called *relative interpretability*) and interpretability from Rabin, called *model interpretability*. All theories are supposed to be consistent in this section.

Let $\mathtt{P}$ be a unary predicate symbol not occurring in T_1. The formula $\varphi^{\mathtt{P}}$, the $\mathtt{P}$-*relativized* of a formula φ, results from φ by replacing all subformulas of the form $\forall x\alpha$ by $\forall x(\mathtt{P}\,x \to \alpha)$. For open φ nothing happens, that is, $\varphi^{\mathtt{P}} = \varphi$. A strict definition of $\varphi^{\mathtt{P}}$ runs by induction: $\varphi^{\mathtt{P}} = \varphi$ if φ is a prime formula, $(\neg\varphi)^{\mathtt{P}} = \neg\varphi^{\mathtt{P}}$, $(\varphi \wedge \psi)^{\mathtt{P}} = \varphi^{\mathtt{P}} \wedge \psi^{\mathtt{P}}$, and $(\forall x\varphi)^{\mathtt{P}} = \forall x(\mathtt{P}\,x \to \varphi^{\mathtt{P}})$. This implies $(\exists x\varphi)^{\mathtt{P}} \equiv \exists x(\mathtt{P}\,x \wedge \varphi^{\mathtt{P}})$ as is easily confirmed. Here a formula equivalent to $(\exists x\varphi)^{\mathtt{P}}$ has been displayed to show more clearly what relativation is intending to mean.

Example. $(\forall x\exists y\, y = \mathtt{S}x)^{\mathtt{P}} \equiv \forall x(\mathtt{P}\,x \to \exists y(\mathtt{P}\,y \wedge y = \mathtt{S}x)) \equiv \forall x(\mathtt{P}\,x \to \mathtt{P}\,\mathtt{S}x)$. As regards the second equivalence observe that $\exists y(\mathtt{P}\,y \wedge y = \mathtt{S}x) \equiv \mathtt{P}\,\mathtt{S}x$ (cf. (12) in **2.4**).

Definition. $T_0 \subseteq \mathcal{L}_0$ is called *interpretable* in $T_1 \subseteq \mathcal{L}_1$ (where for simplicity we assume that T_0 has finite signature) if there is a list Δ containing explicit definitions legitimate in T_1 of the symbols of T_0 not occurring in T_1 and of a unary predicate symbol $\mathtt{P}$, so that $T_0^{\mathtt{P}} \subseteq T_1^{\Delta}$. Here generally $X^{\mathtt{P}} := \{\alpha^{\mathtt{P}} \mid \alpha \in X\}$, and $T_1^{\Delta} := T_1 + \Delta$ is a theory in $\mathcal{L}_1^{\Delta}$ whose signature is $L_0 \cup L_1 \cup \{\mathtt{P}\}$ (L_i is the signature of $\mathcal{L}_i$).

This technically somewhat involved definition only expresses that all notions of T_0 "are understood" in T_1, and all that can be proven in T_0 can also be proven in T_1. Interpretability generalizes the notion of a subtheory: If $T_0 \subseteq T_1$ then T_0 is trivially interpretable in T_1, with $\Delta = \{\mathtt{P}\,x \leftrightarrow x = x\}$. In this case clearly $\alpha^{\mathtt{P}} \equiv \alpha$ modulo Δ.

Let CA be the set of the so-called *closure axioms*

$$\exists x\,\mathtt{P}\,x,\ \mathtt{P}\,c, \text{ and } \forall\vec{x}\,(\textstyle\bigwedge_{i=i}^{n}\mathtt{P}\,x_i \ \to \mathtt{P}\,f\vec{x}), \text{ for all } c, f \in L_0.$$

These sentences are equivalent to $(\exists x\, x = x)^{\mathtt{P}}$, $(\exists x\, x = c)^{\mathtt{P}}$, and $(\forall\vec{x}\,\exists y\, y = f\vec{x})^{\mathtt{P}}$, respectively. Thus, CA is up to equivalence a set of the form $F^{\mathtt{P}}$ for some finite set $F \subseteq \mathit{Taut}_{\mathcal{L}_0}$ and therefore $CA \subseteq T_0^{\mathtt{P}}$ in any case. The sentences of CA guarantee that for a given $\mathcal{B} \vDash \Delta$ there is a well-defined $\mathcal{L}_0$-structure $\mathcal{A}$ whose domain is $A = \mathtt{P}^{\mathcal{B}}$. $\mathcal{A}$'s relations and operations are the ones defined by Δ but restricted to A. This structure $\mathcal{A}$ will be denoted by $\mathcal{B}_{\Delta}$. It is a substructure of the $\mathcal{L}_0$-reduct of $\mathcal{B}$ whose role will become clear in the next lemma. Examples will be given later.

Lemma 6.1. *Let $\mathcal{B} \models CA$. Then $\mathcal{B}_\Delta \models \alpha \Leftrightarrow \mathcal{B} \models \alpha^{\mathsf{P}}$, for all sentences $\alpha \in \mathcal{L}_0^0$.*

Proof. $\mathcal{A} := \mathcal{B}_\Delta$ is an $\mathcal{L}_0$-structure. Claim: $(\mathcal{A}, w) \models \varphi \Leftrightarrow (\mathcal{B}, w) \models \varphi^{\mathsf{P}}$, for any $w\colon Var \to A$. This proves the lemma since α is a sentence. We prove the claim by induction on $\varphi \in \mathcal{L}_0$. It is clear for prime formulas α since $\alpha^{\mathsf{P}} = \alpha$. The induction steps for $\wedge, \neg$ proceed without difficulty, and the one for $\forall$ is obtained as follows:

$$\begin{aligned}(\mathcal{A}, w) \models \forall x\varphi &\Leftrightarrow (\mathcal{A}, w^a_x) \models \varphi \text{ for all } a \in A\\ &\Leftrightarrow (\mathcal{B}, w^a_x) \models \varphi^{\mathsf{P}} \text{ for all } a \in A \quad \text{(induction hypothesis)}\\ &\Leftrightarrow (\mathcal{B}, w^a_x) \models \mathsf{P}\,x \to \varphi^{\mathsf{P}}, \text{ for all } a \in B \quad (\text{because } \mathsf{P}^{\mathcal{B}} = A)\\ &\Leftrightarrow (\mathcal{B}, w) \models \forall x(\mathsf{P}\,x \to \varphi^{\mathsf{P}}) = (\forall x\varphi)^{\mathsf{P}}. \quad \square\end{aligned}$$

If T_0 is axiomatized by X_0 then in the above definition it suffices to require just $X_0^{\mathsf{P}} \cup CA \subseteq T_1^\Delta$ instead of $T_0^{\mathsf{P}} \subseteq T_1^\Delta$. This fact is highly important for applications and follows immediately from

$$(*) \qquad S \vdash \alpha \Rightarrow S^{\mathsf{P}} \cup CA \vdash \alpha^{\mathsf{P}} \quad (S \cup \{\alpha\} \subseteq \mathcal{L}_0^0).$$

For proving $(*)$ let $S \vdash \alpha$ and $\mathcal{B} \models S^{\mathsf{P}} \cup CA$. Then $\mathcal{B}_\Delta \models S$ by the lemma. Thus, $\mathcal{B}_\Delta \models \alpha$ since $S \vdash \alpha$. Consequently $\mathcal{B} \models \alpha^{\mathsf{P}}$. Since $\mathcal{B}$ was arbitrary, $S^{\mathsf{P}} \cup CA \vdash \alpha^{\mathsf{P}}$.

Theorem 6.2. *Let T_0 be interpretable in T_1. If T_0 is strongly undecidable so is T_1.*

Proof. Let $T \subseteq \mathcal{L}_1$ be compatible with T_1. Then $T + T_1$ is consistent and so is $(T + T_1)^\Delta$. Now, $S := \{\alpha \in \mathcal{L}_0^0 \mid \alpha^{\mathsf{P}} \in T^\Delta + CA\}$ is a theory (consider $(*)$ and $S^{\mathsf{P}}, CA \subseteq T^\Delta + CA$). Let $\mathcal{B} \models (T + T_1)^\Delta$ $(\supseteq T^\Delta, T_1^\Delta, CA)$. Then also $\mathcal{B} \models T_0^{\mathsf{P}}, S^{\mathsf{P}}$, since $T_0^{\mathsf{P}} \subseteq T_1^\Delta$. Thus, $\mathcal{B}_\Delta \models T_0, S$ by Lemma 6.1, hence S is compatible with T_0 and so undecidable. If T were decidable, then so would be T^Δ (Exercise 3 in **6.5**). Hence also $T^\Delta + CA$ (Lemma 5.3), and so clearly S. This is a contradiction. $\square$

Example. Q is interpretable in the theory T_d of discretely ordered rings, i.e., ordered rings $\mathcal{R} = (R, 0, +, \times, <)$ with a smallest positive element e, which need not be the unit element of $\mathcal{R}$. Ring multiplication is denoted by $\times$, in order to distinguish it from multiplication in Q. Choose the following definitions for P, S, $\cdot$:

$$\mathsf{P}\,x \leftrightarrow x \geqslant 0 \ \wedge\ x \times e = e \times x \ \wedge\ \forall y \exists z\ z \times e = y \times x,$$

$$y = \mathsf{S}x \leftrightarrow y = x + e, \quad z = x \cdot y \leftrightarrow z \times e = x \times y \ \vee\ \forall u(u \times e \neq x \times y \ \wedge\ z = x).$$

Since $x = e \leftrightarrow 0 < x \ \wedge\ \forall y(0 < y \to x \leqslant y)$, e is eliminable from all these formulas. $0, +$ remain unaltered. With a somewhat patient calculation, all the axioms of Q relativized to P can be proved in T_d^Δ, as can all the closure axioms. Thus, T_d is strongly undecidable according to Theorem 6.2.

While Q is not directly interpretable in the theory T_R of rings or in the theory T_F of fields, it is in a certain finite extension of T_F (Julia Robinson), whereby T_F and T_R are shown to be undecidable. The same also holds for the theory of groups T_G as is shown in [TMR]. However, none of these theories is strongly undecidable.

Q and also PA are interpretable in ZFC, as is nearly every other theory (some exceptions are considered in **7.6**). Let $\mathtt{P}\,x \leftrightarrow x \in \omega$, and define $\mathtt{S}, +, \cdot$ within ZFC such that their restrictions to ω coincide with the usual operations. In particular, $\mathtt{S}$ may be defined by $y = \mathtt{S}x \leftrightarrow y = x \cup \{x\}$. This immediately yields the incompleteness and the undecidability of ZFC, assuming of course its consistency. Q is also interpretable in very weak subtheories of ZFC, for instance in the theory $T_\in$ with the following three axioms.[5] Hence, like Q, the theory $T_\in$ is strongly undecidable.

$$\begin{array}{ll} \exists x \forall y\, y \notin x & (\emptyset \text{ exists}), \\ \forall x \forall y (\forall z (z \in x \leftrightarrow z \in y) \to x = y) & (\text{extensionality}), \\ \forall x \forall y \exists z \forall u (u \in z \leftrightarrow u \in x \vee u = y) & (x \cup \{y\} \text{ exists}). \end{array}$$

In particular, the set of tautologies in a binary relation is undecidable, indeed even without identity in the language; for notice that $=$ can conservatively be eliminated from $T_\in$ by means of the explicit definition $x = y \leftrightarrow \forall z (z \in x \leftrightarrow z \in y)$.

Q is surely interpretable in $Th\mathcal{N}$ and $Th\mathcal{N}$ in turn in $Th\,\mathcal{Z}$ with $\mathcal{Z} = (\mathbb{Z}, 0, 1, +, \cdot)$. This is a consequence of Lagrange's theorem. Hence, $Th\,\mathcal{Z}$ is strongly undecidable, and thus every subtheory is undecidable, e.g., the theory of commutative rings.

$Th\mathcal{N}$ and $Th\,\mathcal{Z}$ have the same degree of complexity, because $Th\,\mathcal{Z}$ is (in various ways) interpretable in $Th\mathcal{N}$; for instance, let the even and odd numbers play the role of nonnegative and negative integers, respectively. Highly interesting also is the mutual interpretability of PA and $\mathsf{ZFC}_{\text{fin}}$. This is the *theory of finite sets*, resulting from ZFC by replacing the axiom of infinity with the axiom "all sets are finite."

We now describe a stricter notion of interpretability, though for simplicity we omit some details. Let $\boldsymbol{K}_0$ and $\boldsymbol{K}$ be nonempty classes of $\mathcal{L}_0$- and $\mathcal{L}$-structures, respectively. Further, let Δ be a list of definitions of the $\mathcal{L}_0$-symbols and a predicate symbol P, and $\mathcal{L}^\Delta$, CA and $\mathcal{B}_\Delta$ for $\mathcal{B} \vDash CA$ as above. $\mathcal{A}^\Delta$ denotes the expansion of $\mathcal{A} \in \boldsymbol{K}$ in $\mathcal{L}^\Delta$ according to the definition list Δ (the Δ-*expansion* of $\mathcal{A}$). For a fixed sentence $\gamma \in \mathcal{L}^\Delta$ set $\boldsymbol{K}_\gamma := \{\mathcal{A}^\Delta \mid \mathcal{A} \in \boldsymbol{K},\ \mathcal{A}^\Delta \vDash \gamma\}$. For each sentence $\beta \in \mathcal{L}^\Delta$ we can effectively construct, as in **2.6**, a reduced formula $\beta^{rd} \in \mathcal{L}$ such that

(0) $\mathcal{A}^\Delta \vDash \beta \ \Leftrightarrow\ \mathcal{A} \vDash \beta^{rd}$, for all $\mathcal{A} \in \boldsymbol{K}$.

Definition. Call $\boldsymbol{K}_0$ (or $Th\,\boldsymbol{K}_0$) *model interpretable* in $\boldsymbol{K}$ ($Th\,\boldsymbol{K}$, respectively.) if with respect to suitable Δ and γ the following conditions hold:

(1) $\boldsymbol{K}_\gamma \vDash CA$ and $\mathcal{B}_\Delta \in \boldsymbol{K}_0$ for all $\mathcal{B} \in \boldsymbol{K}_\gamma$,

(2) For every $\mathcal{A} \in \boldsymbol{K}_0$ there is a $\mathcal{B} \in \boldsymbol{K}_\gamma$ such that $\mathcal{A} \simeq \mathcal{B}_\Delta$.

Theorem 6.3. *Let $\boldsymbol{K}_0$ be model interpretable in $\boldsymbol{K}$. If $Th\,\boldsymbol{K}_0$ is undecidable then so too is $Th\,\boldsymbol{K}$.*

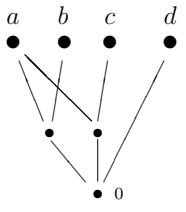
[5] Claimed in [TMR]. The very lengthy proof is presented in [Mo, pp. 283–290].

Proof. It suffices to show $(*)$: $\boldsymbol{K}_0 \vDash \alpha \;\Leftrightarrow\; \boldsymbol{K} \vDash \hat{\alpha}$, where $\hat{\alpha} := (\gamma \to \alpha^{\mathsf{P}})^{rd}$, because a decision procedure for $Th\,\boldsymbol{K}$ would by $(*)$ also mean one for $Th\,\boldsymbol{K}_0$. Let $\boldsymbol{K}_0 \vDash \alpha$, $\mathcal{A} \in \boldsymbol{K}$, and $\mathcal{A}^{\Delta} \vDash \gamma$ so that $\mathcal{A} \vDash \gamma^{rd}$ and $\mathcal{B} := \mathcal{A}^{\Delta} \in \boldsymbol{K}_{\gamma}$. By (1), $\mathcal{B}_{\Delta} \in \boldsymbol{K}_0$, thus $\mathcal{B}_{\Delta} \vDash \alpha$, hence $\mathcal{B} \vDash \alpha^{\mathsf{P}}$ by Lemma 6.1, and so $\mathcal{A} \vDash (\alpha^{\mathsf{P}})^{rd}$. This proves $\mathcal{A} \vDash \hat{\alpha}$ for all $\mathcal{A} \in \boldsymbol{K}$, i.e., $\boldsymbol{K} \vDash \hat{\alpha}$. The $\Leftarrow$-direction is easily proved by contraposition. ❑

Example. Let $\boldsymbol{K}_0$ be the class of all graphs (A, R) and $\boldsymbol{K}$ that of all simple graphs (B, S), that is, S is irreflexive and symmetrical. The figure shows an $\mathcal{A} \in \boldsymbol{K}_0$ such

that aRa, aRb, bRa, and bRc, and on the right the simple graph $\mathcal{B}$ corresponding to $\mathcal{A}$ according to (2) with $\mathcal{A} \simeq \mathcal{B}_{\Delta}$, the "encoding structure" of $\mathcal{A}$. Roughly put, a set N of new points is adjoined to A so that $\mathcal{A}$ is completely described by $\mathcal{B}$. In $\mathcal{B}$, whose domain is $B = A \cup N$, the edges are undirected, but we pay for this by increasing the number of points for coding in $\mathcal{B}$ the directions of the edges in $\mathcal{A}$. The "old," in the figure bold-printed points in $A = \mathsf{P}^{\mathcal{B}}$ neighbor two or three endpoints. These are points in which exactly one S-edge ends. Informally, the definition for R in the structure $\mathcal{B}$ then reads "xRy iff $x, y \in \mathsf{P}^{\mathcal{B}}$ and either $x = y$ and x neighbors three endpoints, or there exists exactly one new point z such that $xSzSy$, or there exist precisely two new points u, v such that $xSuSvSy$ and uSy." γ is rendered informally into "$\exists x\, \mathsf{P}\, x$ and all new points are either endpoints or neighbor exactly two old points or exactly one old and one new point."

In the example, $Th\,\boldsymbol{K}_0$ is the logical theory of a binary relation, already established as undecidable. Accordingly the theory of all simple graphs is undecidable. Now, this can be used to show, e.g., that the theory SL of semilattices is undecidable. By Theorem 5.4 the same then follows for the theory SG of semigroups, since SL is a finite extension of SG. Similar to the last example, it suffices to provide for any simple graph (A, S), the encoding semilattice $(B, \circ)$. The figure on the left shows the ordering diagram of B for $A = \{a, b, c, d\}$ and $S = \{\{a, b\}\{ac\}\}$, where S is understood as a set of edges; cf. **1.5**. The old points are precisely the maximal points of B. By construction, B has a smallest element 0 and is of length 3, i.e., there are at most three consecutive points in B with respect to $<$. This must now be expressed by the sentence γ required in the definition.



The theory of *finite* simple graphs with or without some additional feature (for instance planarity) is undecidable; see e.g. [RZ]. The above construction shows that the undecidable theory of finite simple graphs is model interpretable in the

theory of finite semilattices which hence is undecidable. This clearly implies the undecidability of the theory FSG of finite semigroups. Setting an element on top of the maximal elements in the last figure results in the diagram of a finite lattice, so that the theory of finite lattices turns out to be undecidable. The same holds for the theory FPO of finite partial orders because for the description of (A, S) only the partial order of B is relevant.

Remark. Somewhat more mathematics is required to prove the undecidability of the theory FDL of all *finite distributive lattices*. The figure shows first of all that the theory FPO of finite partial orders $(g, <)$ is also undecidable. But FPO is model interpretable in FDL, in that one identifies the elements of g with the $\cap$-irreducible elements of the lattice, $\mathcal{A}$ say. Here we need to know that $\mathcal{A}$'s structure is completely determined by the partial order of its irreducible elements and that this order can arbitrarily be given.

Positive results are also transferable. For instance the (logical) theory of a unary function is interpretable in the elementary theory of (undirected) trees ([KR]), and with the latter the former is also decidable. The decidability of the theory of a single unary function was first proved by Ehrenfeucht with a different method. Let us mention that the theory of two or more unary functions is undecidable because several undecidable theories are model interpretable in it.

Decidability of the theory of simple trees also follows from the decidability of the second-order monadic theory of binary trees ([Ba, C3]), a very strong result with an immense scope of applications. One of these applications is a relatively simple proof of decidability of a variety of logical systems that expand two-valued propositional logic (see for instance [Ga]), among them all the propositional modal systems considered in Chapter **7**.

Exercises

1. Show (informally) that PA is interpretable in ZFC. The axiom of choice is not involved so that PA is interpretable in ZF as well. More difficult is the proof that PA is interpretable in $\mathsf{ZFC}_{\text{fin}}$.

2. Prove that if T_1 is model-interpretable in T_2 then T_1 is (relatively) interpretable in some finite extension of T_2.

3. Show in detail that $Th(\mathbb{Z}, 0, 1, +, \cdot, \leqslant)$ is interpretable in $Th\mathcal{N}$.

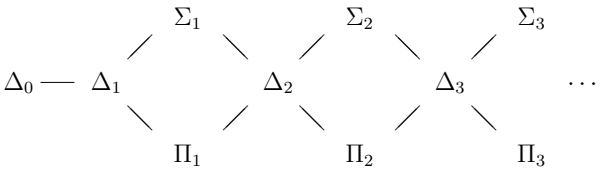
4. Prove that FPO is model-interpretable in the theory FDL of all finite distributive lattices. Thus, $Th\,$FDL is undecidable. (*Hint*: identify the points of A with the $\cap$-irreducible elements of B whose order can be given arbitrarily.)

6.7 The Arithmetical Hierarchy

Finally, we would like to add a little more on the complexity of predicates of $\mathbb{N}$, in particular, of its subsets. The set of the Gödel numbers of all sentences valid in $\mathcal{N}$ is an example of a rather simply defined nonarithmetical subset of $\mathbb{N}$; by Theorem 5.2 it has no definition in $\mathcal{L}_{ar}$.[6] However, relatively simply defined arithmetical sets and predicates may be recursion-theoretically highly complicated. It is useful to classify these according to the complexity of the defining formulas. The result is the *arithmetical hierarchy*, also called the *first-order Kleene–Mostowski hierarchy*. The following definition builds upon the one in **6.3** of the Σ_1- and Π_1-formulas and the Σ_1-, Π_1-, and Δ_1-predicates defined by these.

Definition. A Σ_{n+1}-formula is a formula of the form $\exists\vec{x}\alpha(\vec{x},\vec{y})$, where α is a Π_n-formula ($\in \mathcal{L}_{ar}$); analogously, we call $\forall\vec{x}\beta(\vec{x},\vec{y})$ a Π_{n+1}-formula if β is a Σ_n-formula. Here $\vec{x},\vec{y}$ are arbitrary tuples of variables. A Σ_n*-predicate* (resp. Π_n*-predicate*) is an arithmetical predicate P defined by a Σ_n-formula (resp. Π_n-formula). If P is both Σ_n and Π_n (i.e., a Σ_n- and Π_n-predicate) then we say that P is a Δ_n-predicate, or P *is* Δ_n for short. We denote by Σ_n, Π_n, and Δ_n the sets of the Σ_n-, Π_n- and Δ_n-predicates, respectively. In addition, $\Sigma_0 := \Pi_0 := \Delta_0$.

According to this definition, a Σ_n-formula is a prenex formula φ with n alternating blocks of quantifiers, the first of which is an $\exists$-block. φ's kernel is Δ_0. Obviously, $\Delta_n \subseteq \Sigma_n, \Pi_n$. When considering the hierarchy it is convenient to have Σ_n- and Π_n-formulas closed under equivalence in $\mathcal{N}$. Hence, we say that α *is* Σ_n or Π_n to indicate that α is equivalent to an *original* Σ_n- or Π_n-formula, respectively. Note that since $\exists\vec{x}\varphi \equiv \forall\vec{x}\varphi \equiv \varphi$ in case $var\,\vec{x} \cap var\,\varphi = \emptyset$, every Σ_n- or Π_n-formula is also both Σ_{n+1} and Π_{n+1}. Therefore $\Sigma_n, \Pi_n \subseteq \Delta_{n+1}$. This yields the following inclusion diagram, where all the inclusions, indicated by lines, are proper:



We have already come across Σ_1-, Π_1-, and Δ_1-predicates; for instance, the solvability claims of Diophantine equations are Σ_1, and the unsolvability claims are Π_1. Below we provide an example of a Π_2-predicate. It is also convenient to say that Σ_n- and Π_n-sentences define 0-*ary* Σ_n- and Π_n-predicates, respectively. In this sense the consistency of PA, for example, is a Π_1-predicate and the ω-consistency is Π_3.

[6] $Th\mathcal{N}$ is definable only in second-order arithmetic, which along with variables for numbers has variables for sets of natural numbers. Exercise 3 gives an "approximate" elementary definition.

Each formula φ is equivalent to a Σ_n- or Π_n-formula for a suitable n, for φ can be brought into prenex normal form and the quantifiers can be grouped into blocks of the same quantifiers. The hierarchy serves various purposes. More recent investigations have considered also Δ_0- or Σ_n- or Π_n-*induction*. Here the induction schema IS is restricted to the corresponding class of formulas, closed under equivalence modulo some weak base theory. An example is the theory $I\Delta_0$ mentioned on page 186.

As already shown in **6.4**, the Σ_1-predicates are the recursively enumerable ones, the Π_1-predicates their complements, and the Δ_1-predicates are exactly the recursive predicates, which are the ones whose complements are r.e. as well. Thus, we are provided with a purely recursion-theoretical way of regarding Σ_1, Π_1, and Δ_1. This underscores the importance of the arithmetical hierarchy, which is fairly stable with respect to minor changes in the definition of Δ_0. In view of Theorem 5.6 one could begin, for instance, with a Δ_0 consisting of all polynomially (or equivalently, quantifier-free) definable relations. In some presentations, a system of formulas is effectively enumerated (and denoted by Δ_0), which define exactly the p.r. predicates in $\mathcal{N}$. Section **7.1** will indicate how such a system can be defined. Between these and the Δ_0-formulas (which themselves may still be classified) lie many r.e. sets of formulas which are significant in both the theory and practice of computability, for instance, the class of elementary functions mentioned in the introduction to this chapter. However, by Remark 2 in **6.4** we know that there is no effectively enumerable system of formulas in $\mathcal{L}_{ar}$ through which all recursive, or equivalently all Δ_1-predicates, are defined, so that the definition of the arithmetical hierarchy cannot start in a feasible manner with a representative "set of Δ_1-formulas."

Remark. It should be mentioned that the first-order arithmetical hierarchy considered so far extends in a natural way to that of the the so-called *second-order arithmetic*. The latter is based on a two-sorted language with variables for natural numbers and sets of these. Also this extended hierarchy is closely related to recursion theory (see e.g. [Shoe]). A treatment lies outside the scope of this book.

Similarly to the case $n = 1$, one readily shows that a conjunction or disjunction of Σ_n-formulas is equivalent to some other Σ_n-formula; likewise for Π_n-formulas. The negation of a Σ_n-formula is equivalent to a Π_n-formula, and vice versa; this is certainly correct for $n = 1$, which initiates an easy induction on n. The complement of a Σ_n-predicate is therefore a Π_n-predicate, and vice versa. From this it easily follows that Δ_n is closed under all the mentioned operations, including negation.

By "compression of quantifiers," the idea of which was illustrated in Exercise 1 in **6.3**, one obtains a somewhat simpler presentation of the quantifier blocks. The $\exists$- and $\forall$-blocks can each be collapsed into one quantifier. This procedure is fairly easy, provided we are dealing with equivalence in $\mathcal{N}$ as is the case here, and not in a possibly too weakly axiomatized theory over $\mathcal{N}$ (infact, PA would suffice):

Theorem 7.1. *Each Σ_n-predicate is defined by a formula $\exists x_1 \forall x_2 \cdots Q_n x_n \alpha$, with α a Δ_0-formula, where Q_n is either the $\forall$- or $\exists$-quantifier, depending on whether n is even or odd. Similarly, a Π_n-predicate is defined by a formula $\forall x_1 \exists x_2 \cdots Q_n x_n \alpha$.*

Proof by (simultaneous) induction on n. Exercise 1 in **6.3** formulates the case for Σ_1- and for Π_1-predicates. Assume this is the case for n and let $\exists \vec{x} \alpha$ be the defining formula of a Σ_{n+1}-predicate, where α defines a Π_n-predicate and $\exists \vec{x}$ is a block of length $m \geqslant 1$. Using the (defining Δ_0-formula of the) pairing function, $\exists \vec{x}$ can stepwise be compressed to a single $\exists$-quantifier $\exists x$. The case $m = 0$ can also be included in the argument, using a "vacuous quantifier" $\exists x$ (i.e., $x \notin var\, \alpha$). The Π_{n+1}-formulas are treated completely analogously. One may also use the fact that both Σ_n and Π_n are closed under bounded quantification; Exercise 1. ❑

It is quite often a nontrivial task to determine a well-defined predicate's exact position in the arithmetical hierarchy, or better, like every fastidious game, it requires some training. In the example below, we consider a set that is neither recursive nor r.e. For the sake of simplicity, we apply Church's thesis in one place, although it can be eliminated using a little recursion theory as was demonstrated previously in the proof of Theorem 4.4. The example is also a good preparation for **7.5**.

Example. Let $\mathcal{L}_r$ denote the set of the $\alpha \in \mathcal{L}^1_{ar}$ that represent in Q the recursive subsets of $\mathbb{N}$. For instance, all Δ_0-formulas in $\mathcal{L}^1_{ar}$ belong to this set. Since $\mathbb{N}$ and $\emptyset$ are recursive and $\mathcal{L}^0_{ar} \subseteq \mathcal{L}^1_{ar}$, all members of $\mathsf{Q}^* := \mathsf{Q} \cup \{\alpha \mid \neg\alpha \in \mathsf{Q}\}$ also belong to $\mathcal{L}_r$, because each $\alpha \in \mathsf{Q}$ trivially represents $\mathbb{N}$, and each α with $\neg\alpha \in \mathsf{Q}$ represents $\emptyset$. Conversely, each closed formula of $\mathcal{L}_r$ belongs to Q^*. Obviously then, $\mathsf{Q}^* = \mathcal{L}_r \cap \mathcal{L}^0_{ar}$. We now show that $\mathcal{L}_r$ is arithmetical; more precisely, it is a proper Π_2-set and therefore cannot be recursively enumerable. By definition,

$$\alpha \in \mathcal{L}_r \Leftrightarrow \alpha \in \mathcal{L}^1_{ar} \;\&\; \forall n \exists \Phi [\Phi \text{ is a proof for } \alpha(\underline{n}) \text{ or for } \neg\alpha(\underline{n})].$$

This equivalence readily yields a definition of $\dot{\mathcal{L}}_r$ by a Π_2-formula $\varphi(x)$. Let the p.r. predicate '$a \in \dot{\mathcal{L}}^1_{ar}$' be Σ_1-defined by $\lambda_1(x)$. With $\mathrm{sb} = \mathrm{sb}_{\boldsymbol{v}_0}$, we then set

$$\varphi(x) := \lambda_1(x) \wedge \forall y \exists u [\mathtt{bew}_{\mathsf{Q}}(u, \mathrm{sb}(x, y)) \vee \mathtt{bew}_{\mathsf{Q}}(u, \dot{\neg}\, \mathrm{sb}(u, y))].$$

More precisely, φ should be the reduced in $\mathcal{L}_{ar}$ after eliminating the occurring p.r. function terms using more $\exists$-quantifiers inside the brackets. Thus, φ describes a Π_2-formula, that is $\dot{\mathcal{L}}_r$ is a Π_2-set. It is not Σ_1, because $\mathcal{L}_r$ is not r.e. by Remark 2 in **6.4**, nor is it Π_1. Indeed, assume this were the case; then $\mathsf{Q}^* = \mathcal{L}_r \cap \mathcal{L}^0_{ar}$ would also be Π_1, for $\mathcal{L}^0_{ar}$ is Δ_1. Now Q^* is certainly r.e. and thus Σ_1, and so by Theorem 4.5 Q^* would be recursive. But then we obtain a decision procedure for Q (hence a contradiction) as follows: let $\alpha \in \mathcal{L}^0_{ar}$ be given. If $\alpha \notin \mathsf{Q}^*$ then also $\alpha \notin \mathsf{Q}$; if $\alpha \in \mathsf{Q}^*$, we turn on the enumeration machine for Q and wait until either α or $\neg\alpha$ appears, the former instance of which corresponds to the case $\alpha \in \mathsf{Q}$.

We end this section with a result useful for our purposes in **7.1**. It will be proved that the Σ_1-predicates are definable without refer to Δ_0, using special Σ_1-formulas. To this end somewhat stronger axioms are considered than those of Q, namely the axioms of the theory N presented in **6.3**. All these axioms are provable in PA.

Definition. *Special* Σ_1-formulas are defined as follows:

(a) $\mathsf{S}x{=}y$, $x+y{=}z$, and $x\cdot y{=}z$ are special Σ_1-formulas, where x, y, z denote distinct variables (the special prime formula condition);

(b) if α, β are special Σ_1-formulas then so too are $\alpha\wedge\beta$, $\alpha\vee\beta$, $\alpha\frac{0}{x}$, and $\alpha\frac{y}{x}$, where x, y are distinct and not in $bnd\,\alpha$ (*prime-term substitution*), as well as $\exists x\alpha$ and $(\forall x{<}y)\alpha$ for $y\notin var\,\alpha$.

Theorem 7.2. *Every original* Σ_1*-formula is equivalent to a special* Σ_1*-formula in the theory* N*, thus in* PA *and a fortiori in the standard model* $\mathcal{N}$*.*

Proof. It suffices to verify the claim for all Δ_0-formulas, since the set of special Σ_1-formulas is closed under $\exists$-quantification. Since $s{=}t \equiv \exists x(x{=}s\wedge x{=}t)$ with $x\notin var\,s,t$, it is enough to consider prime formulas of the form $x{=}t$. For prime terms t this clearly follows from $x{=}0 \equiv (x{=}y)\frac{0}{y}$ and $x{=}y \equiv_{\mathsf{N}} (x+z{=}y)\frac{0}{z}$, and the induction steps on the operations $\mathsf{S}, +, \cdot$ follow from $x{=}\mathsf{S}t \equiv \exists y(x{=}\mathsf{S}y\wedge y{=}t)$, $x{=}s+t \equiv \exists y\exists z(x{=}y+z\wedge y{=}s\wedge z{=}t)$, and similarly for $\cdot$.

The claim holds for all literals because of $s\neq t \equiv \exists y\exists z(x\neq y\wedge x{=}s\wedge y{=}t)$, and $x\neq y \equiv_{\mathsf{N}} \exists u\exists z(\mathsf{S}u{=}z\wedge(x+z{=}y\vee y+z{=}x))$. By Exercise 4 in **6.3** we need only carry out induction on $\wedge, \vee, (\forall x{\leqslant}t)$ and $(\exists x{\leqslant}t)$. For $\wedge, \vee$ this is clear. For the remainder note that $(\forall x{\leqslant}t)\alpha$ and $(\exists x{\leqslant}t)\alpha$ are N-equivalent respectively to $\exists y(y{=}t\wedge(\forall x{<}y)\alpha\wedge\alpha\frac{y}{x})$ and $\exists x\exists y\exists z(x+y{=}z\wedge z{=}t\wedge\alpha)$. ❑

Exercises

1. Show that Σ_n, Π_n, and hence Δ_n are closed under bounded quantification.[7]

2. Confirm that $\Delta_0\subset\Delta_1\subset\Sigma_1, \Pi_1$, which therefore shows that these four classes of arithmetical predicates are distinct.

3. Let $Tr_n = \{\alpha\in\mathcal{L}^0_{ar} \mid \mathcal{N}\vDash\alpha \;\&\; \mathrm{qr}\,\alpha\leqslant n\}$, so that $Th\mathcal{N} = \bigcup_{n\in\mathbb{N}} Tr_n$. By Theorem 5.2, $Th\mathcal{N}$ is itself not arithmetical. Prove that Tr_n is arithmetical, more precisely, Tr_n is Δ_{n+1}. In this sense, $Th\mathcal{N}$ is *arithmetically approximable*. (With more effort it can be shown that Tr_n is at most Δ_n).

4. Prove that ω-inconsistency is Σ_3. Theorem 7.5.2 will show that this property is properly Σ_3.

[7] Again, we do not need here that the Σ_n-formulas are closed under $\equiv_{\mathcal{N}}$; it would be sufficient if they are closed under $\equiv_{\mathsf{PA}}$.

Chapter 7

On the Theory of Self-Reference

By self-reference we basically mean the possibility of talking inside a theory T about T itself or related theories. Here we can give merely a glimpse into this recently much advanced area of research. We will prove Gödel's second incompleteness theorem, Löb's theorem and many other results connected with self-reference, while further results are discussed only and elucidated by means of applications. All this is of great interest both for epistemology and the foundations of mathematics.

The mountain we first have to climb is the proof of the so-called derivability conditions for PA and other theories in Section **7.1**. But anyone contented with leafing through **7.1** can begin straight away in **7.2**; from then on we will just be reaping in the fruits of our labor. However, one would forgo a real adventure in doing so, namely the fusion of logic and elementary number theory in the metatheoretical analysis of PA. Who wants to attain a comprehensive understanding of self-reference, should study the material in **7.1** anyway.

Gödel himself tried to interpret the notion "provable" using a modal operator in the framework of the modal system S4. This attempt reflects some of his own results, though not adequately. Only after 1970, when modal logic was sufficiently advanced, could such a program be successfully carried out. A suitable instrument turned out to be the modal logic denoted by G or GL. The Kripke semantics treated in **7.3** is an excellent tool for confirming or refuting self-referential statements as demonstrated in **7.4**. Solovay's completeness theorem, and the completeness theorem of Kripke semantics for G are fortunately of the kind that allows application without knowing the completeness proofs itself (which contain quite a number of technical tricks).

There are several extensions of G important for the analysis of other proof operators or a comparison of these, for example, the bimodal logic in **7.5**. A comprehensive survey can be found in [Bu, Chapter VII], see also [Vi]. In **7.6** we discuss some questions regarding self-reference in axiomatic set theory.

7.1 The Derivability Conditions

Put somewhat simply, Gödel's second incompleteness theorem states that $\vdash_T \mathtt{Con}_T$ cannot hold for a sufficiently strong and consistent axiomatizable theory T. Here $\mathtt{Con}_T$ is a sentence in the language $\mathcal{L}$ of T expressing the consistency of T. In a popular formulation: *If the theory T is consistent, then its consistency is unprovable in T.* As was outlined by Gödel and will be verified in this chapter, the italicized sentence is not only true but even provable in the framework of T.

The easiest way to obtain Gödel's theorem is first to prove the *derivability conditions*, stated below. These conditions deserve some interest on their own. Their formulation supposes the arithmetizability of T, which includes the distinguishing of a sequence $\underline{0}, \underline{1}, \ldots$ of ground terms; see page 194. Let $\mathtt{bew}_T(y, x)$ be a formula that is assumed to represent the recursive predicate *bew*$_T$ in T, exactly as in **6.4**. For $\mathtt{bwb}_T(x) = \exists y\, \mathtt{bew}_T(y, x)$ we write $\Box(x)$, and $\Box\alpha$ is to mean $\mathtt{bwb}_T(\ulcorner \alpha \urcorner)$. We may read $\Box\alpha$ as "box α" or more suggestively "α is provable in T," because it formalizes the property $\vdash_T \alpha$ *within* T. If $\Box$ refers to some theory $T' \neq T$ then $\Box$ has to be indexed correspondingly. For instance, $\Box_{\mathsf{ZFC}}\varphi$ for $\varphi \in \mathcal{L}_\in$ can easily expressed also in PA. Note that $\Box\alpha$ is always a sentence, even if α contains free variables.

Further, set $\Diamond\alpha := \neg\Box\neg\alpha$ for $\alpha \in \mathcal{L}$. If α is a sentence, $\Diamond\alpha$ may be read as α *is compatible with* T, because it formalizes '$\nvdash_T \neg\alpha$' which is, as we know, equivalent to the consistency of $T + \alpha$. First of all, we define $\mathtt{Con}_T$ in a natural way by

$$\mathtt{Con}_T := \neg\Box\bot \,\big(= \neg\, \mathtt{bwb}_T(\ulcorner \bot \urcorner)\big),$$

where $\bot$ is a contradiction, $0 \neq 0$ for instance. We shall see in a moment that $\mathtt{Con}_T$ is independent modulo T of the choice of $\bot$. The mentioned derivability conditions then read as follows:

$$D1\colon\ \vdash_T \alpha \ \Rightarrow \vdash_T \Box\alpha, \quad D2\colon\ \vdash_T \Box\alpha \wedge \Box(\alpha \to \beta) \to \Box\beta, \quad D3\colon\ \vdash_T \Box\alpha \to \Box\Box\alpha.$$

Here α, β run through all sentences of $\mathcal{L}$. Sometimes $D2$ is written in the equivalent form $\Box(\alpha \to \beta) \vdash_T \Box\alpha \to \Box\beta$, and $D3$ as $\Box\alpha \vdash_T \Box\Box\alpha$. These conditions are due to Löb, but they were considered in a slightly different setting already in [HB].

A consequence of $D1$ and $D2$ is $D0\colon\ \alpha \vdash_T \beta \ \Rightarrow\ \Box\alpha \vdash_T \Box\beta$. This implication holds since $\alpha \vdash_T \beta \ \Rightarrow \vdash_T \alpha \to \beta \ \Rightarrow \vdash_T \Box(\alpha \to \beta) \ \Rightarrow \vdash_T \Box\alpha \to \Box\beta$. From $D0$ it clearly follows that $\alpha \equiv_T \beta \ \Rightarrow\ \Box\alpha \equiv_T \Box\beta$. In particular, the choice of $\bot$ in $\mathtt{Con}_T$ is arbitrary as long as $\bot \equiv_T 0 \neq 0$.

Remark 1. Any operator $\partial\colon \mathcal{L} \to \mathcal{L}$ satisfying the conditions $d1\colon\ \vdash_T \alpha \ \Rightarrow \vdash_T \partial\alpha$ and $d2\colon\ \partial(\alpha \to \beta) \vdash_T \partial\alpha \to \partial\beta$ thus satisfies $d0\colon\ \alpha \vdash_T \beta \ \Rightarrow\ \partial\alpha \vdash_T \partial\beta$. It also satisfies $d\wedge\colon\ \partial(\alpha \wedge \beta) \equiv_T \partial\alpha \wedge \partial\beta$, since $\alpha \wedge \beta \vdash_T \alpha, \beta$, and by $d0$, $\partial(\alpha \wedge \beta) \vdash_T \partial\alpha, \partial\beta \vdash_T \partial\alpha \wedge \partial\beta$. Similarly, $\partial\alpha \wedge \partial\beta \vdash_T \partial(\alpha \wedge \beta)$ readily follows from $\alpha \vdash_T \beta \to \alpha \wedge \beta$ by first applying $d0$ and then $d2$. Clearly, $d0$ implies $d00\colon\ \alpha \equiv_T \beta \Rightarrow \partial\alpha \equiv_T \partial\beta$, for all $\alpha, \beta \in \mathcal{L}$.

Whereas $D2$ and $D3$ represent sentence schemas in T, $D1$ is of metatheoretical nature and follows obviously from the representability of bew_T in T. Thus, $D1$ holds even for weak theories such as $T = \mathsf{Q}$. On the other hand, the converse of $D1$,

$D1^*$: $\vdash_T \Box\alpha \;\Rightarrow\; \vdash_T \alpha$, for all $\alpha \in \mathcal{L}^0$,

may fail. Fortunately, it holds for all ω-consistent axiomatic extensions $T \supseteq \mathsf{Q}$, for instance $T = \mathsf{PA}$. Indeed, $\nvdash_T \alpha$ implies $\vdash_T \neg\, \mathtt{bew}_T(\underline{n}, \ulcorner\alpha\urcorner)$ for all n (Theorem 6.4.2). Hence, $\nvdash_T \exists y\, \mathtt{bew}_T(y, \ulcorner\alpha\urcorner) = \Box\alpha$ in view of the ω-consistency of T.

Unlike $D1$, the properties $D2$ and $D3$ are not so easily obtained. T must be able to speak directly or indirectly (via arithmetization) about provability in T. Note that $D3$ is nothing else than *condition* $D1$ *formalized within* T, while $D2$ formalizes (7) from page 178, meaning the closure under MP in arithmetical terms. Let us first show that $D2$ holds provided it has been shown that

(1) $\quad \mathtt{bew}_T(u, x) \wedge \mathtt{bew}_T(v, x \mathbin{\tilde{\to}} y) \vdash_T \mathtt{bew}_T(u * v * \langle y\rangle, y)$,

where the p.r. functions $\mathbin{\tilde{\to}}$, $*$, and $y \mapsto \langle y\rangle$ appearing in (1) must of course be defined in T. Generally speaking, $f \in \mathbf{F}_n$ is called *definable in* T (with respect to a given sequence of terms $(\underline{n})_{n\in\mathbb{N}}$) if there is a formula $\delta(\vec{x}, y) \in \mathcal{L}$ such that

(Δ) (a) $\vdash_T \delta(\underline{\vec{a}}, \underline{f\vec{a}})$ for all $\vec{a}$, (b) $\vdash_T \forall\vec{x}\exists!y\delta(\vec{x}, y)$.

Clearly, f is then also represented by $\delta(\vec{x}, y)$. Because of (b), f may explicitly be defined in T by $\delta(\vec{x}, y)$ (see **2.6**). We will no longer distinguish between T and its definitorial extensions and write simply $\vdash_T y = f\vec{x} \leftrightarrow \delta(\vec{x}, y)$. This and (a) easily yield $\vdash_T f\underline{\vec{a}} = \underline{f\vec{a}}$, for instance $\vdash_T \underline{a \mathbin{\tilde{\to}} b} = \underline{a} \mathbin{\tilde{\to}} \underline{b}$. With $\ulcorner\alpha\urcorner$, $\ulcorner\beta\urcorner$ for x, y, we then obtain from (1) in view of $\ulcorner\alpha \to \beta\urcorner = \underline{\dot\alpha \mathbin{\tilde{\to}} \dot\beta} = \underline{\dot\alpha} \mathbin{\tilde{\to}} \underline{\dot\beta} = \ulcorner\alpha\urcorner \mathbin{\tilde{\to}} \ulcorner\beta\urcorner$,

$$\mathtt{bew}_T(u, \ulcorner\alpha\urcorner) \wedge \mathtt{bew}_T(v, \ulcorner\alpha \to \beta\urcorner) \vdash_T \mathtt{bew}_T(u * v * \langle\ulcorner\beta\urcorner\rangle, \ulcorner\beta\urcorner).$$

Particularization yields $D2$. But the real work, the proof of (1), still lies ahead.

In order to better keep track of things, we restrict our considerations to the theories ZFC and PA, which are of central interest in nearly all foundational questions. ZFC is only briefly discussed. Here the proofs of $D2$ and $D3$ are much easier than in PA. Indeed, (1) and hence $D2$ are clear, because the naive proof of (1) above with $bew_T = bew_{\mathsf{ZFC}}$ can easily be formalized *inside* ZFC. This includes the definability of all functions occurring in (1), for we *did* define them; for instance, the operation $*$ on page 174 (set $a*b = \emptyset$ if not $a, b \in \omega$). We arithmetize $\mathcal{L}_\in$ according to the pattern in **6.2**, encoding formulas as in **6.2** based on prime number factorization,[1] so that $\mathcal{L}_\in$-formulas are encoded within ZFC by certain ω-terms, defined in **3.5**. $\mathcal{L}_{ar}$-formulas are identified with their ω-relativized in $\mathcal{L}_\in$, called the *arithmetical formulas* of $\mathcal{L}_\in$. Moreover, the arithmetical predicate bew_{ZFC} is certainly representable in ZFC by

[1] This is not actually necessary, since in ZFC one can talk directly about finite sequences and hence about $\mathcal{L}_\in$-formulas, but we do so in order to maintain coherence with the exposition in **6.2**.

Theorem 6.4.2, since this theorem can be viewed, just like every other theorem in this book, as a theorem *within* ZFC. Thus, the naive proof of $D1$ based on this theorem (up to Corollary 6.4.3) can as a whole be carried out in ZFC, and so $D3$ is proved. Roughly speaking, $D2$ and $D3$ hold for ZFC because ordinary mathematics, in particular the material in Chapter **6**, is formalizable in ZFC.

In all of the above, no typically set-theoretical constructions like ordinal recursion are needed. Only relatively simple combinatorial facts are required. Hence there is some hope that the proofs of D2 and D3 can also be carried out in sufficiently strong arithmetical theories like PA. This is indeed so and such a result is considerably more interesting for a critical foundation of mathematics. We already encoded $\mathcal{L}_{ar}$-formulas and proofs within PA by their corresponding Gödel terms on page 191. However, it is not obvious how to define in PA the functions appearing in (1) and other relevant functions.[2] Hence, our first goal is to show that all occurring p.r. functions are provably recursive in the following sense, which considerably strengthens the definition (Δ) from the previous page:

Definition. An n-ary recursive function f is called Σ_1-*definable in* PA, or *provably recursive*, if there is a Σ_1-formula $\delta_f(\vec{x}, y)$ such that

(2) (a) $\vdash_{\mathsf{PA}} \delta_f(\underline{\vec{a}}, \underline{f\vec{a}})$ for all $\vec{a} \in \mathbb{N}^n$; (b) $\vdash_{\mathsf{PA}} \forall \vec{x}\, \exists! y \delta_f(\vec{x}, y)$.

Because of the Σ_1-completeness of PA, (a) is equivalent to $\mathcal{N} \vDash \delta_f(\underline{\vec{a}}, \underline{f\vec{a}})$ for all $\vec{a}$. We will prove stepwise that *all* p.r. functions are Σ_1-definable, and derive also the recursion equations belonging to them in PA. Thereafter we may treat all occurring p.r. functions in PA as if they had been available in the language right from the outset. Essentially this fundamental fact allows a treatment of elementary number theory and combinatorics within the boundaries of PA. $D3$ demands additional preparation, and even good textbooks do not carry out all of the proof steps. However, all steps described here and not handled in detail can easily be completed in full by the sufficiently assiduous reader. Life could be made easier through the mutual interpretability of PA and $\mathsf{ZFC}_{\mathrm{fin}}$, though this is itself not easy to prove.

The Σ_1-definability in PA of some functions, including the $\boldsymbol{\beta}$-function, is straightforwardly verified; see Exercise 1. But in order to recognize as legitimate in PA, for instance, the definition of the exponential function by δ_{exp} in Remark 1 from **6.4**, Lemma 6.4.1, and hence also Euclid's lemma and the Chinese remainder theorem have to be proved *within* PA. As regards Euclid's lemma, there is no problem. Just follow the proof in **6.4**. Clearly, some basic arithmetical laws are applied that must be proven first, including those on the difference $a - b$ for $a \geqslant b$.

[2] In [Go2], Gödel presented a list of 45 p.r. functions, of which the last was χ_{bew}. Following [WR], he considered a kind of higher-order arithmetical theory. That Gödel's theorems also hold in first-order arithmetic was probably first noticed in [HB].

As for the Chinese remainder theorem, at present even its formalizability in $\mathcal{L}_{ar}$ is not evident, because we quantify over finite sequences which can take place in PA only *after* it has been shown that PA is capable of talking about such sequences. In order to surmount this obstacle, we use c, d to denote for the time being unary provably recursive functions, which may depend on further parameters. Each such c determines for given n the sequence $\mathsf{c}_0, \ldots, \mathsf{c}_n$, with $\mathsf{c}_\nu = \mathsf{c}(\nu)$ for $\nu \leqslant n$. With the Δ_0-definable relation $\perp$ of coprimeness, the Chinese remainder theorem can provisionally be stated as follows: for arbitrary c, d holds[3]

(3) $\quad \vdash_{\mathsf{PA}} \forall n[(\forall \nu, i, j \leqslant n)(\mathsf{c}_\nu < \mathsf{d}_\nu \wedge (i \neq j \to \mathsf{d}_i \perp \mathsf{d}_j)) \to \exists a(\forall \nu \leqslant n)\, \mathrm{rem}(a : \mathsf{d}_\nu) = \mathsf{c}_\nu]$.

To convert the original proof of the remainder theorem to one for (3) we require, for given provably recursive d, the term $\mathrm{lcm}\{\mathsf{d}_\nu | \, \nu \leqslant n\}$, the least common multiple of $\mathsf{d}_0, \ldots, \mathsf{d}_n$. Then $f : n \mapsto \mathrm{lcm}\{\mathsf{d}_\nu | \, \nu \leqslant n\}$ is defined in PA by the Σ_1-formula

$$\delta_f(x, y) := (\forall \nu \leqslant x)\mathsf{d}_\nu \,|\, y \wedge (\forall z < y)(\exists \nu \leqslant x)\, \mathsf{d}_\nu \not| \, z.$$

More precisely, $\delta_f(x, y)$ *describes* a Σ_1-formula in the original language, similarly as does δ_{exp} on page 190. Since $\mathcal{N} \vDash \delta(\underline{n}, \underline{\mathrm{lcm}\{\mathsf{d}_\nu | \, \nu \leqslant n\}})$ for all n, 2(a) holds. With the least-number schema (see Exercise 3 in **3.3**) applied to $\beta(x, y) := (\forall \nu \leqslant x)\mathsf{d}_\nu \,|\, y$, we obtain $\vdash_{\mathsf{PA}} \exists! y \delta_f(x, y)$, provided it has been shown that $\vdash_{\mathsf{PA}} \exists y \beta(x, y)$ ('finitely many numbers have a common multiple'). This follows by induction on x. Clearly $\vdash_{\mathsf{PA}} \exists y \beta(0, y)$, and the induction step has already been carried out in Example 1 from **2.5**. We then obtain the proof of (3) by following the proof of the remainder theorem in **6.2**, and, writing $\boldsymbol{\beta} st$ for $\boldsymbol{\beta}(s, t)$, a suitable version of Lemma 6.4.1 about the the basic property of the $\boldsymbol{\beta}$-function:

(4) $\quad \vdash_{\mathsf{PA}} \forall v \exists u (\forall \nu \leqslant v)\, \mathsf{c}_\nu = \boldsymbol{\beta} u \nu$, for every provably recursive c.

Theorem 1.1. *Each p.r. function f is provably recursive. Moreover, the recursion equations for f are provable in* PA *whenever $f = \boldsymbol{Op}(g, h)$.*

Proof. For the initial functions and $+, \cdot$ the formulas $\boldsymbol{v}_0 = 0$, $\boldsymbol{v}_1 = \mathsf{S}\boldsymbol{v}_0$, $\boldsymbol{v}_n = \boldsymbol{v}_\nu$ along with $\boldsymbol{v}_2 = \boldsymbol{v}_0 + \boldsymbol{v}_1$ and $\boldsymbol{v}_2 = \boldsymbol{v}_0 \cdot \boldsymbol{v}_1$ are defining Σ_1-formulas. (2) is here obvious. For $f = h[g_1, \ldots, g_m]$, let $\delta_f(\vec{x}, y)$ be $y = h(g_1\vec{x}, \ldots, g_m\vec{x})$. In this case (2) is clear, because we might think of the symbols $h, g_1, \ldots, g_m$ as having already been introduced via explicit definition, so that the last formula simply belongs to the language. Only the definition of δ_f for $f = \boldsymbol{Op}(g, h)$ requires some skill. In noting that formula `beta` in **6.4** is just what we require concerning the $\boldsymbol{\beta}$-function, let

(5) $\quad \delta_f(\vec{x}, y, z) := \exists u [\underbrace{\boldsymbol{\beta} u 0 = g\vec{x} \wedge (\forall v < y)\boldsymbol{\beta} u \mathsf{S} v = h(\vec{x}, v, \boldsymbol{\beta} u v) \wedge \boldsymbol{\beta} u y = z}_{\gamma(u, \vec{x}, y, z)}]$.

δ_f is Σ_1. Metainduction over b shows that $\mathcal{N} \vDash \delta_f(\underline{\vec{a}}, \underline{b}, \underline{f\vec{a}})$ and hence 2(a). Uniqueness in 2(b) follows with a glance at (5) from $\vdash_{\mathsf{PA}} \gamma(u, \vec{x}, y, z) \wedge \gamma(u', \vec{x}, y, z') \to z = z'$,

[3] For suggestive reasons from now on also letters such as $n, \nu, \ldots$ denote variables in $\mathcal{L}_{ar}$.

obtained using induction on y. Also, $\vdash_{\mathsf{PA}} \exists z \delta_f(\vec{x}, y, z)$ is shown by induction on y. For $y = 0$ consider $\vdash_{\mathsf{PA}} \exists u \boldsymbol{\beta} u 0 = g\vec{x}$ according to (4). Choose there, for instance, the provably recursive $\mathsf{c} : v \to w$ defined by $v = 0 \wedge w = g\vec{x} \vee v \neq 0 \wedge w = 0$. The less simple inductive step $(*)$: $\exists z \delta_f(\vec{x}, y, z) \vdash_{\mathsf{PA}} \exists z' \delta_f(\vec{x}, \mathtt{S}y, z')$ will be verified informally: Suppose $\exists z \gamma(u, \vec{x}, y, z)$, or equivalently, $\gamma(u, \vec{x}, y, \boldsymbol{\beta} uy)$. Then the Σ_1-formula

$$\varphi(v, w, u, \vec{x}, y) := v \neq \mathtt{S}y \wedge w = \boldsymbol{\beta} uv \vee v = \mathtt{S}y \wedge w = h(\vec{x}, v, \boldsymbol{\beta} uy)$$

defines in PA a function $\mathsf{c} : v \mapsto w$ with parameters $u, \vec{x}, y$. So by (4) (taking $\mathtt{S}y$ for v) there is some u' with $\boldsymbol{\beta} u' v = \mathsf{c}_v = \boldsymbol{\beta} uv$ for all $v \leqslant y$ and $\boldsymbol{\beta} u' \mathtt{S}y = h(\vec{x}, y, \boldsymbol{\beta} uy)$. With this u' and $z' = \boldsymbol{\beta} u' \mathtt{S}y$ we obtain $\gamma(u', \vec{x}, \mathtt{S}y, z')$ and therefore $\exists z' \delta_f(\vec{x}, \mathtt{S}y, z')$. This proves $(*)$ and hence 2(b). We finally also verify that

$$\text{(a)} \vdash_{\mathsf{PA}} f(\vec{x}, 0) = g\vec{x}, \quad \text{(b)} \vdash_{\mathsf{PA}} f(\vec{x}, \mathtt{S}y) = h(\vec{x}, y, f(\vec{x}, y)).$$

(a) follows from 2(b) since (5) readily yields $\vdash_{\mathsf{PA}} \delta_f(\vec{x}, 0, g\vec{x})$. For (b) show first that $\gamma(u, \vec{x}, y, z) \vdash_{\mathsf{PA}} (\forall v \leqslant y) f(\vec{x}, \mathtt{S}v) = \boldsymbol{\beta} u \mathtt{S}v$ using induction on y. From this one easily infers $\gamma(u, \vec{x}, y, z) \vdash_{\mathsf{PA}} \varphi := (\forall v \leqslant y) f(\vec{x}, \mathtt{S}v) = h(\vec{x}, y, f(\vec{x}, y))$. Now, because of $\vdash_{\mathsf{PA}} \exists z \gamma(u, \vec{x}, y, z)$, we obtain $\vdash_{\mathsf{PA}} \varphi$, which obviously includes (b). ❑

We thus have achieved our first goal. Now the properties of $*, \ell, \ldots$ stated in the remark on page 178 along with the basic properties of bew_{PA} and bwb_{PA} stated on the same page are also easily proved *within* PA. This is a little extra program that includes the proof of the unique prime factorization; Exercise 3. Thus, (1) is indeed provable for $T = \mathsf{PA}$. It implies $D2$ and, with $\Box = \mathtt{bwb}_{\mathsf{PA}}$, moreover

(6) $\Box(x \dot{\to} y) \vdash_{\mathsf{PA}} \Box(x) \to \Box(y)$.[4]

Remark 2. The formalized equations of Exercise 4 in **6.4** are now also seen to be provable in PA. For instance, (b) now reads $\vdash_{\mathsf{PA}} \mathrm{sb}_{\vec{x}}(\ulcorner \varphi \urcorner, \vec{x}) = \mathrm{sb}_{\vec{x}'}(\ulcorner \varphi \urcorner, \vec{x}')$ for $\varphi = \varphi(\vec{x})$, where $\vec{x}' \subseteq \vec{x}$ enumerates the free variables of φ and may be empty. To prove item (c), consider a special case. Let φ be $\mathtt{S}x = y$. Then $\mathrm{sb}_{xy}(\dot{\varphi}, x, \mathtt{S}x) = \mathrm{sb}_x((\varphi \frac{\mathtt{S}x}{y})^{\cdot}, x)$, formalized: $\mathrm{sb}_{xy}(\ulcorner \varphi \urcorner, x, y) \frac{\mathtt{S}x}{y} = \mathrm{sb}_x(\ulcorner \varphi \frac{\mathtt{S}x}{y} \urcorner, x)$. Following the example on page 193, one requires for the proof of this equation in PA just $\vdash_{\mathsf{PA}} \mathrm{cf}\, \mathtt{S}x = \tilde{\mathtt{S}}\, \mathrm{cf}\, x$, which holds by Theorem 1.1.

Now we are suitably equipped to prove $D3$. We first generalize the notation $\Box\varphi$.

Definition. For $\varphi = \varphi(\vec{x})$ let $\Box[\varphi] := \Box(\mathrm{sb}_{\vec{x}}(\ulcorner \varphi \urcorner, \vec{x}))$ $\big(= \mathtt{bwb}_{\mathsf{PA}} \frac{\mathrm{sb}_{\vec{x}}(\ulcorner \varphi \urcorner, \vec{x})}{x}\big)$.

By Remark 2, $\vdash_{\mathsf{PA}} \mathrm{sb}_{\vec{x}}(\ulcorner \varphi \urcorner, \vec{x}) = \mathrm{sb}_{\vec{x}'}(\ulcorner \varphi \urcorner, \vec{x}')$ whenever $var\, \vec{x}' = free\, \varphi$. Therefore, we may assume w.l.o.g. $free \Box[\varphi] = free\, \varphi$. Moreover, for $\alpha \in \mathcal{L}^0_{ar}$ we may identify $\Box[\alpha]$ and $\Box\alpha$, because $\vdash_{\mathsf{PA}} \mathrm{sb}_{\vec{x}}(\ulcorner \alpha \urcorner, \vec{x}) = \mathrm{sb}_{\emptyset}(\ulcorner \alpha \urcorner) = \ulcorner \alpha \urcorner$. By $\vdash_{\mathsf{PA}} \forall \vec{x} \Box[\varphi]$, the schema '$\vdash_{\mathsf{PA}} \varphi(\underline{\vec{a}})$ for all $\vec{a} \in \mathbb{N}^n$' is expressed in PA as a single sentence. $\vdash_{\mathsf{PA}} \forall \vec{x} \Box[\varphi]$ reflects in PA the existence of a *collection of proofs* which, due to the ω-incompleteness of PA, can be less than $\vdash_{\mathsf{PA}} \Box\varphi$ (or equivalently, $\vdash_{\mathsf{PA}} \Box \forall \vec{x} \varphi$).

[4] $\Box$ may even denote $\mathtt{bwb}_T$ for *any* axiomatizable and arithmetizable theory T.

Example. Let $\varphi(x,y)$ be $\mathtt{S}x\mathbf{=}y$. We prove $\varphi \vdash_{\mathsf{PA}} \Box[\varphi]$, or equivalently, $\vdash_{\mathsf{PA}} \Box[\varphi]\,\frac{\mathtt{S}x}{y}$, where w.l.o.g. x, y are not bounded in $\Box(x)$. In view of Remark 2, we then obtain $\Box[\varphi]\,\frac{\mathtt{S}x}{y} = \Box(\mathrm{sb}_{xy}(\ulcorner\varphi\urcorner, x, y))\,\frac{\mathtt{S}x}{y} = \Box(\mathrm{sb}_{xy}(\ulcorner\varphi\urcorner, x, \mathtt{S}y)) \equiv_{\mathsf{PA}} \Box(\mathrm{sb}_x(\ulcorner\varphi\,\frac{\mathtt{S}x}{y}\urcorner, x))$. Thus, $\Box[\varphi]\,\frac{\mathtt{S}x}{y} \equiv_{\mathsf{PA}} \Box[\mathtt{S}x\mathbf{=}\mathtt{S}x]$. Hence, it suffices to verify $\vdash_{\mathsf{PA}} \Box[\mathtt{S}x\mathbf{=}\mathtt{S}x]$. This reflects in PA 'for arbitrary n, $\vdash_{\mathsf{PA}} \mathtt{S}\underline{n}\mathbf{=}\mathtt{S}\underline{n}$'. Let $\alpha(x) := \mathtt{S}x\mathbf{=}\mathtt{S}x$. We prove $\vdash_{\mathsf{PA}} \Box[\alpha]$ in detail. Consider the p.r. function $\tilde{\alpha}\colon n \mapsto \mathrm{sb}_x(\dot{\alpha}, n)$ (the Gödel number of $\alpha(\underline{n})$). By axiom $\Lambda 9$, $\langle\tilde{\alpha}(n)\rangle$ is for each n a simple arithmetized proof of length 1. Stated within PA, $\vdash_{\mathsf{PA}} \mathtt{bew}_{\mathsf{PA}}(\langle\tilde{\alpha}(x)\rangle, \tilde{\alpha}(x))$, which yields $\vdash_{\mathsf{PA}} \exists y\,\mathtt{bew}_{\mathsf{PA}}(y, \tilde{\alpha}(x)) = \Box(\tilde{\alpha}(x)) = \Box[\alpha]$.

The following generalizations of $D1, D2$ for $\alpha = \alpha(\vec{x})$ and $\beta = \beta(\vec{x})$ hold:

(7) (a) $\vdash_{\mathsf{PA}} \alpha \;\Rightarrow \vdash_{\mathsf{PA}} \Box[\alpha]$; (b) $\Box[\alpha \to \beta] \vdash_{\mathsf{PA}} \Box[\alpha] \to \Box[\beta]$.

To see this let $\vdash_{\mathsf{PA}} \alpha$, so that also $\vdash_{\mathsf{PA}} \Box\alpha$. Just as in the above example, a proof for α provides one for $\alpha_{\vec{x}}(\underline{\vec{a}})$ for every $\vec{a} \in \mathbb{N}^n$ in a p.r. way, or stated *within* PA: $\vdash_{\mathsf{PA}} \Box(u) \to \Box(\mathrm{sb}_{\vec{x}}(u, \vec{x}))$. Thus, choosing $\ulcorner\alpha\urcorner$ for u, $\vdash_{\mathsf{PA}} \Box(\mathrm{sb}_{\vec{x}}(\ulcorner\alpha\urcorner, \vec{x}))$ $(= \Box[\alpha])$. (b) follows from (6) with $\mathrm{sb}_{\vec{x}}(\ulcorner\alpha\urcorner, \vec{x}), \mathrm{sb}_{\vec{x}}(\ulcorner\beta\urcorner, \vec{x})$ for x, y, taking into account that $\vdash_{\mathsf{PA}} \mathrm{sb}_{\vec{x}}(\ulcorner\alpha \to \beta\urcorner, \vec{x})\mathbf{=} \mathrm{sb}_{\vec{x}}(\ulcorner\alpha\urcorner, \vec{x}) \mathbin{\tilde{\to}} \mathrm{sb}_{\vec{x}}(\ulcorner\beta\urcorner, \vec{x})$ (Exercise 3 in **6.4**). Additionally, item (c) of this exercise, provable in PA, yields

(8) $\Box[\alpha]\,\frac{t}{x} \equiv_{\mathsf{PA}} \Box[\alpha\,\frac{t}{x}\,]$ $(t \in \{0, y, \mathtt{S}y\}$ and $y \notin \mathit{bnd}\,\alpha)$.

Obviously, $D3$ is only a special case of the *provable* Σ_1*-completeness* of PA:

(9) $\varphi \vdash_{\mathsf{PA}} \Box[\varphi]$ (equivalently, $\vdash_{\mathsf{PA}} \varphi \to \Box[\varphi]$), for all Σ_1-formulas φ.

To make $D3$ evident, choose in (9) for φ the Σ_1-sentence $\Box\alpha$, for any given $\alpha \in \mathcal{L}^0_{ar}$. It follows that $\Box\alpha \vdash_{\mathsf{PA}} \Box[\Box\alpha] \equiv \Box\Box\alpha$, and $D3$ is proved. We obtain (9) by applying the following theorem, because the operator $\partial\colon \alpha \mapsto \Box[\alpha]$ satisfies the conditions of the theorem by (7), (8), and because $\mathit{free}\,\alpha = \mathit{free}\,\Box[\alpha]$ may be assumed.

Theorem 1.2. *Let $\partial\colon \mathcal{L}_{ar} \to \mathcal{L}_{ar}$ be any operator with $\mathit{free}\,\partial\alpha \subseteq \mathit{free}\,\alpha$ satisfying*

$d1$: $\vdash_{\mathsf{PA}} \alpha \;\Rightarrow \vdash_{\mathsf{PA}} \partial\alpha$,

$d2$: $\partial(\alpha \to \beta) \vdash_{\mathsf{PA}} \partial\alpha \to \partial\beta$,

ds: $\partial\alpha\,\frac{t}{x} \equiv_{\mathsf{PA}} \partial(\alpha\,\frac{t}{x}\,)$ $(t \in \{0, y, \mathtt{S}y\},\ y \notin \mathit{bnd}\,\alpha)$.

Then $\vdash_{\mathsf{PA}} \varphi \to \partial\varphi$ holds for all Σ_1-formulas $\varphi \in \mathcal{L}_{ar}$.

Proof. ∂ satisfies also $d0$, $d00$, and $d\wedge$; see Remark 1. Due to $d00$, it is enough to carry out the proof for the special Σ_1-formulas defined in **6.7**. First let φ be $\mathtt{S}x\mathbf{=}y$. $\varphi \vdash_{\mathsf{PA}} \partial\varphi$ is equivalent to $\vdash_{\mathsf{PA}} \partial\varphi\,\frac{\mathtt{S}x}{y}$, and this to $\vdash_{\mathsf{PA}} \partial\,\mathtt{S}x\mathbf{=}\mathtt{S}x$ by ds, which is obvious from $d1$. Similarly, $y\mathbf{=}z \vdash_{\mathsf{PA}} \partial\,y\mathbf{=}z$, which we need in the inductive proof of $\vdash_{\mathsf{PA}} \varphi \to \partial\varphi$ for $\varphi := x{+}y\mathbf{=}z$ on x: $\varphi\,\frac{0}{x} \vdash_{\mathsf{PA}} y\mathbf{=}z \vdash_{\mathsf{PA}} \partial\,y\mathbf{=}z \vdash_{\mathsf{PA}} \partial(\varphi\frac{0}{x}) \vdash_{\mathsf{PA}} \partial\varphi\,\frac{0}{x}$. Thus, $\vdash_{\mathsf{PA}} (\varphi \to \partial\varphi)\,\frac{0}{x}$. Note that $\varphi\frac{\mathtt{S}y}{y} \equiv_{\mathsf{PA}} \varphi\frac{\mathtt{S}x}{x}$ and so $\partial\varphi\,\frac{\mathtt{S}y}{y} \equiv_{\mathsf{PA}} \partial\varphi\,\frac{\mathtt{S}x}{x}$, by $d00, ds$. Then the induction step $\forall yz(\varphi \to \partial\varphi) \vdash_{\mathsf{PA}} \forall yz(\varphi \to \partial\varphi)\frac{\mathtt{S}x}{x}$ obviously follows from $\forall yz(\varphi \to \partial\varphi) \vdash \varphi\,\frac{\mathtt{S}y}{y} \to \partial\varphi\,\frac{\mathtt{S}y}{y} \vdash_{\mathsf{PA}} \varphi\,\frac{\mathtt{S}x}{x} \to \partial\varphi\,\frac{\mathtt{S}x}{x} = (\varphi \to \partial\varphi)\frac{\mathtt{S}x}{x}$. The formula $x \cdot y\mathbf{=}z$ is left to the reader. We now treat the logical connectives. The induction steps for

$\wedge, \vee, \exists$ are simple: $\alpha\wedge\beta \vdash_{\mathsf{PA}} \alpha, \beta \vdash_{\mathsf{PA}} \partial\alpha\wedge\partial\beta \vdash_{\mathsf{PA}} \partial(\alpha\wedge\beta)$ by $d\wedge$. For $\vee$ observe that $\alpha \vdash_{\mathsf{PA}} \partial\alpha \vdash_{\mathsf{PA}} \partial(\alpha\vee\beta)$, and similarly for β. Further, because $\varphi \vdash_{\mathsf{PA}} \exists x\varphi$ we have $\varphi \vdash_{\mathsf{PA}} \partial\varphi \vdash_{\mathsf{PA}} \partial\exists x\varphi$, and since $x \notin \mathit{free}\,\partial\exists x\varphi$, it follows that $\exists x\varphi \vdash_{\mathsf{PA}} \partial\exists x\varphi$. The step for prime-term substitution (t is prime in $\frac{t}{x}$) runs also straightforwardly: $\varphi \vdash_{\mathsf{PA}} \partial\varphi$ yields $\varphi\frac{t}{x} \vdash_{\mathsf{PA}} \partial\varphi\frac{t}{x} \vdash_{\mathsf{PA}} \partial(\varphi\frac{t}{x})$.

It remains to show the step for bounded quantification. Suppose $\alpha \vdash_{\mathsf{PA}} \partial\alpha$ and let $y \notin \mathit{var}\,\alpha$. We show that $\varphi := (\forall x{<}y)\,\alpha \vdash_{\mathsf{PA}} \partial\varphi$ by induction on y. The initial step is clear: $\vdash_{\mathsf{PA}} \varphi\frac{0}{y}$, thus $\vdash_{\mathsf{PA}} \partial(\varphi\frac{0}{y}) \vdash_{\mathsf{PA}} \partial\varphi\frac{0}{y}$, and a fortiori $\vdash_{\mathsf{PA}} \varphi\frac{0}{y} \to \partial\varphi\frac{0}{y}$. Clearly, $\varphi\frac{\mathsf{S}y}{y} \equiv_{\mathsf{PA}} \varphi\wedge\alpha\frac{y}{x}$. Hence $\alpha \vdash_{\mathsf{PA}} \partial\alpha$ yields $\alpha\frac{y}{x} \vdash_{\mathsf{PA}} \partial\alpha\frac{y}{x} \vdash_{\mathsf{PA}} \partial(\alpha\frac{y}{x})$. That leads to

$$\varphi\tfrac{\mathsf{S}y}{y} \wedge(\varphi\to\partial\varphi) \vdash_{\mathsf{PA}} \varphi\wedge\alpha\tfrac{y}{x} \wedge(\varphi\to\partial\varphi) \vdash_{\mathsf{PA}} \partial\varphi\wedge\partial(\alpha\tfrac{y}{x}) \vdash_{\mathsf{PA}} \partial(\varphi\wedge\alpha\tfrac{y}{x}) \vdash_{\mathsf{PA}} \partial(\varphi\tfrac{\mathsf{S}y}{y}).$$

Thus $\varphi\to\partial\varphi \vdash_{\mathsf{PA}} \varphi\frac{\mathsf{S}y}{y} \to\partial(\varphi\frac{\mathsf{S}y}{y})$, which is equivalent to the inductive step. ❑

$D1$–$D3$ are also provable for much weaker theories than PA, e.g., for so-called *elementary arithmetic* $\mathsf{EA} = I\Delta_0 + \forall xy\exists z\delta_{exp}(x,y,z)$. Here we take δ_{exp} to be a Δ_0-formula according to Remark 1 in **6.3**, with $I\Delta_0$ likewise defined there. An equivalent formulation of EA can be found in [FS].

It is noteworthy that the provable recursive functions in EA are precisely the elementary ones (shown in [Si]). If EA is augmented by the Π_2-induction schema without parameters, then exactly the p.r. functions are provably recursive. This beautiful result was proved in [Be4]. Further theories are discussed in [Ba, Part D].

Exercises

1. Prove in PA (using basic laws of arithmetic, e.g. the axioms of N page 86) the definability of the pairing and remainder function: $\forall xy\exists!z\,\underline{2}z = (x+y)^2+\underline{3}x+y$ and $\forall xy\exists!z(\exists v\, x = y\cdot v + z\wedge z < y \vee y = z = 0)$, and also $\forall xy\exists!z\,\mathtt{beta}(x,y,z)$.

2. Prove *in* PA
   (a) Euclid's lemma $(\forall ab{>}0)\exists xy\, ax + \underline{1} = by$,
   (b) $(\forall a{>}1)\exists p(\mathsf{prim}\,p\wedge p|a)$,
   (c) $\vdash_{\mathsf{PA}} \forall abp(\mathsf{prim}\,p\wedge p|ab\to p|a\vee p|b)$.

3. $(\forall k{\geqslant}\underline{2})\exists u\exists n(k = \prod_{\nu\leqslant n} p_i^{\beta u\nu} \wedge \beta un \neq 0)$ can be viewed as a formalization of the prime factorization.[5] Prove this in PA, as well as its uniqueness.

4. Let $T' = T + \alpha$ and T satisfy $D1$–$D4$. Show that
   (a) $\vdash_T \Box_{T'}\varphi \leftrightarrow \Box_T(\alpha\to\varphi)$ (the formalized deduction theorem),
   (b) $D1$–$D4$ hold also for T'.

[5] There are several equivalent formalizations of the prime factorization in PA. Particularly nice is $(\forall n{\geqslant}\underline{2})(\exists m{\geqslant}\underline{2})n = \prod_{\nu\leqslant\ell m} p_i^{(m)_i}$. Here m serves as a variable for the sequence of prime exponents.

7.2 The Theorems of Gödel and Löb

We are now in a position to harvest the yields of our efforts. As long as not stated otherwise, let T denote any arithmetizable axiomatic theory in a language $\mathcal{L}$, which satisfies the derivability conditions $D1$–$D3$ of **7.1** along with the fixed-point lemma of **6.5**. We direct attention straight away to the uniqueness statement of Lemma 2.1(b) below. According to this claim, at most $\Box\alpha \to \alpha$ can be the fixed-point of the formula $\Box(x) \to \alpha$, up to equivalence in T. The proof of Theorem 2.2 will show that $\neg\Box(x)$ too has only one fixed point modulo T. Beneath all this lies, as we shall see from Corollary 4.6, a completely general result.

Lemma 2.1. *Let T be as arranged above, and $\alpha, \gamma \in \mathcal{L}^0$ such that $\gamma \equiv_T \Box\gamma \to \alpha$. Then hold* (a) $\Box\gamma \equiv_T \Box\alpha$ *and* (b) $\gamma \equiv_T \Box\alpha \to \alpha$.

Proof. The supposition yields $\Box\gamma \vdash_T \Box(\Box\gamma \to \alpha) \vdash_T \Box\Box\gamma \to \Box\alpha$, by $D0$ and $D2$. Now, by $D3$, we clearly obtain $\Box\gamma \vdash_T \Box\Box\gamma$ and so $\Box\gamma \vdash_T \Box\alpha$. Since obviously $\alpha \vdash_T \Box\gamma \to \alpha \equiv_T \gamma$ and so $\alpha \vdash_T \gamma$, it follows by $D0$ that $\Box\alpha \vdash_T \Box\gamma$. This, together with the already verified $\Box\gamma \vdash_T \Box\alpha$, proves (a). Using (a) we may replace $\Box\gamma$ with $\Box\alpha$ in $\gamma \equiv_T \Box\gamma \to \alpha$ which results in (b). ❏

Theorem 2.2 (Second incompleteness theorem). PA *satisfies alongside the fixed-point lemma also $D1$–$D3$. For every theory T with these properties,*

(1) $\nvdash_T \mathtt{Con}_T$ *provided T is consistent,*

(2) $\vdash_T \mathtt{Con}_T \to \neg\Box\,\mathtt{Con}_T$.

Proof. $D1$–$D3$ were proved for PA in **7.1**. (1) follows from (2). Assume $\vdash_T \mathtt{Con}_T$. Then $\vdash_T \Box\,\mathtt{Con}_T$ by $D1$, as well as $\vdash_T \neg\Box\,\mathtt{Con}_T$ by (2). Thus, T is inconsistent. To verify (2), let γ be a fixed point of $\neg\Box(x)$, so that

(∗) $\gamma \equiv_T \neg\Box\gamma \equiv \Box\gamma \to \bot$.

By Lemma 2.1(b) with $\alpha = \bot$, we obtain $\gamma \equiv_T \Box\bot \to \bot \equiv \neg\Box\bot = \mathtt{Con}_T$. Replacing γ in (∗) with $\mathtt{Con}_T$ gives $\mathtt{Con}_T \equiv_T \neg\Box\,\mathtt{Con}_T$. Half of this is the claim (2). ❏

Thus, by (1), no sufficiently strong consistent theory can prove its own consistency. In particular, $\nvdash_{\mathsf{PA}} \mathtt{Con}_{\mathsf{PA}}$ since PA is assumed to be consistent. The proof shows that $\mathtt{Con}_T$ is the only fixed point of $\neg\,\mathtt{bwb}_T$ modulo T. It shows also a bit more, namely

(3) $\mathtt{Con}_T \equiv_T \neg\Box\,\mathtt{Con}_T$.

This strengthens (2), but only by a little: $\neg\Box\,\mathtt{Con}_T \vdash_T \mathtt{Con}_T$ is just a special case of

(4) $\neg\Box\alpha \vdash_T \mathtt{Con}_T$ (equivalently, $\neg\,\mathtt{Con}_T \vdash_T \Box\alpha$), for every $\alpha \in \mathcal{L}$.

This follows from $\bot \vdash_T \alpha$ since $\neg\,\mathtt{Con}_T \equiv \Box\bot \vdash_T \Box\alpha$ by $D0$. (4) reflects in T the fact 'If T is inconsistent then every formula is provable'. From (1) and (3) we get $\nvdash_{\mathsf{PA}} \neg\Box\,\mathtt{Con}_{\mathsf{PA}}$, although '$\mathtt{Con}_{\mathsf{PA}}$ is unprovable in PA' is true according to (1).

All these claims hold independently of the "truth content" of the $\alpha \in T$. Namely, a consequence of the second incompleteness theorem is the existence of consistent arithmetical theories $T \supseteq \mathsf{PA}$ in which along with claims true in $\mathcal{N}$ also false ones are provable, i.e., in which truth and untruth live in peaceful coexistence with each other. Such "dream theories" are highly rich in content, for all of them include ordinary number theory. An example is $\mathsf{PA}^\perp := \mathsf{PA} + \neg\,\mathtt{Con}_{\mathsf{PA}}$. This theory is consistent because *the consistency of* $\mathsf{PA}^\perp$ *is equivalent to the unprovability of* $\mathtt{Con}_T$ *in* PA. The italicized sentence is even provable in PA as (5) below will show. By the reflection of the deduction theorem in PA (Exercise 4(a) in **7.1** with $T = \mathsf{PA}$ and $\Box = \Box_{\mathsf{PA}}$), $\Box_{\mathsf{PA}+\alpha}\bot \equiv_{\mathsf{PA}} \Box(\alpha \to \bot) \equiv \Box\neg\alpha$, hence $\neg\Box_{\mathsf{PA}+\alpha}\bot \equiv_{\mathsf{PA}} \neg\Box\neg\alpha$, and so

(5) $\mathtt{Con}_{\mathsf{PA}+\alpha} \equiv_{\mathsf{PA}} \Diamond\alpha$ (in particular, $\mathtt{Con}_{\mathsf{PA}^\perp} \equiv_{\mathsf{PA}} \Diamond(\neg\,\mathtt{Con}_{\mathsf{PA}}) \equiv \neg\Box\,\mathtt{Con}_{\mathsf{PA}}$).

Now, the special case under (5) and (3) clearly yield

(6) $\mathtt{Con}_{\mathsf{PA}} \equiv_{\mathsf{PA}} \mathtt{Con}_{\mathsf{PA}^\perp}$ (hence also $\mathtt{Con}_{\mathsf{PA}} \equiv_{\mathsf{PA}^\perp} \mathtt{Con}_{\mathsf{PA}^\perp}$).

Put together, $\mathsf{PA}^\perp$ contains ordinary number theory as known to us, but also proves the indubitably false sentence $\mathtt{bwb}_{\mathsf{PA}}(\ulcorner 0 \neq 0 \urcorner)$. Moreover, because of $\vdash_{\mathsf{PA}^\perp} \neg\,\mathtt{Con}_{\mathsf{PA}}$, hence $\vdash_{\mathsf{PA}^\perp} \neg\,\mathtt{Con}_{\mathsf{PA}^\perp}$ by (6), $\mathsf{PA}^\perp$ proves its own inconsistency, although $\mathsf{PA}^\perp$ is consistent. It claims to have a mysterious proof of $\bot$. Thus, consistency of T can have a different meaning within T and seen from outside, similar as the meanings of *countable* diverge, depending on whether one is situated in ZFC or is looking at it from outside. One may even say that $\mathsf{PA}^\perp$ is lying to us with the claim $\neg\,\mathtt{Con}_{\mathsf{PA}^\perp}$. However this phenomenum is paraphrased, we learn that for a consistent theory T, the extension $T + \mathtt{Con}_T$ need not be consistent. $T = \mathsf{PA}^\perp$ is an example, and in fact only one of many others. More will be said on this in Theorem 2.4.

We now discuss what is, along with (3), the most famous example of a self-referential sentence. Clearly, a fixed point α of $\Box(x) = \mathtt{bwb}_T(x)$ claims just its own provability, that is, $\alpha \equiv_T \Box\alpha$. A trivial example is $\alpha = \top$, because $\vdash_T \Box\top \to \top$, and since $\vdash_T \top$, clearly $\vdash_T \Box\top$ so that $\top \equiv_T \Box\top$. What is surprising here is that $\top$ turns out to be the only fixed point of $\Box(x)$ modulo T. By $D4^\circ$ below, $\vdash_T \Box\alpha \to \alpha$ implies $\vdash_T \alpha$ and so $\alpha \equiv_T \top$ (which confirms the uniqueness), although one might perhaps expect $\vdash_T \Box\alpha \to \alpha$ for all $\alpha \in \mathcal{L}^0$ because $\Box\alpha \to \alpha$ is intuitively true.

Theorem 2.3 (Löb's theorem). *Take T to satisfy $D1$–$D3$ and the fixed-point lemma. Then T has the properties*

$$D4\colon\ \vdash_T \Box(\Box\alpha \to \alpha) \to \Box\alpha, \qquad D4^\circ\colon\ \vdash_T \Box\alpha \to \alpha \ \Rightarrow\ \vdash_T \alpha \qquad (\alpha \in \mathcal{L}^0).$$

Proof. Let γ be a fixed point of $\Box(x) \to \alpha$, i.e., $\gamma \equiv_T \Box\gamma \to \alpha$. Then $\gamma \equiv_T \Box\alpha \to \alpha$ by Lemma 2.1(b). This and $D0$ imply $\Box\gamma \equiv_T \Box(\Box\alpha \to \alpha)$. Lemma 2.1(a) states $\Box\gamma \equiv_T \Box\alpha$, hence $\Box\alpha \equiv_T \Box(\Box\alpha \to \alpha)$. Half of this is $D4$. Now suppose $\vdash_T \Box\alpha \to \alpha$. Then by $D1$, $\vdash_T \Box(\Box\alpha \to \alpha)$. Using $D4$ results in $\vdash_T \Box\alpha$, and $\vdash_T \Box\alpha \to \alpha$ finally yields $\vdash_T \alpha$, thus proving $D4^\circ$. ❏

$D4$ is just $D4^\circ$, formalized in T. One of many applications of Löb's theorem is a very easy proof of $\nvdash_{\mathsf{PA}} \mathtt{Con}_{\mathsf{PA}}$. Indeed, $\vdash_{\mathsf{PA}} \mathtt{Con}_{\mathsf{PA}}$ ($\equiv \Box\bot \to \bot$) implies $\vdash_{\mathsf{PA}} \bot$ by $D4^\circ$. That's all. Similarly, $D4$ implies (2) for $\alpha = \bot$ by contraposition. Thus, Theorem 2.3 is stronger than Theorem 2.2, which is not obvious at the first glance.

Unlike $\mathsf{PA}^\bot$, $\mathsf{PA}+\mathtt{Con}_{\mathsf{PA}}$ conforms to truth. Unfortunately it is not quite clear what $\mathtt{Con}_{\mathsf{PA}}$ means in number-theoretical terms. This is clear, however, for an arithmetical statement discovered by Paris and Harrington, which implies $\mathtt{Con}_{\mathsf{PA}}$; this statement is provable in ZFC but not in PA in view of (1). Since then, many such sentences have been found, mostly of a combinatorial nature. A popular example is

Goodstein's theorem. *Every Goodstein sequence ends in* 0.

A Goodstein sequence is a number sequence $(a_n)_{n\in\mathbb{N}}$ with arbitrary a_0 given in advance, such that a_{n+1} is obtained from a_n as follows: Let $b_n = n+2$, so that $b_0 = 2$, $b_1 = 3$, etc. Expand a_n in b-adic base for $b := b_n$, so that for suitable k,

$$(*) \quad a_n = \textstyle\sum_{i\leqslant k} b^{k-i}c_i, \text{ with } 0 \leqslant c_i < b.$$

Also the powers $k-i$ are represented in b-adic form, so too the powers of powers, and so on. Now replace b everywhere with $b+1$ ($= b_{n+1}$) and subtract 1 from the output. The result is a_{n+1}. The table below gives an example beginning with $a_0 = 11$; already a_5 has the value 134 217 728. As one sees from this example, a_n initially increases enormously, and it is hardly believable that the sequence ever starts to decrease and ends in 0. But the proof of the theorem is not particularly difficult; one estimates a_n from above by the ordinal number λ_n, which, crudely put, results from a_n if replacing the basis b in $(*)$ by ω. With some ordinal arithmetic it can readily be shown that $\lambda_{n+1} < \lambda_n$ as long as $\lambda_n \neq 0$. Since there is no properly decreasing infinite sequence of ordinal numbers (these are well-ordered), the sequence $(a_n)_{n\in\mathbb{N}}$ must eventually end in 0. For more detailed information see for instance [HP].

$a_0 = 11 = 2^{2+1} + 2 + 1$	$2 \rightsquigarrow 3$	$3^{3+1} + 3 + 1 = 85$
$a_1 = 84 = 3^{3+1} + 3$	$3 \rightsquigarrow 4$	$4^{4+1} + 4 = 1028$
$a_2 = 1027 = 4^{4+1} + 3$	$4 \rightsquigarrow 5$	$5^{5+1} + 3 = 15\,628$
$a_3 = 15\,627 = 5^{5+1} + 2$	$5 \rightsquigarrow 6$	$6^{6+1} + 2 = 279\,938$
$a_4 = 279\,937 = 6^{6+1} + 1$	$6 \rightsquigarrow 7$	$7^{7+1} + 1 = 5\,764\,802$

Many metatheoretical properties can be expressed using the provability operator $\Box$ in T, often using sentence schemas. The following are examples that facilitate a better understanding of the meaning of $\neg\,\mathtt{Con}_T$ within T. By Theorem 6.5.1′, none of the following properties holds for a consistent T when seen from the outside:

(i) $\neg\,\mathtt{Con}_T$: $\Box\bot$ (provable inconsistency),
(ii) SyComp : $\Box\alpha \vee \Box\neg\alpha$ (syntactic completeness),
(iii) SeComp : $\alpha \to \Box\alpha$ (semantic completeness),
(iv) ω-Comp : $\forall x\Box[\varphi(x)] \to \Box\forall x\varphi(x)$ (ω-completeness).

Theorem 2.4. *The properties* (i)–(iv) *are all equivalent in a theory T satisfying the properties named at the beginning of this section.*

Proof. By (4) (i)⇒(ii),(iii),(iv) are clear. (ii)⇒(i): By Rosser's theorem formulated in T (see **7.4**), $\mathtt{Con}_T \vdash_T \neg\Box\alpha \wedge \neg\Box\neg\alpha$ for some α. Thus, $\Box\alpha \vee \Box\neg\alpha \vdash_T \neg\,\mathtt{Con}_T$. (iii)⇒(i): For $\alpha := \mathtt{Con}_T$, SeComp and Theorem 2.2 give $\alpha \vdash_T \Box\alpha, \neg\Box\alpha$, so $\vdash_T \neg\alpha$. (iv)⇒(i): By (9) in **7.1** we obtain $\neg\,\mathtt{bew}_T(x, \ulcorner\bot\urcorner) \vdash_T \Box[\neg\,\mathtt{bew}_T(x, \ulcorner\bot\urcorner)]$. Therefore, $\mathtt{Con}_T = \forall x \neg\,\mathtt{bew}_T(x, \ulcorner\bot\urcorner) \vdash_T \forall x\Box[\neg\,\mathtt{bew}_T(x, \ulcorner\bot\urcorner)]$. From ω-Comp and (2) easily follows $\mathtt{Con}_T \vdash_T \Box\forall x \neg\,\mathtt{bew}(x, \ulcorner\bot\urcorner) = \Box\,\mathtt{Con}_T \vdash_T \neg\,\mathtt{Con}_T$. Hence $\vdash_T \neg\,\mathtt{Con}_T$. ❏

Remark. $\mathtt{Con}_T$ is also equivalent in T to other properties, for example to the schema $\Box\alpha \to \alpha$ for Π_1-formulas α (the *local* Π_1-reflection principle) as well as the *uniform* Π_1-reflection principle $\forall x\Box[\alpha(x)] \to \forall x\alpha(x)$ for Π_1-formulas α. Both the theorems of Paris–Harrington and of Goodstein are equivalent in PA to the uniform Σ_1-reflection, or equivalently, to the consistency of PA plus all true Π_1-sentences; see e.g. [Ba, D8].

We define inductively $T^0 = T$ and $T^{n+1} = T^n + \mathtt{Con}_{T^n}$. This *$n$-times-iterated consistency extension* T^n can be written as $T^n = T + \neg\Box^n\bot$ with $\Box = \mathtt{bwb}_T$, $\Box^0\alpha = \alpha$ and $\Box^{n+1}\alpha = \Box\Box^n\alpha$; Exercise 3. Thus, the consistency of T^n can be expressed by an iterated consistency statement in the basis theory T. Moreover, let $T^\omega := \bigcup_{n\in\omega} T^n$. By definition, $T^n \subseteq T^{n+1}$. Thus, because of $T^n = T + \neg\Box^n\bot$, the following three items are equivalent:

(i) T^ω is consistent, (ii) T^n is consistent for all n, (iii) $\nvdash_T \Box^n\bot$ for all n.

Like $\mathsf{PA}^1 = \mathsf{PA} + \mathtt{Con}_{\mathsf{PA}}$, also PA^ω conforms to truth if one is looking at PA from outside. When considered more closely, this means only that PA^ω is relatively consistent with respect to ZFC. In other terms, $\vdash_{\mathsf{ZFC}} \mathtt{Con}_{\mathsf{PA}^\omega}$. The argument (to be formalized in ZFC) runs as follows: $\vdash_{\mathsf{PA}^\omega} \bot$ implies $\vdash_{\mathsf{PA}^n} \bot$ for some n as was noticed above, hence $\vdash_{\mathsf{PA}} \Box^n\bot$. But this is impossible, as is shown by a repeated application of $D1^*$ on PA (see page 211). Alternatively, one may apply Exercise 4.

Exercises

1. Prove $D4^\circ$ for T by applying Theorem 2.2 to $T' = T + \neg\alpha$.

2. Show by means of Löb's theorem that $\mathtt{Con}_{\mathsf{PA}} \to \neg\Box\neg\,\mathtt{Con}_{\mathsf{PA}}$ is unprovable in PA, although this formula is true if seen from outside.

3. Let T^n recursively be defined as in the text below. Prove that $T^n = T + \neg\Box^n\bot$ and $\mathtt{Con}_{T^n} \equiv_T \neg\Box^{n+1}\bot$, where $\Box$ is $\mathtt{bwb}_T$.

4. Show that $\vdash_{\mathsf{ZFC}} \Box_{\mathsf{PA}}\alpha \to \alpha$ for every arithmetical sentence α from $\mathcal{L}^0_\in$.

7.3 The Provability Logic G

In **7.2** first-order logic was hardly required. It comes then as no surprise that many of the results there can be obtained propositionally, more precisely, in a certain modal propositional calculus. This calculus contains alongside $\wedge, \neg$ the falsum symbol $\bot$, and a further unary connective $\Box$ to be interpreted as the proof operator in $\mathcal{L}_{ar}$, which we denoted by $\Box$ as well. First we define a propositional language $\mathcal{F}_\Box$, whose formulas are denoted by H, G, F: (a) the propositional variables $p_1, p_2, \ldots$ and $\bot$ belong to $\mathcal{F}_\Box$; (b) if $H, G \in \mathcal{F}_\Box$ then so too $(H \wedge G), \neg H, \Box H \in \mathcal{F}_\Box$. No other strings belong to $\mathcal{F}_\Box$ in this context. $H \vee G$, $H \to G$ and $H \leftrightarrow G$ are defined as in **1.1**, $\top := \neg\bot$. Further, $\Box^0 H := H$, $\Box^{n+1} H := \Box\Box^n H$, and $\Diamond H := \neg\Box\neg H$.

Let G be the set of those formulas in $\mathcal{F}_\Box$ derivable using substitution in $\mathcal{F}_\Box$, modus ponens MP and the rule MN: $H/\Box H$ from the tautologies of two-valued propositional logic augmented by the axioms

$$\Box(p \to q) \to \Box p \to \Box q, \quad \Box p \to \Box\Box p, \quad \Box(\Box p \to p) \to \Box p \ (\textit{Löb's axiom}).$$

Strictly speaking, the axiom $\Box p \to \Box\Box p$ is not necessary; it is provable from the remaining axioms; see [Boo] or [Ra1]. For $H \in \mathsf{G}$ we write $\vdash_\mathsf{G} H$ (read "H is derivable in G"). MN obviously corresponds to condition $D1$. The first axiom of G reflects $D2$, the middle $D3$, and the last $D4$, hence its name *provability logic*. The connection between G and PA will be described in **7.4**. Here we are concerned with the formal system G and its semantics, known as *Kripke semantics*. For simplicity, we restrict ourselves to finite Kripke frames, which is just another name for finite, directed graphs. We begin without further ado with the following

Definition. A G-*frame* or *Kripke frame for* G is a finite strict partial order $(g, <)$. A *valuation* is a mapping w that assigns to every variable p a subset wp of g. The relation $P \Vdash H$, dependent on w, between points $P \in g$ and formulas $H \in \mathcal{F}_\Box$ (read "P accepts H") is defined inductively by

$$P \Vdash p \ \Leftrightarrow\ P \in wp, \qquad P \nVdash \bot, \qquad P \Vdash H \wedge G \ \Leftrightarrow\ P \Vdash H \ \&\ P \Vdash G,$$
$$P \Vdash \neg H \ \Leftrightarrow\ P \nVdash H, \quad P \Vdash \Box H \ \Leftrightarrow\ P' \Vdash H \text{ for all } P' > P.$$

If $P \Vdash H$ for all $P \in g$, all G-frames g, and all w, we write $\vDash_\mathsf{G} H$ and say that H *is* G-*valid*. The G-frame on the right, consisting of two points P_1, P_2 with $P_1 < P_2$, shows that $\nvDash_\mathsf{G} p \to \Box p$. Indeed, Let $wp = \{P_1\}$. Then $P_1 \Vdash p$, but $P_1 \nVdash \Box p$ since $P_2 \nVdash p$. Note also that $P_2 \nVdash \Box p \to p$ because $P_2 \Vdash \Box p$.

One easily sees that $P \Vdash \Diamond H$ iff $P' \Vdash H$ for some $P' > P$. Let us write $H \equiv_\mathsf{G} H'$ for $\vDash_\mathsf{G} H \leftrightarrow H'$. This relation is a congruence in $\mathcal{F}_\Box$ that conservatively extends the logical equivalence of formulas without $\Box$. Examples are $\neg\Box H \equiv_\mathsf{G} \Diamond\neg H$ and $\neg\Diamond H \equiv_\mathsf{G} \Box\neg H$. Many more interesting examples are presented in the following. These will later be translated into statements about self-reference.

Examples. (a) Although always $P \nVdash \bot$, $P \Vdash \Box\bot$ obviously holds iff P is maximal in g, that is, if no $P' > P$ exists. Likewise, $\Box\neg\Box\bot$ is accepted precisely at the maximal points of g. Therefore, $\Box\neg\Box\bot \equiv_{\mathsf{G}} \Box\bot$, or $\neg\Box\neg\Box\bot \equiv_{\mathsf{G}} \neg\Box\bot$. This equivalence reflects in G the second incompleteness theorem as will be seen in **7.4**.
(b) Let $\{P_0, \ldots, P_n\}$ be the ordered G-frame with $P_n < \cdots < P_0$. Induction on n shows that $P_n \Vdash \Box^m\bot$ for $m > n$, but $P_n \nVdash \Box^n\bot$, and moreover $P_n \nVdash \Box^{n+1}\bot \to \Box^n\bot$. Hence, $\nvDash_{\mathsf{G}} \Box^{n+1}\bot \to \Box^n\bot$, and a fortiori $\nvDash_{\mathsf{G}} \Box^n\bot$ for all n.
(c) $\vDash_{\mathsf{G}} \Box(\Box p \to p) \to \Box p$. For take an arbitrary g and $P \in g$. If $P \nVdash \Box p$ then there is, since g is finite, some $Q > P$ with $Q \Vdash \neg p$ and $Q' \Vdash p$ for all $Q' > Q$. Thus $Q \Vdash \Box p$; hence $Q \nVdash \Box p \to p$ and so $P \nVdash \Box(\Box p \to p)$. Consequently, $P \Vdash \Box(\Box p \to p) \to \Box p$, which proves our claim, since g was an arbitrary G-frame.
(d) $\vDash_{\mathsf{G}} \neg\Box^{n+1}\bot \to \Diamond R_n$, where $R_n := \bigwedge_{i=1}^{n}(\Box p_i \to p_i)$. Indeed, suppose $P \Vdash \neg\Box^{n+1}\bot$, $P \in g$. Then there must be a chain $P = P_0 < P_1 < \cdots < P_{n+1}$ in g. Now, it is a nice separate exercise to verify that each conjunct of R_n fails to be accepted by at most one of the $n+1$ points $P_1, \ldots, P_{n+1}$. Thus, at least one of these accepts all conjuncts. In other words, $P_i \Vdash R_n$ for some $i > 0$, hence $P \Vdash \Diamond R_n$. This nontrivial example will essentially be used in the proof of Theorem 6.1.

It is easy to prove by induction on $\vdash_{\mathsf{G}} H$ that $\vdash_{\mathsf{G}} H \Rightarrow \vDash_{\mathsf{G}} H$; example (c) is a part of the initial step. The induction step over the rule MN is verified by contraposition: if $P \nVdash \Box H$ then there must be some $P' > P$ in some G-frame with $P' \nVdash H$.

The converse, $\vDash_{\mathsf{G}} H \Rightarrow \vdash_{\mathsf{G}} H$, is not so easily shown. It is part of the following theorem, used is the sequel without proof. It tells us that $\vdash_{\mathsf{G}} H$ can be confirmed by showing that $\vDash_{\mathsf{G}} H$, and vice versa. The particular import of this theorem will become clear only in Theorem 4.2. As for the relatively simple formulas considered in the sequel, we check directly whether they are G-valid. For a proof of Theorem 3.1, based on the finite model property of G, see e.g. [Boo], [Ra1], or [CZ].

Theorem 3.1 (Completeness of Kripke semantics for G). $\vdash_{\mathsf{G}} H \;\Leftrightarrow\; \vDash_{\mathsf{G}} H$.

Both the formulas provable in G and those refutable there are obviously recursively enumerable, thanks to the finite model property of $\vDash_{\mathsf{G}}$. Thus, in complete analogy to Exercise 2 in **3.6**, we obtain the following result:

Theorem 3.2. G *is decidable.*

Remark. Let $\mathfrak{F}^0_\Box$ be the set of variable-free formulas of $\mathfrak{F}_\Box$. An important fragment of G is $\mathsf{G}^0 := \mathsf{G} \cap \mathfrak{F}^0_\Box$. The most interesting formulas in $\mathfrak{F}^0_\Box$ are the $\neg\Box^n\bot$ ($\equiv_{\mathsf{G}} \Diamond^n\top$), for these form a Boolean base in G^0. One proves this statement most easily by showing that G^0 is complete with respect to all (totally) ordered G-frames, including the infinite ones, and applying the base theorem 5.2.3 accordingly: two ordered G-frames satisfying the same "base formulas" $\Box^n\bot$ are either both finite and isomorphic, or both infinite and indistinguishable by means of the formulas $H \in \mathfrak{F}^0_\Box$.

7.4 The Modal Treatment of Self-Reference

Let T be a theory in $\mathcal{L}$ as in **7.2**. A mapping $\imath : p_i \mapsto \alpha_i$ ($\in \mathcal{L}^0$) will be called an *insertion*. $\imath$ assigns to every H an $\mathcal{L}$-sentence $H^\imath$ by extending it to the whole of $\mathfrak{F}_\Box$ by $\bot^\imath = \bot$, $(\neg H)^\imath = \neg H^\imath$, $(H \wedge G)^\imath = H^\imath \wedge G^\imath$, and $(\Box H)^\imath = \Box H^\imath$. In other words, $H^\imath$ results from $H = H(p_1, \ldots, p_n)$ by replacing the p_ν by the α_ν, denoted also by $H^\imath = H(\alpha_1, \ldots, \alpha_n)$. For instance, $(\Box p \wedge \neg\Box\bot)^\imath = \Box\alpha \wedge \neg\Box\bot$ if $p^\imath = \alpha$. In particular, $(\neg\Box\bot)^\imath = \neg\Box\bot = \mathtt{Con}_T$. The following lemma shows that $\vdash_\mathsf{G}$ is "sound" for $\vdash_T$. This already considerably simplifies proofs of self-referential statements.

Lemma 4.1. *For every H such that $\vdash_\mathsf{G} H$ and every insertion $\imath$ in $\mathcal{L}$, $\vdash_T H^\imath$.*

Proof by induction on $\vdash_\mathsf{G} H$. If H is a propositional tautology then $H^\imath \in \mathit{Taut}_L \subseteq T$. If H is one of the modal axioms of G, then $\vdash_T H^\imath$ by $D2$, $D3$, and $D4$. If $\vdash_\mathsf{G} H$ and $\sigma : \mathfrak{F}_\Box \to \mathfrak{F}_\Box$ is a substitution, then $\vdash_T H^{\sigma\imath}$, because $H^{\sigma\imath} = H^{\imath'}$ with $\imath' : p \mapsto p^{\sigma\imath}$, and $\vdash_T H^{\imath'}$ holds by the induction hypothesis. As regards the induction step over MP, consider $(F \to G)^\imath = F^\imath \to G^\imath$. If MN is applied, and $\vdash_T H^\imath$ by the induction hypothesis, then $\vdash_T \Box H^\imath = (\Box H)^\imath$, due to $D1$. ❑

Example 1. We prove (3) of Theorem 2.2 with the calculus $\vdash_\mathsf{G}$. By Lemma 4.1 and Theorem 3.1 it suffices to show that $\vDash_\mathsf{G} \neg\Box\bot \leftrightarrow \neg\Box\neg\Box\bot$. But this holds by Example (a) in **7.3**. Next example: $\vDash_\mathsf{G} \Box(p \leftrightarrow \Diamond p) \to \neg\Diamond p$ is easily confirmed. Thus, $\Box(\alpha \leftrightarrow \Diamond\alpha) \to \neg\Diamond\alpha$ is provable in PA. This formula tells us "a sentence claiming its own consistency with PA is incompatible with PA", which hardly seems plausible. Even the converse is provable in PA since $\vDash_\mathsf{G} \neg\Diamond p \to \Box(p \leftrightarrow \Diamond p)$.

We now explain certain facts that expand upon the reasoning of above. For PA and related theories the converse of Lemma 4.1 holds as well. That is to say, the derivability conditions and Löb's theorem already contain everything worth knowing about self-referential formulas or schemas. For the subtle proofs of Theorems 4.2, 4.4, and 4.5, the reader is referred to [Boo].

Theorem 4.2 (Solovay's completeness theorem). *For all formulas $H \in \mathfrak{F}_\Box$: $\vdash_\mathsf{G} H$ (equivalently $\vDash_\mathsf{G} H$) if and only if $\vdash_\mathsf{PA} H^\imath$ for all insertions $\imath$.*

Example 2 (applications). (a) $\nvdash_\mathsf{PA} \Box^{n+1}\bot \to \Box^n\bot$ because by Example (b) in **7.3**, $\nvDash_\mathsf{G} \Box^{n+1}\bot \to \Box^n\bot$. In particular we get $\nvdash_\mathsf{PA} \mathtt{Con}_\mathsf{PA}$ ($\equiv \Box\bot \to \bot$). (b) $\nvdash_\mathsf{PA} \neg\Box^{n+1}\bot$, since $\nvDash_\mathsf{G} \neg\Box^{n+1}\bot$. (c) It is easily verified with the 2-point frame on page 221 that $\nvDash_\mathsf{G} \neg\Box p \to \Box\neg\Box p$, in particular $\nvDash_\mathsf{G} \neg\Box\bot \to \Box\neg\Box\bot$. Therefore, $\nvdash_\mathsf{PA} \mathtt{Con}_\mathsf{PA} \to \Box\,\mathtt{Con}_\mathsf{PA}$. (d) $\mathsf{PA}_n := \mathsf{PA} + \Box^n\bot$ is consistent for $n > 0$ by (b), but is ω-inconsistent. Otherwise, by $D1^*$ (page 211), $\vdash_{\mathsf{PA}_n} \Box^n\bot \Rightarrow \vdash_{\mathsf{PA}_n} \Box^{n-1}\bot \Rightarrow \cdots \Rightarrow \vdash_{\mathsf{PA}_n} \bot$, contradicting $\nvdash_{\mathsf{PA}_n} \bot$. Since $\vdash_\mathsf{PA} \Box^n\bot \to \Box^{n+1}\bot$ by $D3$, we obtain $\mathsf{PA}_n \supseteq \mathsf{PA}_{n+1}$, and since $\mathsf{PA}_n \neq \mathsf{PA}_{n+1}$ by (a), it follows that $\mathsf{PA}_0 \supset \mathsf{PA}_1 \supset \cdots \supset \mathsf{PA}$. Observe that PA_1 is just $\mathsf{PA}^\bot$.

Note also since $\not\vdash_{\mathsf{G}} \Box p \to p$, there must be some $\alpha \in \mathcal{L}^0_{ar}$ with $\vdash_{\mathsf{PA}} \Box\alpha \to \alpha$ (which one?) The above examples point out that Theorem 4.2 and the decidability of G are very efficient instruments in deciding the provability of self-referential statements.

Many other theories have the same provability logic as PA, where in general a modal propositional logic H is the *provability logic* for T when the analogue of Theorem 4.2 holds with respect to T and H. For some theories, the provability logic may be a proper extension of G. For example, the ω-inconsistent theory PA_n from Example 2(d) has the provability logic $\mathsf{G}_n := \mathsf{G} + \Box^n\bot$, the smallest extension of G closed under all rules of G with the additional axiom $\Box^n\bot$. This follows directly from Theorem 4.2 (Exercise 1). By the following theorem (due to A. Visser), other extensions of G to be considered as provability logics are out of the question.

Theorem 4.3. *Let T be at least as strong as PA and T^ω as on page 220. Then*
(a) *whenever T^ω is consistent, then G is the provability logic of T* (proof in **7.6**),
(b) *if $\vdash_{T^\omega} \bot$ and n is minimal such that $\vdash_{T^n} \bot$, then T's provability logic is G_n.*

The formulas $H \in \mathfrak{F}_\Box$ such that $\mathcal{N} \vDash H^\imath$ for all insertions $\imath$ in $\mathcal{L}_{ar}$ can also be surprisingly easily characterized. All $H \in \mathsf{G}$ are obviously included; but in addition also $\Box p \to p$, because obviously $\mathcal{N} \vDash \Box\alpha \to \alpha$ for all $\alpha \in \mathcal{L}^0_{ar}$.

Let GS ($\supseteq \mathsf{G}$) be the set of all formulas in $\mathfrak{F}_\Box$ that can be obtained from those in $\mathsf{G} \cup \{\Box p \to p\}$ using substitution and modus ponens only. Induction in GS readily yields $H \in \mathsf{GS} \Rightarrow \mathcal{N} \vDash H^\imath$ for all $\imath$. Again, the converse holds as well:

Theorem 4.4 ([So]). *$H \in \mathsf{GS}$ if and only if $\mathcal{N} \vDash H^\imath$ for all insertions $\imath$.*

GS is decidable as well. For it can be shown that $H \in \mathsf{GS} \Leftrightarrow H^* \in \mathsf{G}$, where

$$H^* := [\textstyle\bigwedge_{\Box G \in \mathrm{Sf}^\Box H}(\Box G \to G)] \to H.$$

Here $\mathrm{Sf}^\Box H$ is the set of subformulas of H of the form $\Box G$. By Theorem 4.4, many questions concerning the relations between *provable* and *true* are effectively decidable. For instance, $H(p) := \neg\Box(\neg\Box\bot \to \neg\Box p \wedge \neg\Box\neg p) \notin \mathsf{GS}$ can straightforwardly be verified. By Theorem 4.4 then $\mathcal{N} \vDash \neg H(\alpha) \equiv \Box(\neg\Box\bot \to \neg\Box\alpha \wedge \neg\Box\neg\alpha)$ for some $\alpha \in \mathcal{L}^0_{ar}$. Translated into English: *It is provable in* PA: *the consistency of* PA *implies the independence of α for some sentence α.* This is exactly Rosser's theorem which in this way turns out to be provable in PA. As was shown in [Be1], the box in the formulas $H \in \mathsf{GS}$ in Theorem 4.4 may denote $\mathtt{bwb}_T$ for *any* axiomatizable $T \supseteq \mathsf{PA}$, provided $T \subseteq Th\mathcal{N}$. However, if T proves false sentences (as does e.g. $\mathsf{PA}^\bot$) then GS has to be redefined in a feasible manner and is always decidable.

A variable p in H is called *modalized in H* if every occurrence of p is contained within the scope of a $\Box$, as is the case in $\neg\Box p$, $\neg\Box\neg p$, and $\Box(p \to q)$. By contrast, p is not modalized in $\Box p \to p$. Another particularly interesting theorem is

Theorem 4.5 (DeJongh–Sambin fixed-point theorem). *Let p be modalized in $H(p, q_1, \ldots, q_n)$, $n \geqslant 0$. Then there exist a formula $F = F(\vec{q})$ from $\mathfrak{F}_\Box$ such that*

(a) $F \equiv_{\mathsf{G}} H(F, \vec{q})$, (b) $\vdash_{\mathsf{G}} \bigwedge_{i=1}^{2}[(p_i \leftrightarrow H(p_i, \vec{q})) \wedge \Box(p_i \leftrightarrow H(p_i, \vec{q}))] \to (p_1 \leftrightarrow p_2)$.

From this theorem we easily obtain a corresponding result for theories T:

Corollary 4.6. *If p is modalized in $H = H(p, \vec{q})$ and T satisfies $D1$–$D4$, then there is an $F = F(\vec{q}) \in \mathfrak{F}_\Box$ with $F(\vec{\alpha}) \equiv_T H(F(\vec{\alpha}), \vec{\alpha})$ for all $\vec{\alpha} = (\alpha_1, \ldots, \alpha_n)$, $\alpha_i \in \mathcal{L}^0$. For each $\vec{\alpha}$ there is only one $\beta \in \mathcal{L}^0$ modulo T such that $\beta \equiv_T H(\beta, \vec{\alpha})$.*

Proof. Choose F according to (a) of the theorem. Then $F(\vec{\alpha}) \equiv_T H(F(\vec{\alpha}), \vec{\alpha})$ by Lemma 4.1 ($\vec{q}^{\,s} = \vec{\alpha}$). To prove uniqueness let $\beta_i \equiv_T H(\beta_i, \vec{\alpha})$ for $i = 1, 2$. By $D1$, $\vdash_T (\beta_i \leftrightarrow H(\beta_i, \vec{\alpha})) \wedge \Box(\beta_i \leftrightarrow H(\beta_i, \vec{\alpha}))$. Inserting β_i for p_i and α_i for q_i in the formula under (b) in the theorem then yields $\vdash_T \beta_1 \leftrightarrow \beta_2$ by Lemma 4.1. ❏

Example 3. For $H = \neg\Box p$ ($n = 0$), $F = \neg\Box\bot$ is a "solution" of (a) in Theorem 4.5 because $\neg\Box\bot \equiv_{\mathsf{G}} \neg\Box(\neg\Box\bot)$. According to Corollary 4.6, $\mathtt{Con}_T$ ($= \neg\Box\bot$) is modulo T the only fixed point of $\neg\,\mathtt{bwb}_T$. For $H = \Box p \to q$ (here n equals 1), $F = \Box q \to q$ is a solution of $F \equiv_{\mathsf{G}} H(F, \vec{q})$. The corollary states that $\Box\alpha \to \alpha$ is modulo T the only fixed point of $\mathtt{bwb}_T(x) \to \alpha$. This is exactly what was shown in Lemma 2.1.

Many special cases of the corollary represent older self-reference results from Gödel, Löb, Rogers, Jeroslow, and Kreisel which, stated in terms of modal logic, concern fixed points of $\neg\Box p$, $\Box p$, $\neg\Box\neg p$, $\Box\neg p$, $\Box p \to q$, and $\Box(p \to q)$ (these are, in order, $\neg\Box\bot$, $\top$, $\bot$, $\Box\bot$, $\Box q \to q$, and $\Box q$). Incidentally, for the listed formulas one gets fixed points according to a simple recipe. All listed formulas are of the form

$$H = G\tfrac{\Box H'}{p} \quad (p \text{ not modalized in } G(p, \vec{q});\ H'(p, \vec{q}) \text{ chosen appropriately}).$$

Then $F = H\tfrac{G(\top, \vec{q})}{p}$ is the fixed point of H, as can be seen after some calculation. For $H = \neg\Box p$ is $G = \neg p$. Hence, $F = \neg\Box p\tfrac{\neg\top}{p} = \neg\Box\neg\top \equiv_{\mathsf{G}} \neg\Box\bot$. For $H = \Box p \to q$ is $G = p \to q$, and so $F = (p \to q)\tfrac{\Box(\top \to q)}{p} = \Box(\top \to q) \to q \equiv_{\mathsf{G}} \Box q \to q$. For Kreisel's formula $\Box(p \to q)$ we have $G = p$. Therefore, $F = p\tfrac{\Box(\top \to q)}{p} = \Box(\top \to q) \equiv_{\mathsf{G}} \Box q$.

Exercises

1. Prove that the theory PA_n from Example 2(d) has the provability logic G_n.
2. Show that $\mathsf{PA}^n_\bot := \mathsf{PA}^n + \neg\,\mathtt{Con}_{\mathsf{PA}^n}$ equals $\mathsf{PA} + \Box^{n+1}\bot \wedge \neg\Box^n\bot$ ($\Box = \Box_{\mathsf{PA}}$) and has the provability logic $\mathsf{G}_1 = \mathsf{G} + \Box\bot$. Show the same for the theory $T = \mathsf{PA} + \Box(\Box\,\mathtt{Con}_{\mathsf{PA}} \vee \Box\neg\,\mathtt{Con}_{\mathsf{PA}}) \wedge \neg(\Box\,\mathtt{Con}_{\mathsf{PA}} \vee \Box\neg\,\mathtt{Con}_{\mathsf{PA}})$.
3. Prove that the recipe given in the text above is correct.
4. (Mostowski's theorem). Let $T \supseteq \mathsf{PA}$ be axiomatizable and suppose $T \vDash \mathcal{N}$. Show that there are two mutually independent Σ_1-sentences α, β in T, that is, α, $\neg\alpha$, $\alpha \to \beta$, $\alpha \to \neg\beta$, $\neg\alpha \to \beta$, $\neg\alpha \to \neg\beta$ are unprovable in T.

7.5 A Bimodal Provability Logic for PA

Hilbert remarked jokingly that the incompleteness phenomenon can be forcefully removed from the world by use of the so-called *ω-rule* ρ_ω : $\dfrac{X \vdash \varphi(\underline{n}) \text{ for all } n}{X \vdash \forall x\varphi}$. ρ_ω has infinitely many premises. It is an easy exercise to derive with the aid of ρ_ω every sentence α valid in $\mathcal{N}$ from the axioms of PA, even from those of Q. Indeed, all sentences can (up to equivalence) be obtained from variable-free literals with $\wedge, \vee, \forall, \exists$, bypassing formulas with free variables. Due to the Σ_1-completeness of Q, all valid variable-free literals are derivable. The inductive steps for $\wedge, \vee, \exists$ are simple, applying Σ_1-completeness in the $\exists$-step once again. Only in the $\forall$-step rule ρ_ω is used. Clearly, the unrestricted use of the infinitistic rule ρ_ω contradicts Hilbert's own intention of giving mathematics a finitistic foundation. However, things look different if we restrict ρ_ω each time to a *single* application. In view of Remark 1 in **6.2**, we no longer distinguish between φ and $\dot{\varphi}$, so that $\ulcorner\varphi\urcorner = \underline{\varphi}$. Let us define

$$1bwb_{\mathsf{PA}}(\alpha) := (\exists\varphi{\in}\mathcal{L}^1_{ar})[bwb_{\mathsf{PA}}(\forall x\varphi \to \alpha) \wedge \forall n\ bwb_{\mathsf{PA}}(\varphi(\underline{n}))].$$

This predicate is arithmetical; more precisely, it is Σ_3 because of the $\exists$-quantifier hidden in bwb_{PA}. We read $1bwb_{\mathsf{PA}}(\alpha)$ as "α is 1-provable." Let $\mathtt{1bwb}(x)$ be the Σ_3-formula in $\mathcal{L}_{ar}$ defining $1bwb_{\mathsf{PA}}$. Write $\boxed{1}\alpha$ for $\mathtt{1bwb}(\ulcorner\alpha\urcorner)$ and $\diamondsuit\alpha$ for $\neg\boxed{1}\neg\alpha$. As we know, $\Box\alpha$ for $\alpha \in \mathcal{L}^0_{ar}$ can be read '$\mathsf{PA} + \neg\alpha$ is inconsistent', while $\boxed{1}\alpha$, by Lemma 5.1, formalizes '$\mathsf{PA} + \neg\alpha$ is ω-inconsistent'. Therefore, $\diamondsuit\top$ ($\equiv \neg\boxed{1}\bot$) means 'PA ($= \mathsf{PA} + \neg\bot$) is ω-consistent'. This explains the interest in the operator $\boxed{1}$.

If $bwb_{\mathsf{PA}}(\alpha)$ then certainly $1bwb_{\mathsf{PA}}(\alpha)$ (choose α for φ). The italicized statement is reflected in PA as '$\vdash_{\mathsf{PA}} \Box\alpha \to \boxed{1}\alpha$ for every $\alpha \in \mathcal{L}^0_{ar}$'. The converse implication fails, because we know, $\nvdash_{\mathsf{PA}} \mathtt{Con}_{\mathsf{PA}}$, while $\mathtt{Con}_{\mathsf{PA}}$ is easily 1-provable. Indeed, with $\varphi(x) := \neg\,\mathtt{bew}_{\mathsf{PA}}(x, \bot)$ holds $\vdash_{\mathsf{PA}} \varphi(\underline{n})$ for all n, and trivially $\vdash_{\mathsf{PA}} \forall x\varphi(x) \to \mathtt{Con}_{\mathsf{PA}}$.

Define $\Omega := \{\forall x\varphi \mid \varphi = \varphi(x)\ \&\ \vdash_{\mathsf{PA}} \varphi(\underline{n}) \text{ for all } n\}$, and $\mathsf{PA}^\Omega := \mathsf{PA}{+}\Omega$. According to its definition, Ω and hence also PA^Ω are formally Σ_3. As Theorem 5.2 will show, PA^Ω is properly Σ_3, and therefore no longer recursively axiomatizable.

Lemma 5.1. *The following properties are equivalent for $\alpha \in \mathcal{L}^0_{ar}$:*

(i) $1bwb_{\mathsf{PA}}(\alpha)$, (ii) $\vdash_{\mathsf{PA}^\Omega} \alpha$, (iii) $\mathsf{PA} + \neg\alpha$ *is ω-inconsistent.*

Proof. (i)⇒(ii) follows with a glance at the definitions (read (i) naively). (ii)⇒(iii): Let $\vdash_{\mathsf{PA}^\Omega} \alpha$. Since Ω is closed under conjunctions, there is some $\forall x\varphi(x) \in \Omega$ with $\forall x\varphi \vdash_{\mathsf{PA}} \alpha$, hence $\vdash_{\mathsf{PA}} \neg\alpha \to \exists x\neg\varphi$ and so $\vdash_{\mathsf{PA}+\neg\alpha} \exists x\neg\varphi$. Now, $\forall x\varphi \in \Omega$, therefore $\vdash_{\mathsf{PA}} \varphi(\underline{n})$ and a fortiory $\vdash_{\mathsf{PA}+\neg\alpha} \varphi(\underline{n})$, for all n. Thus, $\mathsf{PA} + \neg\alpha$ is ω-inconsistent. (iii)⇒(i): Let $\vdash_{\mathsf{PA}+\neg\alpha} \beta(\underline{n})$ for all n, but $\vdash_{\mathsf{PA}+\neg\alpha} \exists x\neg\beta$. Then $\vdash_{\mathsf{PA}} \forall x\beta \to \alpha$. With $\varphi(x) := \neg\alpha \to \beta(x)$ clearly $\vdash_{\mathsf{PA}} \varphi(\underline{n})$ for all n. Now, $\forall x\varphi \equiv \alpha \vee \forall x\beta \vdash_{\mathsf{PA}} \alpha$. Hence $\vdash_{\mathsf{PA}} \forall x\varphi \to \alpha$. Thus, altogether $1bwb_{\mathsf{PA}}(\alpha)$. ❏

Theorem 5.2 (the 1-provable Σ_3-completeness of PA). *All true Σ_3-sentences are 1-provable. Moreover, for every β of this kind, $\vdash_{\mathsf{PA}} \beta \to \boxed{1}\beta$.*

Proof. Let $\mathcal{N} \vDash \beta := \exists x \forall y \gamma(x,y)$ where $\gamma(x,y)$ is Σ_1. Then there is some m such that $\mathcal{N} \vDash \gamma(\underline{m},\underline{n})$ for all n. Therefore, $\vdash_{\mathsf{PA}} \gamma(\underline{m},\underline{n})$ for all n, because PA is Σ_1-complete. Hence, $\forall y \gamma(\underline{m},y) \in \Omega$ and so $\vdash_{\mathsf{PA}^\Omega} \exists x \forall y \gamma$, or equivalently, $1bwb_{\mathsf{PA}}(\beta)$ by Lemma 5.1. Because of the provable Σ_1-completeness of PA, this argumentation is comprehensible in PA, so that also $\vdash_{\mathsf{PA}} \beta \to \boxed{1}\beta$. ❑

$D1$–$D4$ are also valid for the operator $\boxed{1} : \mathcal{L}^0_{ar} \to \mathcal{L}^0_{ar}$. Indeed, $D1$ holds because $\vdash_{\mathsf{PA}} \alpha \Rightarrow \vdash_{\mathsf{PA}} \Box\alpha \Rightarrow \vdash_{\mathsf{PA}} \boxed{1}\alpha$, and $D2$ formalizes '$\vdash_{\mathsf{PA}^\Omega} \alpha, \alpha \to \beta \Rightarrow \vdash_{\mathsf{PA}^\Omega} \beta$' in view of Lemma 5.1. $D3$ is an application of Theorem 5.2 with $\beta = \boxed{1}\alpha$. The proof of $D4$ in **7.2** uses, along with the fixed-point lemma, only $D1$–$D3$; so $D4$ holds as well. Therefore, nearly everything said in **7.2** on $\Box$ applies also to $\boxed{1}$; in particular Theorem 2.2, which now reads $\nvdash_{\mathsf{PA}} \neg\boxed{1}\bot$ ($\equiv \Diamond\top$). To put it more concisely, although the consistency of PA is provable with the extended means, ω-consistency is not. Hence, this property, which has a Π_3-Definition according to Exercise 4 in **6.7**, cannot be Σ_3 by Theorem 5.2, and must therefore be properly Π_3.

Alongside $\Box\alpha \to \boxed{1}\alpha$, there are other noteworthy interactions between $\Box$ and $\boxed{1}$, in particular $\vdash_{\mathsf{PA}} \neg\Box\alpha \to \boxed{1}\neg\Box\alpha$. This formalizes 'If $\nvdash_{\mathsf{PA}} \alpha$ then $\neg\Box\alpha$ is 1-provable'. To verify the latter notice that $\nvdash_{\mathsf{PA}} \alpha$ implies $\vdash_{\mathsf{PA}} \varphi(\underline{n})$ for all n, where $\varphi(x)$ is $\neg\,\mathsf{bew}_{\mathsf{PA}}(x, \ulcorner\alpha\urcorner)$, and since $\vdash_{\mathsf{PA}} \forall x\varphi \to \neg\Box\alpha$, we obtain $\vdash_{\mathsf{PA}} \boxed{1}\neg\Box\alpha$. On the other hand, $\vdash_{\mathsf{PA}} \neg\Box\alpha \to \Box\neg\Box\alpha$ is false in general; see Example 2(c) on page 223.

The language of the bimodal propositional logic GD now to be defined results from $\mathfrak{F}_\Box$ by adding a further connective $\boxed{1}$ to $\mathfrak{F}_\Box$, which is treated syntactically just as $\Box$. The axioms of GD are those of G both for $\Box$ and $\boxed{1}$, augmented by the axioms

$$\Box p \to \boxed{1}p \quad \text{and} \quad \neg\Box p \to \boxed{1}\neg\Box p.$$

The rules of GD are the same as those for G. Insertions $\imath$ to $\mathcal{L}^0_{ar}$ are defined as in **7.4**, but with the addional clause $(\boxed{1}H)^\imath = \boxed{1}H^\imath$ $\big(= 1bwb_{\mathsf{PA}}(\ulcorner H^\imath \urcorner)\big)$. By the reasoning above, all axioms and rules of GD are sound. This proves (the easier) half of the following remarkable theorem from *Dzhaparidze* (1985):

Theorem 5.3. *$\vdash_{\mathsf{GD}} H \Leftrightarrow \vdash_{\mathsf{PA}} H^\imath$ for all insertions $\imath$. Further,* GD *is decidable.*

Thus, GD completely captures the interaction between bwb_{PA} and $1bwb_{\mathsf{PA}}$; also Theorem 4.5 carries over. However, GD no longer has an adequate Kripke semantics, which complicates the decision procedure. For further references see [Boo], [Be3] .

As an exercise, the reader should derive $\boxed{1}(\Box p \to p)$ from the axioms of GD. Thus, $\vdash_{\mathsf{PA}} \boxed{1}(\Box\alpha \to \alpha)$ for *every* $\alpha \in \mathcal{L}^0_{ar}$, while $\vdash_{\mathsf{PA}} \Box(\Box\alpha \to \alpha)$ does hold only if $\vdash_{\mathsf{PA}} \alpha$. In other words, the *local reflection principle* $\{\Box\alpha \to \alpha \mid \alpha \in \mathcal{L}^0_{ar}\}$ is 1-provable in PA. *Be careful*: GD expands G conservatively, so that $\nvdash_{\mathsf{GD}} \Box p \to p$.

7.6 Modal Operators in ZFC

Considerations regarding self-reference in ZFC are technically more easy, but from the foundational point of view more involved because there is no superordinate theory. If ZFC is consistent as we assume it is, then $\mathtt{Con}_{\mathsf{ZFC}}$ is a true arithmetical statement but is not provable in ZFC. Thus, true arithmetical statements may even be unprovable in ZFC. It makes sense, therefore, to consider $\mathsf{ZFC}^+ := \mathsf{ZFC} + \mathtt{Con}_{\mathsf{ZFC}}$, because after all, we want set theory to embrace as many facts about numbers and sets as possible from which interesting consequences may result.

As **7.2** shows, the consistency assumption for ZFC alone does not guarantee that ZFC^+ is consistent. The second incompleteness theorem excludes $\vdash_{\mathsf{ZFC}} \mathtt{Con}_{\mathsf{ZFC}}$ but does not preclude the possibility $\vdash_{\mathsf{ZFC}} \mathtt{Con}_{\mathsf{ZFC}} \to \mathtt{Con}_{\mathsf{ZFC}^+}$. But then $\vdash_{\mathsf{ZFC}^+} \mathtt{Con}_{\mathsf{ZFC}^+}$, and ZFC^+ would be inconsistent. From certain assumptions regarding the existence of large cardinals, the consistency of ZFC^+ follows fairly easily. These assumptions would then have to be jettisoned, and in the framework of ZFC the consistency of ZFC would no longer have its external sense; although consistent, ZFC would then prove along with true arithmetical facts also false ones. This sounds strange, but there is no convincing argument that this cannot be so.

Even if $\nvdash_{\mathsf{ZFC}} \neg\, \mathtt{Con}_{\mathsf{ZFC}}$, it may still be that one of the sentences from the sequence $\Box\neg\, \mathtt{Con}_{\mathsf{ZFC}}, \Box\Box\neg\, \mathtt{Con}_{\mathsf{ZFC}}, \ldots$ is provable in ZFC. We exclude the latter only if we assume that the ω-iterated consistency extension ZFC^ω is consistent, i.e., $\nvdash_{\mathsf{ZFC}} \Box^n\bot$ for all n (see page 220), so that by Theorem 4.3 G would be the provability logic of ZFC. In fact, $(\forall n{\in}\mathbb{N})\ \nvdash_{\mathsf{ZFC}} \Box^n\bot$ is equivalent to G's being the provability logic of ZFC, by the general Theorem 6.1 below. Therein $Rf_T := \{\Box\alpha \to \alpha \mid \alpha \in \mathcal{L}^0\}$ denotes the already encountered local reflection principle. Theorem 4.3 is also a corollary of the theorem, since $(\forall n{\in}\mathbb{N})\ \nvdash_T \Box^{n+1}\bot$ is equivalent to the consistency of T^ω.

Theorem 6.1. *For a sufficiently expressive theory T* [6] *the following are equivalent:* (i) *T^ω is consistent,* (ii) *$T + Rf_T$ is consistent,* (iii) *G is the provability logic of T.*

Proof. (i)⇒(ii) indirect: Suppose $T+Rf_T$ is inconsistent. Then there are $\alpha_0, \ldots, \alpha_n$ such that $\vdash_T \neg\varphi$, $\varphi := \bigwedge_{i=1}^n(\Box\alpha_i \to \alpha_i)$. Hence $\vdash_T \Box\neg\varphi \equiv_T \neg\Diamond\varphi$. Now, because $\vdash_{T^\omega} \neg\Box^{n+1}\bot$, by Example (d) in **7.3** and Lemma 4.1, we get $\vdash_{T^\omega} \Diamond R_n^i$ $(p_i^i = \alpha_i)$. Clearly, $R_n^i = \varphi$ and so $\vdash_{T^\omega} \Diamond\varphi$. Since also $\vdash_{T^\omega} \neg\Diamond\varphi$, T^ω is inconsistent. (ii)⇒(iii): The proof of Theorem 4.2 for PA, as presented in [Boo], runs nearly the same for T, because PA is transgressed in one place only: one uses the fact that $\mathcal{N} \vDash Rf_{\mathsf{PA}}$. However, the existence of a corresponding T-model is ensured by (ii). (iii)⇒(i): $\nvdash_{\mathsf{G}} \Box^{n+1}\bot$, hence $\nvdash_T \Box^{n+1}\bot \equiv_T \neg\, \mathtt{Con}_{T^n}$ for all n, and so T^ω is consistent. ❑

[6] By such a T we mean that the proof steps of Theorem 4.2 that do not transgress PA, can be carried out in T, which does not yet imply the provability of the theorem itself.

The equivalence (i)⇔(ii) is a purely proof-theoretical one and called *Goryachev's Theorem*; see [Gor] or [Be2]. We obtained it using essentially some modal logic. For $T = \mathsf{ZFC}$, perhaps a bit more interesting than (i) or (ii) is the assumption

$(*)$ $\quad \vdash_{\mathsf{ZFC}} (\exists x{\in}\omega)\varphi \Rightarrow \nvdash_{\mathsf{ZFC}} \neg\varphi(\underline{n})$ for some $n \quad (\varphi(x) \in \mathcal{L}_\in)$,

the *ω-consistency* of ZFC. It implies $D1^*$ which in turn ensures $\nvdash_{\mathsf{ZFC}} \Box^{n+1}\bot$, that is, (i) and hence all other conditions in Theorem 6.1 hold for ZFC.

Remark. It is worthwhile to notice that the consistency of $\mathsf{ZFC} + Rf_{\mathsf{ZFC}}$ and thereby the proof of Solovay's completeness theorem for ZFC follows directly from $(*)$, without appealing to Goryachev's theorem. What is needed to see this is the following

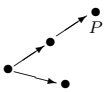
Lemma. *If* ZFC *is ω-consistent then there exists a model* $\mathcal{V} \vDash \mathsf{ZFC}$ *with* $\mathcal{V} \vDash Rf_{\mathsf{ZFC}}$.

Proof. Let $\Omega := \{(\forall x \in \omega)\alpha \mid \alpha = \alpha(x), \vdash_{\mathsf{ZFC}} \alpha(\underline{n}) \text{ for all } n\}$. Then $\mathsf{ZFC} + \Omega$ is consistent. Otherwise $\vdash_{\mathsf{ZFC}} \neg(\forall x \in \omega)\alpha \equiv (\exists x \in \omega)\neg\alpha$ for some $(\forall x \in \omega)\alpha \in \Omega$ because Ω is closed under conjunction, in contradiction to $(*)$. Any $\mathcal{V} \vDash \mathsf{ZFC} + \Omega$ satisfies the reflection principle Rf_{ZFC} as well, for if $\mathcal{V} \nvDash \alpha$ then $\nvdash_{\mathsf{ZFC}} \alpha$ and therefore $\vdash_{\mathsf{ZFC}} \neg\,\mathtt{bew}_{\mathsf{ZFC}}(\underline{n}, \ulcorner\alpha\urcorner)$ for all n. That means $(\forall y \in \omega)\neg\,\mathtt{bew}_{\mathsf{ZFC}}(y, \ulcorner\alpha\urcorner) \in \Omega$, which clearly implies $\mathcal{V} \nvDash \Box\alpha$.

Now we interpret the modal operator $\Box$ no longer as *provable in* ZFC which is equivalent to *valid in all* ZFC *models*, but rather as *valid in particular classes of* ZFC*-models.* For the following undefined notions we refer to [Ku]. Particularly interesting is a *transitive* model. This is a model $\mathcal{V} = (V, \in^{\mathcal{V}}) \vDash \mathsf{ZFC}$, where the set V is *transitive* (i.e., $a \in b \in V \Rightarrow a \in V$). Then $\in^{\mathcal{V}}$ is the usual $\in$-relation restricted to V, a set in our metatheory (which itself is essentially ZFC). We write V for $\mathcal{V}$. Like any set, V has an ordinal rank, denoted by ρV, and $\rho U < \rho V$ whenever $U \in V$. To prove the soundness half of Theorem 6.3 we use the following

Lemma 6.2. ([JK]) *Let* V, W *be transitive models such that* $\rho V < \rho W$ *and suppose* $V \vDash \alpha$. *Then* $W \vDash$ '*there is a transitive model* U *with* $U \vDash \alpha$'.[7]

Let Gi result from augmenting G by the axiom $\Box(\Box p \to \Box q) \vee \Box(\Box q \to p)$. In the same sense that G is complete with respect to all finite partial orders, Gi is complete with respect to all *preference orders.* This is a finite (strict) partial order $(g, <)$ for which there is some $h : g \to \mathbb{N}$ with $P < Q \;\Leftrightarrow\; hP < hQ$, for all $P, Q \in g$. As for G, the finite model property ensures the decidability of Gi. The figure shows a partial order, which is easily seen not to be a preference order and in which the adjoined axiom is easily refuted choosing $wp = \{P\}$ and $wq = \emptyset$. Thus, the additional axiom does not belong to G; hence $\mathsf{Gi} \supset \mathsf{G}$. We mention that in [So] and [Boo] a somewhat more complex axiom is considered.



[7] In transitive models W the sentence in ' ' (which with some encoding can be formulated in $\mathcal{L}_\in$) is absolute, and therefore equivalent to the existence of a transitive model $U \in W$ with $U \vDash \alpha$. The latter is much stronger than the consistency assumption of ZFC, but for the direction in which the proof of the theorem is to go the stronger assumption is not needed.

We define insertions $\imath : \mathcal{F}_\Box \to \mathcal{L}^0_\in$ as in **7.4** with the clause $(\Box H)^\imath = \Box H^\imath$, where $\Box\alpha$ for $\alpha = H^\imath \in \mathcal{L}^0_\in$ is now to mean 'α is valid in all transitive models', more precisely, the formalization of this property in the language $\mathcal{L}_\in$. Accordingly, $\Diamond\alpha = \neg\Box\neg\alpha$ states 'it is not the case that in all transitive models holds $\neg\alpha$', or in equivalent terms, 'α holds in some transitive model'.

Theorem 6.3. $\vdash_{\mathsf{Gi}} H$ *iff* $\vdash_{\mathsf{ZFC}} H^\imath$ *for all insertions* $\imath$.

We prove only the direction $\Rightarrow$, that is, soundness. As regards the axioms of Gi, since $\Box p \to \Box\Box p$ is provable from the other axioms of G (see **7.3**), it suffices to prove

$$\text{(A) } \Box(\alpha \to \beta) \wedge \Box\alpha \vdash_{\mathsf{ZFC}} \Box\beta, \quad \text{(B) } \Box(\Box\alpha \to \alpha) \vdash_{\mathsf{ZFC}} \Box\alpha,$$
$$\text{(C) } \vdash_{\mathsf{ZFC}} \Box(\Box\alpha \to \Box\beta) \vee \Box(\Box\beta \to \alpha), \quad \text{for all } \alpha, \beta \in \mathcal{L}^0_\in .$$

(A) is trivial, because the sentences valid in any class of models are closed under MP. (B) is equivalent to (B′) $\Diamond\neg\alpha \vdash_{\mathsf{ZFC}} \Diamond(\Box\alpha \wedge \neg\alpha)$. Here is the proof: if a transitive model exists in which $\neg\alpha$ holds, then there is also one with minimal rank, V say. We claim that $V \vDash \Box\alpha$. Otherwise $V \vDash \Diamond\neg\alpha$, and hence there would be a transitive model $U \in V$ with $U \vDash \neg\alpha$ and $\rho U < \rho V$, contradicting our choice of V. This proves $V \vDash \Diamond(\Box\alpha \wedge \neg\alpha)$ and verifies (B′). Finally, (C) is verified by contraposition: suppose there are transitive models V, W and sentences α, β such that

(a) $V \vDash$ 'α holds in all transitive models and in some transitive model holds $\neg\beta$',

(b) $W \vDash$ 'β holds in all transitive models', (c) $W \vDash \neg\alpha$.

From these assumptions it follows first of all that $\rho W < \rho V$. For suppose by (a) that $U \in V$ is a transitive model for $\neg\beta$. If $\rho V \leqslant \rho W$ then $\rho U < \rho W$. Hence, by Lemma 6.2, $W \vDash$ 'there is a transitive model for $\neg\beta$', contradicting (b). Now, since $W \vDash \neg\alpha$ by (c) and because of $\rho W < \rho V$, in V holds 'there is some transitive model for $\neg\alpha$' by Lemma 6.2, in contradiction to (a). This proves (C). For the substitution rule, soundness follows as for G in **7.4**. MN is trivially sound, because if α is provable in ZFC then of course α is valid in all transitive models.

Another interesting model-theoretical interpretation of $\Box\alpha$ is 'α is valid in all V_κ'. Here κ runs through all inaccessible cardinal numbers. The adequate modal logic for this interpretation of $\Box$ is $\mathsf{Gj} := \mathsf{G} + \Box(\Box p \wedge p \to q) \vee \Box(\Box q \to p)$ according to [So] (provided there are sufficiently many inaccessibles). This logic, often denoted by G.3, is complete with respect to all finite strict linear orders, which of course are also

frames for Gi, so that $\mathsf{Gi} \subseteq \mathsf{Gj}$. The figure shows a Gi-frame on which the additional axiom is easily refuted with $wp = \{P\}$ and $wq = \emptyset$, hence it is not a Gj-frame and so $\mathsf{Gi} \subset \mathsf{Gj}$. As usual, the finite model property of Gj implies its decidability. This modal logic is sound for the above interpretation of $\Box$, and we recommend that the advanced reader carry out the proof, which is similar to that of Gi. All one needs to know besides Lemma 6.2 is that V_κ is a transitive model and that $V_\kappa \in V_\lambda$ or $V_\lambda \in V_\kappa$, for arbitrary inaccessible cardinals $\kappa \neq \lambda$.

Hints to the Exercises

Section 1.1

1. (a): Note that x_k is a fictional variable in f iff $a_k = 0$. (b): Because of the uniqueness, 2^{n+1} (= number of subsets of $\{0, \dots, n\}$) is the number sought.

2. Proof by formula induction on φ. Consider the property $\mathcal{E}$: 'ξ is a prime formula or there are $\alpha, \beta \in \mathcal{F}$ with $\xi = \neg\alpha$ or $\xi = (\alpha \wedge \beta)$ or $\xi = (\alpha \vee \beta)$'.

3. Verify by induction on φ the stronger property that no proper initial segment of φ is a formula nor can φ be an initial segment of a formula. Let for instance $\varphi = \neg\alpha$ (Exercise 2). A proper initial segment of $\neg\alpha$ is either the one-element string $\neg$ or a proper initial segment of α.

4. Assume $(\alpha \circ \beta) = (\alpha' \circ' \beta')$, hence $\alpha \circ \beta = \alpha' \circ' \beta'$. Assume $\alpha \neq \alpha'$. Then α is a proper initial segment of α' or conversely. This is impossible according to Exercise 3. Consequently $\alpha = \alpha'$, hence $\circ = \circ'$ and $\beta = \beta'$.

Section 1.2

2. $\neg p \equiv p + 1,\ 1 \equiv p + \neg p,\ p \leftrightarrow q \equiv p + \neg q$, and $p + q \equiv p \leftrightarrow \neg q$.

3. Induction on α shows that $\alpha^{(n)}$ is monotonic; for if $f, g \in \boldsymbol{B}$ are monotonic then so is $\vec{a} \mapsto f\vec{a} \circ g\vec{a}$, $\circ \in \{\wedge, \vee\}$. Converse: Induction on the arity. If $f \in \boldsymbol{B}_{n+1}$ is monotonic then also $f_k : \vec{x} \mapsto f(\vec{x}, k)$ for $k = 0, 1$. Let f_k be represented by α_k ($k = 0, 1$, induction hypothesis). Then f is represented by $\alpha_0 \vee \alpha_1 \wedge p_{n+1}$.

4. A not representable f is not monotonic by Exercise 3. But then a suitable instantiation of constants for all but one argument of φ easily yields negation.

Section 1.3

1. (a): MP easily yields $p \to q \to r,\ p \to q,\ p \vDash r$. Apply (D) three times.

2. With the deduction theorem one easily verifies $(\alpha \to \beta) \to (\gamma \to \alpha) \to (\gamma \to \beta)$.

5. $X \vdash \bar{X} \vdash \alpha \Rightarrow X \vdash \alpha \Rightarrow \alpha \in \bar{X}$. Thus, $\bar{X}$ is deductively closed.

Section 1.4

1. $X \cup \{\neg\alpha \mid \alpha \in Y\} \vdash \bot \Rightarrow X \cup \{\neg\alpha_0, \dots, \neg\alpha_n\} \vdash \bot$ for some $\alpha_0, \dots, \alpha_n \in Y$. This yields $X \vdash (\bigwedge_{i \leqslant n} \neg\alpha_i) \to \bot$, or equivalently, $X \vdash \bigvee_{i \leqslant n} a_i$.

2. Supplement Lemma 4.4 by the proof of $X \vdash \alpha \vee \beta \ \Leftrightarrow\ X \vdash \alpha$ or $X \vdash \beta$.

3. Let $X \nvdash \varphi$, $X \vdash' \varphi$, say, and $Y \supseteq X \cup \{\neg\varphi\}$ be maximally consistent in $\vdash$. Further define σ by $p^\sigma = \top$ for $p \in Y$ and $p^\sigma = \bot$ otherwise. Simultaneous induction on $\alpha, \neg\alpha$ shows that $\alpha \in Y \Rightarrow \vdash \alpha^\sigma$ and $\alpha \notin Y \Rightarrow \vdash \neg\alpha^\sigma$. Hence $\vdash \neg\varphi^\sigma$. Thus $\vdash' \neg\varphi^\sigma$, and so $X^\sigma \vdash' \neg\varphi^\sigma$. But $X \vdash' \varphi$, therefore $X^\sigma \vdash' \varphi^\sigma$. Thus, $X^\sigma \vdash' \alpha$ for all α according to $(\neg 1)$. Hence $\vdash'$ is inconsistent and so $\vdash$ is maximal.

4. There is a smallest consequence relation with the properties $(\wedge 1)$ – $(\neg)$, namely the calculus $\vdash$ of this section. Since $\vdash \subseteq \models$ and $\vdash$ is already maximal according to Exercise 3, $\vdash$ and $\models$ must coincide.

Section 1.5

1. Add to the formulas in Example 1 the set of formulas $\{p_{ab} \mid a \leqslant_0 b\}$.

2. $\Rightarrow$: Assume $M, N \notin F$. Then $\neg M, \neg N \in F$; hence $\setminus(M \cup N) = \setminus M \cap \setminus N \in F$. Therefore $M \cup N \notin F$. $\Leftarrow$: $M \in F$ implies $M \cup N \in F$ by condition (b).

3. $\Rightarrow$: Let U be trivial. Then $E \in U$ for some finite $F \subseteq I$. Let $E = E_1 \cup \{i\}$ for some $i \in E$ so that $E_1 \in U$ or $\{i\} \in U$ (cf. Exercise 2). If $\{i\} \in U$ we are done. Otherwise replace E by the smaller E_1 and repeat the argument. This consideration leads to $\{i_0\} \in U$ for some $i_0 \in I$. $\Leftarrow$: proved already in the text.

Section 1.6

1. First verify the deduction theorem, which holds for each calculus with MP as the only rule and A1, A2 among the axioms; cf. Lemma 6.3. X is consistent iff $X \nvdash \bot$, for $X \vdash \bot$ implies $X \vdash (\alpha \to \bot) \to \bot = \neg\neg\alpha$ by A1, hence $X \vdash \alpha$ by A3. Now prove $X \vdash \alpha \to \beta$ iff $X \vdash \alpha \Rightarrow X \vdash \beta$ for maximally consistent X. This allows you to proceed along the lines of Lemma 4.5 and Theorem 4.6.

2. Apply Zorn's lemma on the partially ordered set $H := \{Y \supseteq X \mid Y \nvdash \alpha\}$.

3. (a): Such a X satifies $(*)$: $X \vdash \varphi \to \alpha$ for all α. For otherwise $X, \varphi \to \alpha \vdash \varphi$, hence $X \vdash (\varphi \to \alpha) \to \varphi$, and so $X \vdash \varphi$ by Peirce's axiom. Suppose $\alpha \notin X$. Then $X, \alpha \vdash \varphi, \varphi \to \beta$ by $(*)$, and so $X, \alpha \vdash \beta$. This confirms (a). (b): With (a) follows $X \vdash \alpha \to \beta$ iff $X \vdash \alpha \Rightarrow X \vdash \beta$. Proceed with an adaptation of Lemma 4.5.

4. Prove (m) by induction. For example, if $\alpha \vdash \alpha\beta$ then $\alpha\gamma \vdash \gamma\alpha \vdash \gamma\alpha\beta \vdash \alpha\beta\gamma$. Although by (4) and (5) no parantheses in $\alpha\beta\gamma$ are needed, it is tricky to prove $\alpha(\beta\gamma)\delta \vdash (\alpha\beta)\gamma\delta$. (M) implies $(*)$: $X, \alpha \vdash \gamma$ & $X, \beta \vdash \gamma \Rightarrow X, \alpha\beta \vdash \gamma$, because $X, \alpha \vdash \gamma \Rightarrow X, \alpha\beta \vdash \gamma\beta \vdash \beta\gamma$ and $X, \beta\gamma \vdash \gamma\gamma \vdash \gamma$, therefore $X, \alpha\beta \vdash \gamma$. From $(*)$ follows $\binom{*}{*}$: $X \vdash \alpha\beta \Leftrightarrow X \vdash \alpha$ or $X \vdash \beta$, provided X is φ-maximal, for note that $X \nvdash \alpha$ & $X \nvdash \beta \Rightarrow X, \alpha \vdash \varphi$ & $X, \beta \vdash \varphi \Rightarrow X, \alpha\beta \vdash \varphi \Rightarrow X \nvdash \alpha\beta$. Having $\binom{*}{*}$ one may proceed with a slight modification of Lemma 4.5.

Section 2.1

1. There are 10 essentially binary Boolean functions f. The corresponding algebras $(\{0,1\}, f)$ split into 5 pairs of isomorphic ones, e.g. $(\{0,1\}, \wedge) \simeq (\{0,1\}, \vee)$.

4. Let r and f be unary. Then $ra \Rightarrow ra_j \Rightarrow rha$, and $hfa = h(fa_i)_{i\in I} = fa_j = fha$.

Section 2.2

1. A terminal segment of $f\vec{t}$ has the form $t'_k t_{k+1} \cdots t_n$ (t'_k a terminal segment of t_k).

2. (a): Define $W(\zeta) = 1$ if ζ is a variable or constant, $W(\zeta) = 1 - n$ if ζ is a n-ary function symbol and expand W to all strings of symbols ζ involved in building terms by $W(\zeta_1 \cdots \zeta_n) = \sum_{i=1}^{n} W(\zeta_i)$. Show by term induction that $W(t) = 1$ for all terms t, so that $W(t_1 \cdots t_n) = n$. (b): If not, t would be a concatenation of at least two terms by Exercise 1, which is impossible by (a). (c) derives from (b). (d): $t_1 \neq t'_1$ yields a contradiction to (b).

Section 2.3

1. Let $\mathcal{M}^{cd}_{xy} \vDash \alpha \wedge \alpha\,\frac{y}{x}$ ($c \neq d$). Then for each a there is some $b \neq a$ with $\mathcal{M}^{b}_{y} \vDash \varphi\,\frac{y}{x}$.

3. The Theorems 3.1 and 3.5 yield $\mathcal{A} \vDash \alpha\,[a] \Leftrightarrow \mathcal{A}' \vDash \alpha\,[a] \Leftrightarrow \mathcal{A}' \vDash \alpha_x(\boldsymbol{a})$.

4. (b): $\exists_n \wedge \neg\exists_m$ is for $n \leqslant m$ equivalent to $\bigvee_{k=n}^{m} \exists_{=k}$, and for $n > m$ to $\exists_{=0}(\equiv \bot)$.

Section 2.4

1. $\alpha \equiv \beta \;\Rightarrow \vDash \forall\vec{x}\,(\alpha \leftrightarrow \beta) \;\Rightarrow \vDash (\alpha \leftrightarrow \beta)\,\frac{\vec{t}}{\vec{x}}$ $\big(= \alpha\,\frac{\vec{t}}{\vec{x}} \leftrightarrow \beta\,\frac{\vec{t}}{\vec{x}}\big)$.

4. $\exists x(Px \to \forall y Py) \equiv \forall x Px \to \forall y Py$ according to (10).

Section 2.5

2. Observe that $S \vDash \varphi \to \beta \;\Leftrightarrow\; S, \varphi \vDash \beta$, and (e) page 62.

3. Prove first $T_\alpha = \{\beta \in \mathcal{L}^0 \mid T, \alpha \vdash \beta\}$ is a theory. Then show that $T_\alpha = T + \alpha$.

Section 2.6

1. Follow the proof of Theorem 6.1 (observe that $y = f\vec{t} \equiv_{T_f} \delta_f(\vec{t}, y)$). Hint for the "only if" part: $y \equiv f\vec{t} \equiv_{T_f} \delta_f(\vec{t}, y)$, and $T_f \vDash \forall\vec{x}\exists!y\, y \equiv f\vec{x} \to \forall\vec{x}\exists!y\delta(\vec{x}, y)$.

2. $\mathcal{N} \vDash x = 0 \leftrightarrow \forall y\, x \neq \mathtt{S}y$. An elementary calculation confirms the (quantifier-free) definition $x + y \equiv z \leftrightarrow \mathtt{S}(x \cdot z) \cdot \mathtt{S}(y \cdot z) \equiv \mathtt{S}(z^2 \cdot \mathtt{S}(x \cdot y))$, where $z^2 := z \cdot z$.

3. Let $xy = xz = e$ ($\circ$ and $\vDash$ not written) and choose some y' with $yy' = e$. Then $x = x(yy') = (xy)y' = y'$, hence $yx = e$. $zx = e$ is proved analogously. This yields $y = y(xz) = (yx)z = ez = (zx)z = z(xz) = ze = z$.

4. Were $<$ definable then $<$ would be invariant under automorphisms of $(\mathbb{Z}, 0, +)$. This is not the case for the automorphism $n \mapsto -n$ (Padoa's method).

Section 3.1

1. Let $X \vdash \alpha\,\frac{t}{x}$. Then $X, \forall x\neg\alpha \vdash \alpha\,\frac{t}{x}, \neg\alpha\,\frac{t}{x}$. Hence $X, \forall x\neg\alpha \vdash \exists x\alpha$. Certainly also $X, \neg\forall x\neg\alpha \vdash \exists x\alpha$ (since $\exists x\alpha = \neg\forall x\neg\alpha$). Thus $X \vdash \exists x\alpha$ according to $(\neg 2)$.

2. Let $\alpha' := \alpha\,\frac{y}{x}$, $u \notin var\,\alpha$, $u \neq y$. Then $\forall x\alpha \vdash \alpha'\,\frac{u}{y}$ $(= \alpha\,\frac{u}{x})$ by $(\forall 1)$. Hence we obtain $\forall x\alpha \vdash \forall y\alpha'$ $(= \forall y\alpha\,\frac{y}{x})$ by $(\forall 2)$.

Section 3.2

1. Theorem 2.6 and Exercise 4 in **3.1**.

2. First verify $t^{\mathfrak{T}} = t$ by induction on t. Next prove by induction on $\wedge, \neg$ $(*)$ $\mathfrak{T} \vDash \forall\vec{x}\varphi \Leftrightarrow \mathfrak{T} \vDash \varphi\,\frac{\vec{t}}{\vec{x}}$ for all $\vec{t} \in \mathcal{T}^n$ (φ open). Let $\mathcal{M} \vDash X$. Then clearly $\mathcal{M} \vDash \tilde{X} := \{\varphi\,\frac{\vec{t}}{\vec{x}} \mid \forall\vec{x}\varphi \in X,\ \vec{t} \in \mathcal{T}^n\}$. Finally prove $\mathcal{M} \vDash \varphi \Leftrightarrow \mathfrak{T} \vDash \varphi$, for all open φ (induction on $\wedge, \neg$). Thus, $\mathfrak{T} \vDash \tilde{X}$ and so $\mathfrak{T} \vDash X$ according to $(*)$.

3. Theorem 2.7 and the finiteness theorem for $\vdash$.

Section 3.3

1. Prove $\forall z\, x + (y + z) = (x + y) + z$ in PA by induction on z, then $\forall y\, x + \mathsf{S}y = \mathsf{S}x + y$ and $\forall y\, x + y = y + x$ by induction on y. Quantify free variables at the end.

2. Informally: $x < y$ implies $\exists z\, \mathsf{S}z + x = y$. Therefore $\exists z\, z + \mathsf{S}x = y$. The converse $\mathsf{S}x \leqslant y \to x < y$ is clear since $\vdash_{\mathsf{PA}} x < \mathsf{S}x$. Connexity: The induction hypothesis may be written as $x < y \vee y \leqslant x$. If $x < y$ then $\mathsf{S}x \leqslant y$, hence $\mathsf{S}x \leqslant y \vee y \leqslant \mathsf{S}x$ (induction claim). We get the same in case $y \leqslant x$, since then also $y \leqslant \mathsf{S}x$.

3. (a): We have to prove that $\forall x(\varphi \to \alpha) \vdash_{\mathsf{PA}} \forall x\alpha$, where $\varphi := (\forall y{<}x)\alpha\,\frac{y}{x}$. By Exercise 2, $y{<}\mathsf{S}x \equiv_{\mathsf{PA}} y{\leqslant}x$. Thus, $\varphi, \forall x(\varphi \to \alpha) \vdash_{\mathsf{PA}} \varphi \wedge \alpha \vdash_{\mathsf{PA}} (\forall y{<}\mathsf{S}x)\alpha\,\frac{y}{x} = \varphi\,\frac{\mathsf{S}x}{x}$. Therefore $\forall x(\varphi \to \alpha) \vdash_{\mathsf{PA}} \forall x(\varphi \to \varphi\,\frac{\mathsf{S}x}{x})$. Trivially also $\forall x(\varphi \to \alpha) \vdash_{\mathsf{PA}} \varphi\,\frac{0}{x}$. This yields $\forall x(\varphi \to \alpha) \vdash_{\mathsf{PA}} \forall x\varphi \vdash_{\mathsf{PA}} \forall x\varphi\,\frac{\mathsf{S}x}{x} \vdash_{\mathsf{PA}} \forall x\alpha$ by IS. (b): Follows from (a) by contraposition. (c): For $\varphi := (\forall y{<}x)\exists z\alpha \to \exists u(\forall y{<}x)(\exists z{<}u)\alpha$ clearly holds $\vdash_{\mathsf{PA}} \varphi\,\frac{0}{x}$, and one readily shows that $\varphi \vdash_{\mathsf{PA}} \varphi\,\frac{\mathsf{S}x}{x}$. This yields the claim by IS.

Section 3.4

1. $X = T \cup \{\boldsymbol{v}_i \neq \boldsymbol{v}_j \mid i \neq j\}$ is satisfiable because each finite subset is.

2. $X = Th\mathcal{A} \cup \{\boldsymbol{v}_{n+1} < \boldsymbol{v}_n \mid n \in \mathbb{N}\}$ has a model with an infinite descending chain.

4. Let $u \in Var$. The following set (with symbols $\boldsymbol{a}$ for the $a \in V$) is consistent:

$$Th(V, \in^V) \cup \{\boldsymbol{a} \in \boldsymbol{b} \mid a, b \in V,\ a \in^V b\} \cup \{\boldsymbol{a} \notin \boldsymbol{b} \mid a, b \in V,\ a \notin^V b\} \cup \{\boldsymbol{a} \in u \mid a \in V\}.$$

Section 3.5

1. $\beta \in T' \Leftrightarrow \alpha \to \beta \in T$ (deduction theorem).

2. $T \subseteq \bigcap\{T' \supseteq T \mid T' \text{ complete}\}$ follows indirectly: $\alpha \notin T \Rightarrow T + \neg\alpha$ is consistent. Hence there is a completion $T' \supseteq T$ with $\alpha \notin T'$ (it may be that $T' = T$).

3. According to Exercise 2, there is a bijection between the set of consistent extensions of T (including T) and the set of nonempty subsets of $\{T_1, \ldots, T_n\}$ = set of all completions of T. This proves both (ii)⇒(i) and the "Moreover" part.

4. With T also the Lindenbaum completion is effectively enumerable, [TMR, p. 15].

Section 3.6

1. $x = y \not\vdash \forall x\, x = y$. Hence the same holds for $\vdash$ in view of $\vdash \subseteq \models$.

2. (a): Let $(\varphi_n)_{n\in\mathbb{N}}$ and $(\mathcal{A}_n)_{n\in\mathbb{N}}$ be effective enumerations of all sentences and of all finite T-models (up to isomorphy). In step n write down all φ_i for $i \leqslant n$ with $\mathcal{A}_n \not\models \varphi_i$. (b): Let $(\alpha_n)_{n\in\mathbb{N}}$ and $(\beta_n)_{n\in\mathbb{N}}$ be effective enumerations of sentences provable or refutable in T, respectively. Each $\alpha \in \mathcal{L}^0$ occurs in one of these sequences. Exactly in the first case α belongs to T.

3. Condition (ii) from Exercise 2 is then granted because the validity of only finitely many axioms is tested in a finite structure.

Section 3.7

1. For **H**: Let h be a homomorphism. Put $x^{hw} := hx^w$. Then $ht^{\mathcal{A},w} = t^{\mathcal{B},hw}$. For **S**: (3) in **2.3**. For **P**: Let $\mathcal{B} = \prod_{i\in I} \mathcal{A}_i$. Then $t^{\mathcal{B},w} = (t^{\mathcal{A}_i,w_i})_{i\in I}$ with $x^w = (x^{w_i})_{i\in I}$.

2. $\alpha_{\text{unc}} := \mathfrak{O}x\, x = x$ is a sentence in $\mathcal{L}^1_Q$ such that $\mathcal{A} \models \alpha_{\text{unc}} \Leftrightarrow A$ is uncountable. Formalize in $\mathcal{L}_{II}$ 'there is a continuous order without a greatest element'.

3. Informally: $\mathbb{R}$ is a continuously ordered set that has a countable dense subset.

4. Let x be a variable not in $\mathcal{P}, \mathcal{Q}$. A possible definition is provided by the program $x := 0$; `WHILE` $\alpha \wedge x = 0$ `DO` $\mathcal{P}$; $x :=$ `S0 OD`; `WHILE` $x = 0$ `DO` $\mathcal{Q}$; $x :=$ `S0 OD`.

Section 4.1

1. Prove first (a) $(\forall i{\in}I)\mathcal{A}_i \models \pi\,[w_i] \Leftrightarrow \mathcal{B} \models \pi\,[w]$ $(x^w = (x^{w_i})_{i\in I})$ for prime formulas π. Then prove (b) $(\forall i{\in}I)\mathcal{A}_i \models \alpha\,[w_i] \Rightarrow \mathcal{B} \models \alpha\,[w]$ by induction over basic Horn formulas α as in Theorem 1.3. (b) yields the induction steps over $\wedge, \forall, \exists$. Observe that $t^{\mathcal{B},w} = (t^{\mathcal{A}_i,w_i})_{i\in I}$. For the universal case apply Theorem 2.3.2.

2. A set of positive Horn formulas has the trivial, one-element model.

Section 4.2

1. With $w_1 \vDash p_1, p_3, \neg p_2$ and $w_2 \vDash p_2, p_3, \neg p_1$ holds $w_1, w_2 \vDash \mathcal{P}$. Since $w \vDash \mathcal{P}$ implies $w \vDash p_3$ and $w \vDash p_1$ or $w \vDash p_2$, there is no valuation $w \leqslant w_1, w_2$ with $w \vDash \mathcal{P}$.

2. For arbitrary $w \vDash \mathcal{P}$, $w \vDash p_{m,n,m+n}$ follows inductively on n. Hence $w_{st} \leqslant w_{\mathcal{P}}$.

3. (a): resolution theorem. (b): $w_{\mathcal{P}} \nvDash p_{n,m,k}$ if $k \neq n+m$; hence $\mathcal{P}, \neg p_{n,m,k} \nvdash^{HR} \square$.

Section 4.3

2. $\Rightarrow$: $x_i \in var\, t_j \Rightarrow x_i^\sigma = t_j \neq t_j^\sigma = x_i^{\sigma^2}$. $\Leftarrow$: $t_i^\sigma = t_i$ since $x^\sigma = x$ for all $x \in var\, t_i$.

3. Let ω be an unifier of $K_0 \cup K_1$. Then $K_0^\omega = K_1^\omega$. Put $x^{\omega'} = x^{\rho\omega}$ for $x \in var\, K_0^\rho$ and $x^{\omega'} = x^\omega$ else. Then $K_0^{\rho\omega'} = K_0^{\rho^2\omega} = K_0^\omega$ $(\rho^2 = \iota)$, and $K_1^{\omega'} = K_1^\omega$.

Section 4.4

1. Let K_0, K_1 be decomposed as in the definition of UR and let ρ be a separator of K_0, K_1, and ω' defined as in the hint to Exercise 3 in **4.3**.

2. Join $\mathcal{P}_g$ and $\mathcal{P}_h$ and add to the resulting program the rules $r_f(\vec{x}, 0, u) :\!- r_g(\vec{x}, u)$ and $r_f(\vec{x}, Sy, u) :\!- r_f(\vec{x}, y, v), r_h(\vec{x}, y, v, u)$.

3. Add to the programs the rule $r_f \vec{x} u :\!- r_{g_1} \vec{x} y_1, \ldots, r_{g_m} \vec{x} y_m, r_h \vec{y} u$.

Section 5.1

3. Let $a, b, c \in \mathbb{R}$ with $0 \leqslant a < b, c$. There is a linear function that represents an automorphism of the (closed) interval $[a, b]$ onto the interval $[a, c]$.

4. W.l.o.g. let $A \cap B = \emptyset$. It suffices to show that $D_{el}\mathcal{A} \cup D_{el}\mathcal{B}$ is consistent.

5. (a): $\{t^{\mathcal{A}} \mid t \in \mathcal{T}_G\}$ is closed with respect to all $f^{\mathcal{A}}$ and exhausts the domain $\mathcal{A}$. (b): According to (a), choose for each $a \in A \backslash G$ some $t_a \in \mathcal{T}_G$ with $\vdash_T a = t_a$.

Section 5.2

2. $T_{\text{suc}} \vdash$ IS because $(\mathbb{N}, 0, \mathtt{S}) \vDash$ IS and T_{suc} is complete. To prove the "no circle" scheme from IS apply IS to $\alpha(x_n) = \forall x_0 \cdots x_{n-1}(\bigwedge_{i<n} \mathtt{S}x_i = x_{i+1} \to x_n \neq x_0)$.

3. Let $a \in G \vDash T$ and $\frac{a}{n}$ the element with $n\frac{a}{n} = a$, and $\frac{m}{n} : a \mapsto m\frac{a}{n}$ for $\frac{m}{n} \in \mathbb{Q}$. Then G becomes the vector group of a $\mathbb{Q}$-vector space that is $\aleph_1$-categorical.

4. Each consistent extension T' of T is the intersection of its completions in T.

5. Each $\mathcal{A} \vDash T$ has a countable elementary substructure (Theorem 1.5).

Section 5.3

1. For SO_{00}: In the first round player II may play arbitrarily, then according to the winning strategies for models of SO_{01} or SO_{10} in the decomposed segments.

2. If player I starts with $a \in A$ and to the right and the left of a remain at least 2^{n-1} elements, player II should choose correspondingly. Otherwise he should answer with the elements of the same distance from the left or right edge element.

3. For $\mathsf{FO} \subseteq \mathsf{SO}_{11}$: $\mathsf{SO}_{11} \nvdash \alpha \Rightarrow \mathcal{A} \nvDash \alpha$ for a sufficiently large finite $\mathcal{A} \vDash \mathsf{FO}$.

4. Prove first that $\mathsf{SO}_{11} \cup \{\exists_i \mid i > 0\}$ is complete.

Section 5.4

1. Let $h\colon \mathcal{A} \to \mathcal{B}$ be a homomorphism, $\mathcal{M} = (\mathcal{A}, w)$, $\mathcal{M}' = (\mathcal{B}, w')$ with $x^{w'} = hx^{w}$. Show $\mathcal{M} \vDash \varphi\,[\vec{a}] \Rightarrow \mathcal{M}' \vDash \varphi\,[h\vec{a}]$ by induction on φ.

2. Let A be an ordered set. Replace each $a \in A$ by an exemplar of $(\mathbb{Z}, <)$ or $(\mathbb{Q}, <)$, respectively. That results in a discrete or a dense order $B \supseteq \mathcal{A}$, respectively.

3. Clearly $T := T_0 + T_1$ is inductive since both T_0, T_1 and hence T are $\forall\exists$-theories. Let $\mathcal{A}_0 \vDash T_0$. Choose $\mathcal{A}_1$ with $\mathcal{A}_0 \subseteq \mathcal{A}_1 \vDash T_1$, $\mathcal{A}_2$ with $\mathcal{A}_1 \subseteq \mathcal{A}_2 \vDash T_0$ etc. This results in a chain $\mathcal{A}_0 \subseteq \mathcal{A}_1 \subseteq \mathcal{A}_2 \subseteq \cdots$ with $\mathcal{A}_{2i} \vDash T_0$, $\mathcal{A}_{2i+1} \vDash T_1$. Then $\mathcal{A}^* := \bigcup_{i\in\mathbb{N}} \mathcal{A}_{2i} = \bigcup_{i\in\mathbb{N}} \mathcal{A}_{2i+1} \vDash T_0, T_1$ and so $\mathcal{A}^* \vDash T$. Therefore T_0 and T are model compatible. Consequently also T_1 and T.

4. The union S of a chain of inductive theories model compatible with T has again these properties as is readily checked. By Zorn's lemma there exists a maximal, hence in view of Exercise 3 a largest theory of this kind.

Section 5.5

1. Let $(i, j) \neq (0, 0)$. Then DO_{ij} has models $\mathcal{A} \subseteq \mathcal{B}$ with $\mathcal{A} \not\preccurlyeq \mathcal{B}$.

2. (a) Lindström's criterion. T is $\aleph_1$-categorical because a T-model can be understood as a $\mathbb{Q}$-vector space. (b) Each T_0-model G is embeddable in a T-Modell H. One gains such H by defining a suitable equivalence relation on the set of all pairs $\frac{a}{n}$ with $a \in G$ and $n \in \mathbb{Z}\setminus\{0\}$.

3. Uniqueness follows similarly to uniqueness of the model completion.

4. The algebraic closure $\overline{\mathcal{F}_p}$ of the prime field $\mathcal{F}_p$ is identical to $\bigcup_{n\geqslant 1} \mathcal{F}_{p^n}$, where $\mathcal{F}_{p^n}$ denotes the finite field of p^n elements. Moreover, a sentence true in all a.c. fields with prime characteristic holds already in all a.c. fields.

Section 5.6

1. Let $\mathcal{A}, \mathcal{B} \vDash \mathsf{ZG}$, $\mathcal{A} \subseteq \mathcal{B}$. Then also $\mathcal{A}' \subseteq \mathcal{B}'$ for their ZGE expansions because $m\restriction$ has in ZG both an $\forall$- and an $\exists$-Definition. Thus $\mathcal{A}' \preccurlyeq \mathcal{B}'$ and hence $\mathcal{A} \preccurlyeq \mathcal{B}$.

2. Very similiar to quantifier elimination in ZGE but somewhat more simple.

3. Inductively over quantifier-free $\varphi = \varphi(x)$ follows: for each $\mathcal{A} \vDash \mathsf{RCF}^\circ$ is $\varphi^{\mathcal{A}}$ or $(\neg\varphi)^{\mathcal{A}}$ finite. This is not the case for $\alpha(x)$.

4. CS holds in the real closed field $\mathbb{R}$, hence in each $\mathcal{A} \in \mathsf{RCF}$. The proofs from CS of $(\forall x \geqslant 0)\exists y\, x = y \cdot y$, and that each polynomial of odd degree has a zero must be carried out without a theory of continuous functions, which is very instructive.

Section 5.7

1. If F is trivial then there is some $i_0 \in I$ with $i_0 \in J$ for each $J \in F$ (Exercise 3 in **1.5**). For $a, b \in \prod_{i\in I} \mathcal{A}_i$ then $a \approx_F b \Leftrightarrow i_0 \in I_{a=b} \Leftrightarrow a_{i_0} = b_{i_0}$. This implies $\prod^F_{i\in I} \mathcal{A}_i \simeq \mathcal{A}_{i_0}$ (can be shown directly or with the homomorphism theorem).

2. $x \mapsto x^I/F$ $(x \in A)$ is an embedding and moreover an elementary embedding.

3. Let $X \vDash_{\boldsymbol{K}} \varphi$ and I, J_α and F defined as in the proof of Theorem **7.3** and assume that for each $i \in I$ there is some $\mathcal{A}_i \in \boldsymbol{K}$ and $w_i : PV \to A_i$ such that $w_i\alpha \in D^{\mathcal{A}_i}$ for all $\alpha \in i$ but $w_i\varphi \notin D^{\mathcal{A}_i}$. Put $\mathcal{C} := \prod^F_{i\in I} \mathcal{A}_i$ $(\in \boldsymbol{K})$ and $w = (w_i)_{i\in I}$. Then $wX \subseteq D^{\mathcal{C}}$ and $w\varphi \notin D^{\mathcal{C}}$, hence $X \nvDash_{\mathcal{C}} \varphi$, a contradiction to $X \vDash_{\boldsymbol{K}} \varphi$.

4. W.l.o.g. $\mathcal{A} = \mathcal{2}$ and $\mathcal{2} \subseteq \mathcal{B} \subseteq \mathcal{2}^I$ for some I according to Stone's representation theorem quoted in **2.1**. $\mathcal{2} \vDash \alpha \Rightarrow \mathcal{2}^I \vDash \alpha \Rightarrow \mathcal{B} \vDash \alpha$ according to Theorem 7.5.

Section 6.1

1. $b \in ran\, f \;\Leftrightarrow\; (\exists a \leqslant b) f a = b$ (this predicate is p.r. iff f is p.r.).

2. Put $S_m := \sum_{i\leqslant m} i$. *Injectivity*: Let $\wp(a,b) = \wp(a',b')$. Were $a + b < a' + b'$ then $\wp(a,b) < \wp(a,b) + a + 1 = S_{a+b} + a + b + 1 = S_{a+b+1} \leqslant S_{a'+b'} \leqslant \wp(a',b')$. Thus $a+b = a'+b'$. But then $a = \wp(a,b) - S_{a+b} = \wp(a',b') - S_{a'+b'} = a'$, hence also $b = b'$. *Surjectivity*: Since $\wp(0,0) = 0 \in ran\, \wp$ it suffices to prove $\wp(a,b) + 1 \in ran\, \wp$, for all a, b. Clear for $b = 0$ because $\wp(a,0) + 1 = S_a + a + 1 = S_{a+1} = \wp(0, a+1)$. In case $b \neq 0$ is $\wp(a,b) + 1 = S_{a+1+b-1} + a + 1 = \wp(a+1, b-1)$. This proof also confirms that the figure for $\wp$ has been drawn correctly.

3. $\Rightarrow$: Let $M = \{a \in \mathbb{N} \mid \exists b Rab\}$, R recursive and $c \in M$ fixed. Put $fn = k$ in case $(\exists m \leqslant n)\, n = \wp(m,k)$ & Rmk, and $fn = c$ otherwise.

4. $\varkappa_1 n = (\mu k \leqslant n)[(\exists m \leqslant n)\wp(k,m) = n]$.

Section 6.2

1. Let $\alpha_0, \alpha_1, \ldots$ be a recursive enumeration of X and let $\beta_n = \underbrace{\alpha_n \wedge \ldots \wedge \alpha_n}_{n}$. By Exercise 1 in **6.1**, $\{\beta_n \mid n \in \mathbb{N}\}$ is recursive and axiomatizes T as well.

3. (a): Let $\Phi_n = (\varphi_0, \ldots, \varphi_n)$ be a proof of $\varphi = \varphi_n$ in $T + \alpha$. Suppose that proofs Φ'_k for $\alpha \to \varphi_k$ from $\Phi_i = (\varphi_0, \ldots, \varphi_i)$ for all $i < n$ have already been constructed. Define a proof Φ'_n for $\alpha \to \varphi$ by p.r. case distinction according to the cases $\varphi = \alpha$, $\varphi \in X \cup \Lambda$ (X is an axiom system for T) and φ_i results from φ_k and φ_m for some $k, m < i$ by applying MP. In other words, follow the proof of Lemma 1.6.3.

4. Prove this first for equations. Construct in a p.r. way for each t a normal form $\mathrm{Nf}(t) = \underline{a_0} + \sum_{1 \leqslant \nu \leqslant n} \underline{a_\nu} \cdot \boldsymbol{v}_0^{k_0^\nu} \cdot \ldots \cdot \boldsymbol{v}_n^{k_n^\nu}$ such that $\mathcal{N} \vDash t_1 = t_2$ iff $\mathrm{Nf}(t_1) = \mathrm{Nf}(t_2)$.

Section 6.3

1. There are several proof methods. A natural way is to proceed stepwise over the length $n > 1$ of $\vec{x}$, using the function $\wp$ which is Δ_0 (Remark 2). It suffices to notice that $\exists x \exists y \alpha \equiv_{\mathcal{N}} \exists z(z = \wp(x, y) \wedge \alpha)$ and $\forall x \forall y \alpha \equiv_{\mathcal{N}} \forall z(z = \wp(x, y) \to \alpha)$, where $z \notin \mathit{var}\,\alpha$. Note that for Σ_1 also works $\exists \vec{x} \alpha \equiv_{\mathcal{N}} \exists x(\exists x_1 \leqslant x) \ldots (\exists x_n \leqslant x)\alpha$ by Exercise 2. In all these equivalences $\equiv_{\mathcal{N}}$ could be replaced by $\equiv_{\mathsf{PA}}$.

2. $(\forall z{<}y)\exists x \alpha \equiv_{\mathsf{PA}} \exists u(\forall z{<}y)(\exists x{<}u)\alpha$ ($u \notin \mathit{var}\,\alpha$, schema of bounds, see Exercise 3 in **3.3**). From this it readily follows that $(\exists z{<}y)\forall x \alpha \equiv_{\mathsf{PA}} \forall u(\exists z{<}y)(\forall x{<}u)\alpha$.

Section 6.4

1. (a): $p \nmid a \Rightarrow a \perp p \Rightarrow \exists xy\, xa{+}1 = yp$ (Euclid's lemma) $\Rightarrow \exists xy\, b = ypb - xab \Rightarrow p \mid b$. (b): Let $m := \mathrm{lcm}\{a_\nu \mid \nu \leqslant n\} = a_\nu c_\nu$ for suitable c_ν. Assume $(\forall \nu \leqslant n) p \nmid a_\nu$. Then $(\forall \nu \leqslant n) p \mid c_\nu$ by (a). Thus $m = pm'$ and $c_\nu = pc'_\nu$ for suitable m', c'_ν. This leads to contradition to the definition of m. (c) easily follows from (b).

2. $\exists u[\mathtt{beta}\, u0 = \underline{2} \wedge (\forall v{<}x)(\exists w, w' \leqslant y)(\mathtt{beta}\, uvw \wedge \mathtt{beta}\, u\mathbf{S}vw' \wedge w < w'$
$\wedge\ \mathsf{prim}\, w \wedge \mathsf{prim}\, w' \wedge (\forall z{<}w')(\mathsf{prim}\, z \to z \leqslant w) \wedge \mathtt{beta}\, uxy)]$.

3. (a): Prove this first for x instead of $\vec{x}$. (b): It suffices to show that $\mathrm{sb}_x(\dot{\varphi}) = \dot{\varphi}$ for $x \notin \mathit{free}\,\varphi$. Observe that $\mathrm{sb}_x((\forall x \alpha)^\cdot, x) = (\forall x \alpha)^\cdot$ for closed α.

Section 6.5

2. (ii)⇒(i): If T is complete and $T' + T$ is consistent then $T' \subseteq T$.

3. Trivial if $T + \Delta$ is inconsistent. Otherwise let $\varkappa$ be the conjunction of all sentences $\forall \vec{x}\, \exists! y \alpha(\vec{x}, y)$, α running through all defining formulas for operations from Δ. If T is decidable than so is $T + \varkappa$. Moreover $\vdash_{T+\Delta} \alpha \Leftrightarrow \vdash_{T+\varkappa} \alpha^{rd}$.

4. $fa = b$ iff Φ is a proof in Q and $\dot{\Phi} = a$ and $(\dot{\Phi})_{last} = b$, or else $b = 0$.

Section 6.7

2. Δ_0 is r.e. but not Δ_1 (Remark 2 in **6.4**). $\dot{\mathsf{Q}}$ is Σ_1 but not Δ_1.

3. The functions $\tilde{\wedge}$, $\tilde{\neg}$, $\tilde{\forall}$, sb_x as well as e.g. $\mathcal{L}_{ar}$ are p.r. and hence Δ_1. The same holds for Tr_0 by Exercise 4 in **6.2**. Clearly

$$\varphi \in Tr_{n+1} \Leftrightarrow \varphi \in Tr_n \vee (\exists\, \alpha, \beta, x \leqslant \varphi)\forall n[\varphi = \forall x\alpha \;\&\; \alpha_x(\underline{n}) \in Tr_n \vee \varphi = \alpha \wedge \beta \;\&\; \alpha, \beta \in Tr_n \vee \varphi = \neg\alpha \;\&\; \alpha \notin Tr_n].$$

Section 7.1

1. For $\wp$: Prove $\vdash_{\mathsf{PA}} \exists z(\underline{2}z = (x+y)^2 + \underline{3}x + y)$ by induction on y.

2. (a): Follow the proof of the lemma in **6.4**. (b): $<$-induction. (c): Use (a).

4. (a): For $\Box_{T+\alpha}\varphi \vdash_T \Box_T(\alpha \to \varphi)$ use Exercise 3b in **6.2**.

Section 7.2

1. $\vdash_T \Box\alpha \to \alpha \Rightarrow \vdash_{T'} \neg\Box\alpha \Rightarrow \vdash_{T'} \mathtt{Con}_{T'}$, since by (5) $\mathtt{Con}_{T'} \equiv_T \neg\Box\neg\neg\alpha \equiv_T \neg\Box\alpha$. Thus, T' is inconsistent by (1), hence $\vdash_T \alpha$.

3. Clear if $n = 0$. Let $T^n = T + \neg\Box^n\bot$ and $\mathtt{Con}_{T^n} \equiv_T \neg\Box^{n+1}\bot$ (induction hypothesis). Since $\Box^n\bot \vdash_T \Box^{n+1}\bot$ by $D2$, $T^{n+1} = (T + \neg\Box^n\bot) + \neg\Box^{n+1}\bot = T + \neg\Box^{n+1}\bot$. Further, by (5), $\mathtt{Con}_{T^{n+1}} \equiv_T \neg\Box\neg(\neg\Box^{n+1}\bot) \equiv_T \neg\Box^{n+2}\bot$.

4. For any (not formalized) arithmetical sentence A the statement 'If A is provable in PA then A is true in $\mathcal{N}$' is provable in ZFC. Formalized: $\vdash_{\mathsf{ZFC}} \Box_{\mathsf{PA}}\alpha \to \alpha$, where α formalizes A.

Section 7.4

1. Prove first $\mathsf{G}_n = \{H \in \mathfrak{F}_\Box \mid \vdash_{\mathsf{G}} \Box^n\bot \to H\}$. We then obtain

$$\begin{aligned} \vdash_{\mathsf{G}_n} H &\Leftrightarrow \vdash_{\mathsf{G}} \Box^n\bot \to H \Leftrightarrow \vdash_{\mathsf{PA}} (\Box^n\bot \to H)^\imath \text{ for all } \imath && \text{(Theorem 4.2)} \\ &\Leftrightarrow \vdash_{\mathsf{PA}} \Box^n\bot \to H^\imath \text{ for all } \imath && \text{(property of } \imath) \\ &\Leftrightarrow \vdash_{\mathsf{PA}_n} H^\imath \text{ for all } \imath && (\mathsf{PA}_n = \mathsf{PA} + \Box^n\bot). \end{aligned}$$

2. Put $\mathsf{PA}^n_\bot := \mathsf{PA}^n + \neg\,\mathtt{Con}_{\mathsf{PA}^n}$. By (6) in **7.2**, $\mathtt{Con}_{\mathsf{PA}^n_\bot} \equiv_{\mathsf{PA}} \mathtt{Con}_{\mathsf{PA}^n} \equiv \neg\Box\Box^n\bot$. Thus, $\mathsf{PA}^n_\bot = (\mathsf{PA} + \neg\Box^n\bot) + \Box^{n+1}\bot$. This theory is consistent ($\nvdash_{\mathsf{PA}} \Box^{n+1}\bot \to \Box^n\bot$). Therefore $\mathsf{PA}^n_\bot$ has the provability logic G_1 (Theorem 4.3). As regards T note that $\Box\,\mathtt{Con}_{\mathsf{PA}} \vee \Box\neg\,\mathtt{Con}_{\mathsf{PA}} \equiv_{\mathsf{PA}} \Box\bot \vee \Box^2\bot \equiv_{\mathsf{PA}} \Box^2\bot$, hence $T = \mathsf{PA} + \Box^3\bot \wedge \neg\Box^2\bot = \mathsf{PA}^2_\bot$.

4. $\nvdash_{\mathsf{G}^*} \neg[\neg\Box(p \to q) \wedge \neg\Box(\neg p \to q) \wedge \neg\Box(p \to \neg q) \wedge \neg\Box(\neg p \to \neg q)]$ and Theorem 4.4.

Literature

[Ba] J. Barwise (editor), *Handbook of Mathematical Logic*, North-Holland 1977.

[BD] A. Berarducci, P. D'Aquino, Δ_0-*complexity of* $y = \prod_{i \leqslant n} F(i)$, Ann. Pure Appl. Logic 75 (1995), 49–56.

[Be1] L. Beklemishev, *On the classification of propositional provability logics*, Izvestiya 35 (1990), 247–275.

[Be2] ———— *Iterated local reflection versus iterated consistency*, Ann. Pure Appl. Logic 75 (1995), 25–48.

[Be3] ———— *Bimodal logics for extensions of arithmetical theories*, Journ. Symb. Logic 61 (1996), 91–124.

[Be4] ———— *Parameter free induction and reflection*, in *Computational Logic and Proof Theory*, LN Comp. Science 1289, Springer 1997.

[Ben] M. Ben-Ari, *Mathematical Logic for Computer Science*, 2nd edition, Springer 2001.

[Bi] G. Birkhoff, *On the structure of abstract algebras*, Proc. Cambridge Phil. Soc. 50 (1935), 433–455.

[BJ] G. Boolos, R. Jeffrey, *Computability and Logic*, 3rd ed. Cambridge Univ. Press 1992.

[BGG] E. Börger, E. Grädel, J. Gurevich, *The Classical Decision Problem*, Springer 1997.

[BM] J. Bell, M. Machover, *A Course in Mathematical Logic*, North-Holland 1977.

[Boo] G. Boolos, *The Logic of Provability*, Cambridge Univ. Press 1993.

[BP] P. Benacerraf, H. Putnam (editor), *Philosophy of Mathematics, Selected Readings*, 2nd ed. Cambridge Univ. Press 1993.

[Bu] S.R. Buss (editor), *Handbook of Proof Theory*, Elsevier 1998.

[Bue] S. Buechler, *Essential Stability Theory*, Springer 1996.

[Ca] G. Cantor, *Gesammelte Abhandlungen*, Springer 1980.

[Ch] A. Church, *A note on the Entscheidungsproblem*, Jour. Symb. Logic 1 (1936), 40–41.

[CK] C.C. Chang, H.J. Keisler, *Model Theory*, 3rd ed. North-Holland 1990.

[CM] W. Clocksin, C. Melnik, *Programming in PROLOG*, 3rd edition, Springer 1987.

[CZ] A Chagrov, M. Zakharyashev, *Modal Logic*, Claredon Press 1997.

[Da] D. van Dalen, *Logic and Structure*, 2nd ed. Springer 1983.

[Dav] M. Davis (editor), *The Undecidable*, Raven New York 1965.

[De] R. Dedekind *Was sind und was sollen die Zahlen?*, (Braunschweig 1888), Vieweg 1969.

[Do] K. Doets *From Logic to Logical Programming*, MIT-Press 1994.

[EFT] H. Ebbinghaus, J. Flum, W. Thomas, *Mathematical Logic*, Springer 1996.

[En] H. Enderton, *A Mathematical Introduction to Logic*, Acad. Press 1972. 2nd edition 2001.

[Fe] W. Felscher, *Lectures on Mathematical Logic*, Gordon and Breach 2000.

[Fr] G. Frege, *Begriffsschrift*, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens (Halle 1879), Olms 1971.

[FS] H. Friedman, M. Sheard, *Elementary decent recursion and proof theory*, Arch. Pure & Appl. Logic 71 (1995), 1–47.

[Ga] D. Gabbay, *Decidability results in non-classical logic III*, Israel Jour. Math 10 (1971), 135–146.

[Ge] G. Gentzen, *Collected Works of Gerhard Gentzen*, North Holland 1969.

[GJ] M. Garey, D. Johnson, *Computers and Intractability*, A Guide to the Theory of NP-Completeness, Freeman 1979.

[Go1] K. Gödel, *Die Vollständigkeit der Axiome des logischen Funktionenkalküls*, Monatshefte Math. u. Physik 37 (1930), 349–360, or *Collected Works I.*

[Go2] ———, *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I*, Monatshefte für Mathematik und Physik 38 (1931), 173–198.

[Go3] ———, *Collected Works*, Oxford Univ. Press, Vol. I 1986, Vol. II 1990, Vol. III 1995.

[Gor] S.N. Goryachev, *On the interpretability of some extensions of arithmetic*, Math. Notes 40 (1986), 561–572.

[Gr] G. Grätzer, *Universal Algebra*, 2nd ed. Springer 1979.

[HA] D. Hilbert, W. Ackermann, *Grundzüge der theoretischen Logik*, Springer 1928.

[HB] D. Hilbert, P. Bernays, *Grundlagen der Mathematik*, Band I Springer 1934, Band II Springer 1939.

[He] L. Henkin, *The completeness of the first-order functional calculus*, Journ. Symb. Logic 14 (1949), 159–166.

[Hej] J. Heijenoort, *From Frege to Gödel*, Harvard Univ. Press 1967.

[HeR] B. Herrmann, W. Rautenberg, *Finite replacement and Hilbert-style axiomatizability,* Zeitsch. Math. Log. Grundl. Math. 38 (1982), 327–344.

[Her] J. Herbrand, *Recherches sur la théorie de la démonstration*, in [Hej].

[Ho] W. Hodges, *Model Theory*, Cambridge Univ. Press 1993.

[HP] P. Hajek, P. Pudlak, *Metamathematics of First-order Arithmetic*, Springer 1993.

[HR] H. Herre, W. Rautenberg, *Das Basistheorem und einige Anwendungen in der Modelltheorie,* Wiss. Zeitsch. Humboldt-Univ. 19 (1970), 575–577.

[Id] P. Idziak, *A characterization of finitely decidable congruence modular varieties*, Trans. Am. Math. Soc. 349 (1997), 903–934.

[Ig] K. Ignatiev, *On strong provability predicates*, Journ. Symb. Logic 58 (1993), 249–290.

[JK] R. Jensen, C. Karp, *Primitive recursive set theory*, in *Axiomatic Set Theory*, AMS 1971, 143–167.

[Ka] R. Kaye, *Models of Peano Arithmetic*, Clarendon Press 1991.

[Ke] H.J. Keisler, *Logic with the quantifier "there exist uncountably many"*, Ann. Pure Appl. Logic 1 (1970), 1–93.

[Kl1] S. Kleene, *Introduction to Metamathematics*, North-Holland 1952, 2nd edition, Wolters-Noordhoff 1988.

[Kl2] ———, *Mathematical Logic*, Wiley & Sons 1968.

[KR] I. Korec, W. Rautenberg, *Model interpretability into trees and applications*, Arch. Math. Logik 17 (1976), 97–104.

[KK] G. Kreisel, J. Krivine, *Elements of Mathematical Logic*, North-Holland 1971.

[Kr] J. Krajíček, *Bounded Arithmetic, Propositional Logic, and Complexity Theory*, Cambridge Univ. Press 1995.

[Ku] K. Kunen, *Set Theory. An Introduction to Independence Proofs*, North-Holland 1980.

[Li] P. Lindström, *On extensions of elementary logic*, Theoria 35 (1969), 1–11.

[Ll] J.W. Lloyd, *Foundations of Logic Programming*, Springer 1987.

[Lo] M. Löb, *Solution of a problem of Leon Henkin*, Journ. Symb. Logic 20 (1955), 115–118.

[Ma] A.I. Malcev, *The Metamathematics of Algebraic Systems*, North-Holland 1971.

[Mal] J. Malitz, *Introduction to Mathematical Logic*, Springer 1979.

[Mar] D. Marker, *Model Theory, An Introduction*, Springer 2002.

[Mat] Y. Matiyasevich, *Hilbert's Tenth Problem*, MIT Press 1993.

[Me] E. Mendelson, *Introduction to Mathematical Logic*, Van Nostrand 1979.

[Mo] D. Monk, *Mathematical Logic*, Springer 1976.

[ML] G. Müller, W. Lenski (editors) *The Ω-Bibliography of Mathematical Logic*, Springer 1987.

[MS] A. Macintyre, H. Simmons, *Gödel's diagonalization technique and related properties of theories*, Colloq. Math. 28, 1973.

[MW] A. Macintyre, A. Wilkie, *On the Decidability of the Real Exponential Field*, Preprint 1995.

[MV] R. McKenzie, M. Valeriote, *The Structure of Decidable Locally Finite Varieties*, Progress in Math. 79, Birkhäuser 1989.

[Po] W. Pohlers, *Proof Theory – An Introduction*, Springer Lecture Notes 1407 (1989)

[Pz] B. Poizat, *A Course in Model Theory. An Introduction to Contemporary Mathematical Logic*, Springer 2000.

[Pr] M. Presburger, *Über die Vollständigkeit eines gewisses Systems der Arithmetik ganzer Zahlen, in welchem die Addition als einzige Operation hervortritt*, Congr. Math. Pays Slaves (1) (1929), 92–101.

[RS] H. Rasiowa, R. Sikorski, *The Mathematics of Metamathematics*, Polish Scientific Publ. 1968.

[Ra1] W. Rautenberg, *Klassische und Nichtklassische Aussagenlogik*, Vieweg 1979.

[Ra2] ———, *Elementare Grundlagen der Analysis,* BI Verlag 1993.

[Ro1] A. Robinson, *Introduction to Model Theory and to the Metamathematics of Algebra*, North-Holland 1974.

[Ro2] A. Robinson, *Non-Standard Analysis*, North-Holland 1970.

[Rob] J. Robinson, *A Machine-oriented logic based on the resolution principle*, Journ. Ass. Comp. Machinery 12 (1965), 23–41.

[Rog] H. Rogers, *Theory of Recursive Functions and Effective Computability*, 2nd ed. MIT Press 1988.

[Ros] J.B. Rosser, *Extensions of some Theorems of Gödel and Church*, Journ. Symb. Logic 1 (1936), 87–91.

[Rot] P. Rothmaler, *Introduction to model theory*, Gordon & Breach, 2000.

[Ry] C. Ryll-Nardzewki, *The role of induction in elemenary arithmetic*, Fund. Math. 39 (1952), 87–91.

[RZ] W. Rautenberg, M. Ziegler *Recursive inseparability in graph theory,* Notices Am. Math. Soc. 22 (1975).

[Sa] G. Sacks, *Saturated Model Theory*, Benjamin Reading 1972.

[Sam] G. Sambin, *An effective fixed point theorem in intuitionistic diagonalizable algebras*, Studia Logica 35 (1976), 345–361.

[Se] A. Selman, *Completeness of calculi for axiomatically defined classes*, Algebra Universalis 2 (1972), 20–32.

[Sh] S. Shelah, *Classification Theory and the Number of Nonisomorphic Models*, North-Holland 1978.

[Shoe] J. Shoenfield, *Mathematical Logic*, Addison-Wesley 1967.

[Si] W. Sieg, *Herbrand Analyses*, Arch. Math. Logic 30 (1991), 409–441.

[Sm] C. Smoryński, *Self-reference and Modal Logic*, Springer 1984.

[Sm1] R. Smullyan, *Theory of Formal Systems*, Princeton Univ. Press 1961.

[So] R. Solovay, *Provability interpretation of modal logic*, Israel Jour. Math. 25 (1976), 287–304.

[Sz] W. Szmielew, *Elementary properties of abelian groups*, Fund. Math. **41** (1955), 287–304.

[Ta1] A. Tarski, *Der Wahrheitsbegriff in den formalisierten Sprachen*, Studia Philosophica 1 (1936), in [Ta3].

[Ta2] ———, *A Decision Method for Elementary Algebra and Geometry*, Berkeley 1948, 1951, Paris 1967.

[Ta3] ———, *Logic, Semantics and Metamathematics*, Clarendon Press 1956.

[Ta4] ———, *Introduction to Logic and and to the Methodology of the Deductive Sciences*, Oxford Univerity Press 1994 (first edition in Polish, 1936).

[Tak] G. Takeuti, *Proof Theory*, North Holland, 1975.

[TMR] A. Tarski, A. Mostowski, R.M. Robinson, *Undecidable Theories*, 2nd ed. North-Holland 1971.

[TV] A. Tarski, R. Vaught, *Arithmetical extensions and relational systems*, Compositio Math. 13 (1957), 81–102.

[Tu] A. Turing, *On computable numbers, with an application to the Entscheidungsproblem*, Proceedings of the London Mathematical Society 43 (1937), reprint in [Dav].

[Vi] A. Visser, *An Overview of Interpretability Logic*, *Advances in Modal Logic 1996*, Lecture Notes CSLI (ed. M. Kracht et al.), Stanford 1998.

[Wae] B. van der Waerden, *Algebra I*, Springer 1964.

[WR] A. Whitehead, B. Russell, *Principia Mathematica,* Cambridge Univ. Press 1910.

[Wi] A. Wilkie, *Model completeness results for expansions of the real field II: the exponential function*, Journ. Am. Math. Soc. 9 (1996), 1051–1094.

[WP] A. Wilkie, J. Paris, *On the scheme of induction for bounded arithmetic Formulas,* Ann. of Pure & Appl. Logic 35 (1987), 261–302.

[Zi] M. Ziegler, *Model theory of moduls*, Ann. Pure Appl. Logic 26 (1984), 149–213.

Index of Terms and Names

Index of Symbols