Pavol Zlatoš

Lectures on Mathematical Logic

2025

Comenius University Bratislava

Publication supported by grant no. K-23-002-00 of the Slovak grant agency KEGA.

Comenius University Bratislava
Faculty of Mathematics, Physics and Informatics
Department of Algebra and Geometry

Reviewers
Prof. RNDr. Miroslav Haviar, PhD.
Prof. RNDr. Ladislav Kvasz, DSc.

https://stella.uniba.sk/texty/FMFI_PZ_mathematical_logic_lectures.pdf

Publisher
Comenius University Bratislava

ISBN 978-80-223-6123-1 (print)
ISBN 978-80-223-6124-8 (online)

Contents

Preface

This book would probably never have seen the light of day if not for the COVID-19 pandemic of 2020 – 2021. I had been lecturing on mathematical logic long before that and had written a series of lecture notes, initially distributed as hand-outs to students and later made available online. I continuously had been editing, supplementing and expanding the notes, but I had no intention of processing them into a book form. After all, there are so many excellent textbooks on logic in general and mathematical logic in particular... Nevertheless, I never used any single one of them as my primary source, even though I drew inspiration from many. As a rule they were, to my taste, too extensive or overloaded with technical details unnecessary for students who do not intend to specialize in logic. This gave me the freedom to select topics based on students' interests and to modify my approach accordingly. Furthermore, I typically taught the first-semester material in Slovak in the fall for students of the major *Mathematics*, and in English in the spring, in a slightly simplified form with fewer proofs, for students of *Cognitive Science*. The Slovak course continued in the spring semester, whereas the English course spanned for just one semester.

However, COVID changed everything. I was taken aback by the (inevitably negative) impact of replacing in-person lectures with online teaching. To mitigate these effects at least partially, I started preparing a structured text, divided into sections corresponding as closely as possible to individual lectures, and kept up with them on a weekly basis. I made the text available to students in electronic form as soon as it was completed. Thus, during the spring of 2020, the first draft of this book came to life. At the same time, I was writing an English version of the first part of the course (dedicated to propositional logic, first-order logic, and Gödel's incompleteness theorems) and a Slovak version of the second part (covering selected topics in model theory). During the fall of 2020 and spring of 2021, I mostly just revised and expanded both parts. The decision to elaborate this series of materials into a full-fledged book in two language mutations matured in me only in the fall of 2021. At the same time, I conceived a conceited intention to distinguish my text from most mathematically oriented logic textbooks by including two introductory chapters: one on the subject and significance of logic and its position within the system of sciences, and the other offering a brief historical overview of logic, presenting it as an integral and indispensable part of human culture.

The content of this book bears the imprint of the circumstances of its creation. Whether one regards this as an advantage or virtue out of necessity, the book's scope is deliberately limited, making it feasible to cover in a two-semester course. The selection of topics, beyond the obligatory core material (propositional and predicate

logic), reflects the interests and preferences of the author. Thus it should come as no surprise that several important topics are not covered here. Just to name a few: logical circuits, non-classical, modal, and many-valued logics, computability theory, complexity theory, several areas of model theory, etc. However, as already mentioned, "there are so many excellent textbooks on logic in general and mathematical logic in particular..." Readers wishing to delve deeper into some of the topics omitted here, will find plenty of resources to choose from. Some hints can be found in the bibliography, but they are by no means exhaustive.

Bratislava, September 2025

Pavol Zlatoš

1 The Subject of Logic and its Position within the System of Sciences

The word *logic* can take on different meanings depending on the context. These meanings are related to each other, but they are certainly not identical. In a broad and commonly used sense in everyday speech, *logic* is understood as something like an internal structure, rather sensed than apparent, which, however, gives sense to a given matter (issue), to interpretation of a set of phenomena, a particular field of knowledge, or to an argumentation, or serves as a guarantee for it. Conversely, when this structure is disrupted, this sense is lost, and its absence or denial turns it into nonsense. This way, we often speak of the *logic of things*, the *logic of law*, the *logic of development*, the *logic of events*, the *logic of conflict*, and so on. This meaning is further refined in the context of scientific theories, particularly in their deductive constituents. Here, *logic* merges with a set of methodological principles for building such theories, which includes, among other things, the analysis of their fundamental principles, methods of processing experimental data and facts, the formulation of hypotheses, and methods of their verification.

In a somewhat more specific meaning, *logic* is a set of rules and principles of correct thinking, reasoning, and inference that must be followed in these activities, and whose observance ensures their correctness. Conversely, their violation serves as a warning sign indicating errors in a given inference. However, it is important to note that *correctness* is not the same as *truth*. A logically correct argument necessarily leads to a true conclusion only if it starts from true premises. On the other hand, a true conclusion can sometimes be reached through a logically correct argument based on false premises, as well as through a logically flawed argument, regardless of the truth of the initial premises.

Logic thus becomes a scientific discipline at the intersection of philosophy, the methodology of science, linguistics, and (as it extensively employs mathematical methods and procedures) also mathematics, with important applications in law and, in turn, in mathematics itself.

For the purposes of our course, however, we will understand *logic* in the following even narrower sense:

> *Logic is a normative scientific discipline examining the form, structure, and laws of correct (i.e., logical) thinking and reasoning, as they manifest in language, whether spoken or written, while abstracting from the content of specific thoughts and inferences.*

This "slogan", however, requires further explanation.

From experience, we know that thinking and reasoning can follow the paths of reason as well as the paths of unreason, they can be fully logical, less logical, or entirely illogical. This is evidenced, for example, by commonly used mocking or even politically incorrect expressions such as *female logic*, *military logic*, or *police logic*, and so on. However, if thinking is illogical, so much the worse for thinking. Similarly, a faulty (illogical) inference does not disprove logic; rather, logic disproves such an inference. Thus, the mission of logic is not to empirically describe the various forms of thinking and reasoning but to establish and formulate *norms* that ensure their correctness. We say that logic performs a *normative function* with respect to language and thought.

The reader has surely noticed the presence of the adjective *logical* in our definition of logic as a scientific discipline, which may have given rise to the feeling that we are somewhat caught in a logical circle. Well, we do not intend to hide this. The author openly admits that without a certain knowledge of logic, more precisely, without the ability to think logically, logic cannot be explained or studied. He will therefore assume that his readers or listeners already are capable to think and argue logically, and he does not intend to teach them this but rather to require it of them. On the other hand, the study of logic, which involves becoming aware of and reflecting on the principles of logical thinking and reasoning, undoubtedly develops and cultivates these abilities.

Let us pause for a moment to consider the relationship between logic and language. Although thinking is probably not entirely exhausted by thinking in language — since it also includes various images and nonverbal representations, vague intuitions and insights, as well as moods and emotions — if we wish to share the content or results of our thinking with others, we must find a way to express and communicate them. The most effective communication tool at our disposal is our language, whether in its spoken or written form. However, the communication in language requires us to give linguistic form even to those components of our thinking that originally lacked it. Of course, there is also the possibility of expressing oneself through music or visual arts, through movement and dance, or through gestures and facial expressions, but in these cases (with the exception of sign language), logic is of little help. Logic also falls short when it comes to many forms of literary creation, especially, though not exclusively, to poetry. However, when we engage in reasoning, where we derive conclusions through logical argumentation based on certain accepted premises (again that vicious circle), these processes take place fully, though not exclusively, in language (since in spoken discourse, elements such as intonation, gestures, and facial expressions also play a role) and fall within the domain of logic.

Finally, let us try to clarify what it actually means to *abstract from the content* of particular thoughts and inferences and to focus exclusively on their *form*. Is something like that possible at all? The attempt to see in diverse phenomena their content and form, and to separate one from the other, runs deep in the tradition of European thought. Although to a great extent artificial, this approach has contributed significantly to the successes of Western civilization in the domains of philosophy, science, and ultimately also technology. This distinction namely allows us to conceptually break down a perceived segment of reality into two components, each assigned a different function: *content* and *form*. By suppressing content, we obtain a considerably

simplified, emptied form of a given phenomenon, which can be more easily subjected to intellectual analysis. Moreover, such an abstract form may also appear in a different domain of phenomena, allowing it to be filled with new content. This, in turn, enables the transfer of knowledge gained in one domain to another. This contributes to the universal character of European science and constitutes one of the foundational pillars of *reductionism*, which is among the leading methodological principles of scientific inquiry. It also facilitates the broad application of mathematical methods while simultaneously providing incentives for further developments in mathematics itself.

The philosophical problem of the separability of form and content first emerges in a developed form in ancient Greece in connection with classical geometry. Certain plane or spatial figures capture our interest due to their shape, particularly its simplicity and symmetry. These shapes were given specific names. In the plane, they include the *triangle* (with various qualifiers such as *equilateral*, *isosceles*, *right-angled*), the *rectangle*, and the *square* as its special case, as well as the *circle*, the *regular pentagon*, the *regular hexagon*, and so on. In space, they are primarily the so-called *Platonic solids*, namely the *regular tetrahedron*, the *cube*, the *octahedron*, the *dodecahedron*, and the *icosahedron*, as well as the *sphere*, the *(quadrilateral) pyramid*, etc. Many of these shapes also prove to be useful and desirable in various fields of human activity: in land surveying, in the construction of dwellings, temples, and palaces, in various crafts, and so forth. At the same time, these shapes, and even more so when skillfully combined, appeal to our aesthetic sense, playing an important role in visual arts (painting and sculpture) and architecture.

Furthermore, human-made objects utilizing these shapes tend to be more perfected the more precisely they succeed in achieving or imitating the given form. With the pursuit of perfection there comes the realization that absolute perfection is, at least for human creators, unattainable. Pure and perfect geometric shapes thus become self-standing, independent entities, transforming into an unreachable ideal. Likewise, geometry itself becomes an independent science devoted to the study of these idealized forms. As this process unfolds, the list of ideal geometric shapes expands to include objects even further removed from reality. Chief among these are *points*, conceived as the smallest possible locations in a plane or space, devoid of any spatial extent, and *lines* or *segments*, understood as perfectly straight paths without any thickness.

The interpretation presented above, according to which ideal geometric objects are creations of human intellect and originate from experience and practical human activity, followed in the spirit of Aristotle. However, it is also possible to adopt the standpoint held by Plato, who argued that ideal geometric objects have an independent existence and reside in an ideal geometric world (somewhere on the halfway between the material world and the realm of pure ideas) which precedes the earthly material world in its existence. Ideal geometric objects participate in real objects of corresponding shapes, and the degree of this participation depends on how faithfully a given material object imitates the corresponding ideal form. However, the purity of an ideal geometric object in its earthy realization is always "tainted" by its material content, and only, so to say, shines through or from behind it.

For the purposes of our course, it does not matter which of these two interpretations we prefer. What is important is to recognize that ancient geometry demonstrates the possibility of conceptually separating form from content in a broad class of

phenomena. Furthermore, geometry, even in its original ancient form, and even more so in its later developmental stages, serves as a compelling and undeniable evidence in favor of the fertility and success of such an approach.

As we will have the opportunity to see several times, logic also allows for a similar separation of form from content. Moreover, compared to geometry, this separation is conceptually much simpler and does not pose analogous metaphysical problems. This is made possible primarily through symbolic notation inspired by mathematics, which provides a concrete representation of logical forms. Let us illustrate what we mean by this with a simple example.

One of the well-known syllogisms is often illustrated with the following reasoning:

All men are mortal.
Socrates is a man.
Therefore*, Socrates is mortal.*

The first statement is commonly referred to as the *major premise*, the second as the *minor premise*, and the third as the *conclusion*. Notice that the conclusion follows from the premises with logical necessity. The validity of this reasoning is not based on empirical proof of Socrates' mortality, such as the execution of his death sentence by drinking a cup of hemlock. The same form appears in the following reasoning, as well:

All kings are horses.
Alexander the Great was a king.
Therefore*, Alexander the Great was a horse.*

However absurd, this reasoning is still logically valid, even though its conclusion is obviously false. Another reasoning of a similar type

All kings are horses.
Alexander the Great was a horse.
Therefore*, Alexander the Great was a king.*

is not logically valid, because its conclusion — although true — does not logically follow from the premises. On the other hand, another absurd reasoning

All kings are horses.
Alexander the Great was not a horse.
Therefore*, Alexander the Great was not a king.*

is logically valid, even though its conclusion is false (as is one of its premises).

A similar nature can be observed in the following two logically valid inferences:

All Greeks are men.
All Athenians are Greeks.
Therefore*, all Athenians are men.*

or

All Spartans are brave warriors.
Some Greeks are Spartans.
Therefore*, some Greeks are brave warriors.*

It would not be difficult to provide further examples of logically valid inferences that follow the same patterns as the last two. However, their form only faintly shines from behind of these individual examples. Understanding it is achieved through illustration with a sufficient number of suitable examples. In Aristotle's works, however, for the first time in history, thanks to the use of subject-predicate variables, its explicit form emerges, for example:

All M are P.
All S are M.
Therefore, *all S are P.*

or alternatively,

All M are P.
Some S are M.
Therefore, *some S are P.*

The middle term M appears in both premises: in the first, it serves as the subject, while in the second it functions as the predicate, and it is absent in the conclusion. The term P appears in both the first premise and the conclusion, always playing the role of the predicate; the term S is present in the second premise and the conclusion, in both cases acting as the subject.

Similar forms appear in the logic of the *Stoic school*, where ordinal numerals take on the role of propositional variables:

If the first, then the second.
But the first.
Therefore, *the second.*

or

If the first, then the second.
But not the second.
Therefore, *not the first.*

Using modern symbolic notation, which, however, developed only in the course of the 19$^{\text{th}}$ and 20$^{\text{th}}$ centuries, we can represent the forms of these inferences in a fairly simple and clear way. Moreover, this allows us to distinguish the form of the first two valid inferences about Socrates and Alexander from the form of the inference about Greeks and Athenians. Let $P(x)$, $Q(x)$ and $R(x)$ represent arbitrary properties (predicates) of a variable object x, and a denote any specific object. Then the following inferences are logically valid:

$$(\forall x)(P(x) \Rightarrow Q(x)) \qquad (\forall x)(P(x) \Rightarrow Q(x))$$
$$P(a) \qquad \neg Q(a)$$
$$\text{Therefore, } Q(a) \qquad \text{Therefore, } \neg P(a)$$

$$(\forall x)(P(x) \Rightarrow Q(x)) \qquad (\forall x)(P(x) \Rightarrow Q(x))$$
$$(\forall x)(R(x) \Rightarrow P(x)) \qquad (\exists x)(P(x) \wedge R(x))$$
$$\text{Therefore,} (\forall x)(R(x) \Rightarrow Q(x)) \qquad \text{Therefore,} (\exists x)(Q(x) \wedge R(x))$$

Finally, let us clarify what we understand by the phrase *mathematical logic*. Logic can be mathematical in two basic ways: through the methods it uses, and through its subject matter. Logic can analyze various fragments of natural or artificial languages that allow such an analysis. In this process, mathematical methods can be helpful to varying degrees, such as symbolic representation of individual linguistic objects or their classes, and modeling relationships between them using mathematical tools. These tools usually facilitate processing using computational techniques. In this sense, terms like *formal logic* or *symbolic logic* are often used. An example of this type of logic is *propositional calculus*, with which our study of mathematical logic will begin.

The subject of logic can also be the languages of mathematical theories, the analysis of their deductive structure, and the relationship between the symbolism used within them and the meaning that the corresponding symbols may acquire. It is hardly surprising that the languages of mathematical theories, thanks to their high degree of precision and unambiguity, are particularly suitable for such an analysis. This type of logic is then mathematical in its subject matter. It is also not surprising that this logic heavily uses mathematical methods. Thus, it is mathematical logic both in its methods and its subject matter. We will see that this "doubly mathematical" logic not only provides a unifying perspective on various areas of mathematics but also often enriches them with new methods and interesting mathematical insights. Starting from the chapter on *first-order* logic, we will focus namely on this kind of logic.

The notation and its structure in a symbolic language is scholarly called *syntax*. The assignment of meaning to symbols and their conglomerates is the subject of *semantics*. The investigation of the relationship between syntax and semantics will be both a key theme as well as a unifying element of our study of various branches of mathematical logic in this course.

2 A Glance into History

We would like to warn the reader in advance that our brief overview of the history of logic makes no claim to completeness, and the author is aware that any attempt to do so would be futile. Our goals are much more modest: at least kaleidoscopically present logic in its development as part of the still-living current of European culture and education, originating from times long past and winding through human history to the present day.

2.1 Antiquity

Logic — much like classical geometry — is one of the fruits of the Greek miracle. Though some germs of logic can be found in the thinking of ancient China and India, too, they did not come close to the level to which logic was elevated in ancient Greece. The very word *logic* is derived from the Greek word *logos* ($\lambda o \gamma o \varsigma$), which means both *word* and *speech*, but also *order*, as the opposite of *chaos*, or even the principle that creates the world by breaking it free from chaos. In a similar sense, the word *Word* is used in the opening verse of the Gospel of John, which bears a distinct Hellenistic influence: *In the beginning was the Word, and the Word was with God, and the Word was God.*

Logic, as an intentional part of human knowledge reflecting manifestations of human thinking in language, could not have arisen until thinking and language became logical. In other words, logic had to first be constituted within language as an "*internal structure, rather sensed than apparent, which, however, gives sense to a given matter* [...] *or serves as a guarantee for it* [...]," and only then could it be discovered and gradually begin to take the form of a "*normative scientific discipline examining the form, structure, and laws of correct (i.e., logical) thinking and reasoning, as they manifest in language,* [...] *abstracting from the content of specific thoughts and inferences.*" We can only speculate as to what contributed to the discovery of logic in the first of the mentioned meanings and to the emergence of logic in the second meaning. To what extent was it the political organization of democratic Greek *polis*, the development of trade, legal systems, philosophy, and mathematics, within which logical thinking and the art of argumentation were cultivated, but also the growing respect for these skills and the demand for people proficient in them, capable, for example, of representing parties in lawsuits or teaching others this art. It seems that all of this played a part in the birth of logic. However, it would be naive to expect that we can provide a sufficiently convincing explanation through purely rational reasons.

Those who dealt with teaching rhetoric as the art of speaking and argumentation were called *sophists*, which roughly translates to *teachers of wisdom.* Sophists were often hired as representatives in lawsuits. Their goal was generally to persuade their

opponent of the correctness of their views. This led them to analyze language and uncover its internal structure, which differentiates universally valid arguments from others. Sophists thus discovered and used the *law of contradiction*, the *law of the excluded middle*, and so on. However, the utilitarian nature of their profession caused that their logic fell under the realm of *eristics*, which can be loosely translated as the art of conducting a dispute. More important than searching for truth or the correctness of judgments became refuting or at least questioning the arguments of the opponent, as well as creating seemingly logically correct arguments that lead to desired conclusions and are difficult to refute. Additionally, as a byproduct primarily for the amusement of the audience, a variety of seemingly correct arguments called *sophisms* emerged, leading to absurd conclusions. Some sophisms, however, took on the character of *paradoxes*, hinting at the limits of logic and conceptual understanding of the world in language. Let us look at a few examples.

Sophism *The Horned*, attributed to Eubulides of Miletus (4^{th} century BCE):

What you have not lost, you have.
You have not lost horns.
Therefore, you have horns.

The reasoning itself is logically correct. However, the first premise is obviously false, and therefore, the conclusion is also false.

Sophism ascribed to the brothers Euthydemus and Dionysodorus (5^{th} century BCE):

This dog is yours.
This dog is a father.
Therefore, your father is a dog.

Here, the mistake is less obvious. The words *father* and *dog* are used in two different roles. They refer both to the properties of "being a father" or "being a dog", and to a specific person (your father) or a specific dog (this dog). The trick lies in the fact that these dual roles are deliberately not distinguished. Let us denote $F(x)$ as the property "being a father", $D(x)$ as the property "being a dog", and $Y(x)$ as the property "belonging to you" (i.e., to the recipient of the argument). Let f represent your father and d represent this dog. The argument then takes the form:

$Y(d)$
$F(d)$
Therefore, $D(f)$

This reasoning is not logically valid because its premises (even if true) do not say anything about the object f that appears in the conclusion. The argument cannot be saved even by adding three additional true premises: $F(f)$, $Y(f)$, and $D(d)$, which are omitted in the argument, but psychologically expected. None of the given premises establishes any relationship between the considered predicates $F(x)$, $D(x)$, and $Y(x)$.

The following sophism, attributed to perhaps the most famous of the sophists, Protagoras of Abdera (cca 485 – 410 BCE), however, has no simple solution.

The Paradox of Protagoras' Student. A young man became a student of Protagoras. According to their agreement, he was to pay his teacher only after completing his studies and winning his first lawsuit. The young man completed his studies but did not engage in any legal disputes, and therefore, according to the agreement, he did not pay his master. Eventually, Protagoras lost patience, demanded payment, and threatened the student with a lawsuit. He explained: "The court will either rule that you must pay or it will not. In the first case, you will have to pay based on the court's decision; in the second case, you will have to pay according to our agreement, since you have won your first case." However, the student, well-trained by Protagoras himself, was not intimidated and responded: "I will not pay in either case. If the court orders me to pay, then I have not won my first case, so according to our agreement, I do not owe anything. If the court does not order me to pay, then I do not have to pay based on the court's ruling."

The most famous sophism of all, however, is the *Liar Paradox*, also known as *Epimenides' Paradox*, again attributed to Eubulides. In this paradox, the Cretan philosopher Epimenides makes the following statement:

"*All Cretans are liars.*"

We will examine this paradox in more detail in the chapter dedicated to Gödel's incompleteness theorems.

Another source of logic, initially called *dialectics*, is Greek philosophy. Within it, the view gradually emerges and prevails that the fundamental mission of philosophy is to distinguish appearance from reality and to attain certain and indisputable knowledge about the world. This primarily means understanding the order of the world (*logos*), that is, what is stable and unchanging in it. Individual phenomena and things, on the other hand, are variable, temporary, and unstable. Therefore, one must penetrate through phenomena to their substances, and, uncover, behind the changes, the laws governing them, and then, in turn, explain the changes they undergo. However, the source of such knowledge cannot be the senses, which often deceive us, nor our everyday experience. Although experience may inspire our reasoning, we can reach true knowledge only through thought and reason.

This tendency is already evident in pre-Socratic philosophy, particularly in Parmenides (cca 540 – 470 BCE) and his followers, members of the school he founded in the city of Elea in southern Italy. Parmenides describes the world as a single, unchanging, and eternal entity, which he calls *Being* or the *One*. This Being *is*, because Being cannot not be, and there *is* only One, whereas Non-Being *is not*. With all due respect, to a modern reader, this whole argument appears like an inflated bundle of tautologies, making it difficult to take seriously.

To support his master's teachings, Parmenides' student Zeno (cca 490 – 430 BCE) developed a sophisticated system of arguments, known as *Zeno's paradoxes*, which demonstrate the impossibility of motion. Zeno always assumes that motion occurs and, based on this assumption (along with additional premises, which we will point out later), arrives at a contradiction in the form of an absurd consequence, from which he concludes that motion is impossible. We should recognize that this conclusion contradicts common experience, so its persuasiveness relies solely on the irrefutability of

the logical arguments used. This is likely the oldest recorded example of argumentation of this kind. Moreover, by insistently pointing out the contradictions in the conceptual grasp of motion, Zeno's paradoxes present a permanent challenge to the sciences born in the spirit of antiquity — mathematics, physics, and philosophy in particular — which has not lost its relevance over time. Ultimately, let the readers judge for themselves.

Dichotomy. There is no motion because whatever moves must first reach the halfway point before reaching the end. But before reaching halfway, it must first reach the halfway point of that halfway, and so on. Thus, motion can never even begin.

Achilles and the Tortoise. Achilles, the fastest of men, can never catch up with a tortoise, the slowest of creatures, if the tortoise starts ahead of him. The pursuer must first reach the point where the fugitive started, then the point where the fugitive was when the pursuer reached the starting point, and so on. Thus, the slow tortoise will always remain a certain distance ahead of the swift-footed Achilles.

The Flying Arrow. Suppose that at every moment, everything is at its own place and is either at rest or in motion. If the flying arrow is at a specific place at every moment, then at each individual moment, it is motionless and therefore cannot move. If it were moving at any given moment, it would not be at that place at that moment, meaning it would not be anywhere at all.

The Stadium. Imagine two rows, each consisting of the same number of objects of equal size, moving in opposite directions along the track of a stadium at the same speed. If we assume that motion occurs through changes in position by equal, indivisible spatial units in equally short, indivisible time units, then in the moment when the first row moves one unit relative to the stadium, the second row also moves one unit relative to the stadium, but in the opposite direction. Thus, one row moves by two units relative to the other in the same time. This means that two spatial units equal one unit, and two time units equal one such unit.

Note that in the first two paradoxes, Zeno assumes the infinite divisibility of space and time, while in the latter two, he assumes the opposite, namely, that space and time consist of smallest, indivisible parts (though this assumption is explicitly stated only in The Stadium). While the first and third paradoxes demonstrate the impossibility of absolute motion, the second and fourth refute the possibility of relative motion.

Unlike the sophists, who primarily viewed logic as *eristics*, i.e., the art of argumentation aimed at refuting an opponent's standpoint, another approach developed, known as *dialectical logic*, which focused on conducting dialogue with the shared goal of seeking the truth. The unsurpassed master of this art was Socrates (469 – 399 BCE), one of the most influential figures in Western thought, emerging at its very dawn. In the *Socratic dialogues*, as presented to us by Socrates' student Plato (427 – 348 BCE), the search typically begins with a question posed by Socrates or one of his guests, who often include a recognized expert on the given topic. The initial answer usually aligns with commonly accepted opinion. Socrates then demands a more detailed explanation and requests clear definitions of the terms used. As he continues questioning, the respondent often arrives at a position that contradicts his original stance. If the

group reaches a consensus and adopts this new view, Socrates once again challenges it with further questions. This pattern repeats multiple times. Occasionally, participants feel they are approaching the truth, but more often, they realize they are merely deepening their awareness of their own ignorance — the ignorance to which Socrates himself openly admitted from the outset.

The first Greek philosophical school focused on logic arose in the Greek city of Megara, halfway between Athens and Corinth. Its founder, Euclid of Megara (cca 435 – 365 BCE), often confused with Euclid of Alexandria (around 300 BCE) the author of the *Elements*, was a student of Socrates and an admirer of Zeno. Among its members was the aforementioned Eubulides of Miletus, known for several famous sophisms and paradoxes. Closely associated with it was the so-called Dialectical School, represented by Diodorus Cronus (? – cca 284 BCE) and Philon of Megara (3rd century BCE). The works of the Megarians and Dialecticians have not survived; we know of them only indirectly through the writings of other authors, such as Plato's dialogue *Theaetetus*, which is believed to have been originally authored by Euclid. It is known that the Megarians enriched eristics by developing a method of deriving absurd conclusions from statements they sought to question or refute. Philon is also credited for having arrived at a conception of implication coinciding with the modern one.

However, the foundations of logic as a scientific discipline were laid by *Aristotle of Stagira* (384 – 322 BCE) in his five-volume work, collectively referred to since the Middle Ages as the *Organon* (i.e., *The Instrument*). The contents of its individual parts can be summarized, in a simplified manner, as follows:

Categories introduce a classification of simple concepts (nouns), which can serve as subjects or predicates in statements, into ten types (categories).

On Interpretation analyzes categorical statements in terms of their fundamental logical structure, including negation, opposition, conversion, and quantitative attributes. Using subject-predicate variables, it provides an initial "combinatorial analysis" of possible types of categorical syllogisms. This work also includes discussions on possibility, contingency, necessity, as well as the validity of statements about the future, marking the beginnings of modal and temporal logic.

Prior Analytics provides a formal analysis of the conditions that guarantee the correctness of logically valid inferences (categorical syllogisms).

Posterior Analytics deals with proofs of statements, definitions of concepts, and the classification of scientific knowledge. While *Prior Analytics* focuses on the formal aspect of syllogistic logic, *Posterior Analytics* examines logical reasoning primarily in terms of its content. It distinguishes between *apodictic syllogisms*, whose premises are certain and true, thus leading to true and certain conclusions, and *dialectical syllogisms*, whose premises and conclusions are uncertain.

Topics explore dialectical proofs in greater depth and seek to establish their degree of reliability. Syllogisms that create a deceptive impression of validity, either in form or content, are called sophistical. The appendix *On Sophistical Refutations* (often considered a separate sixth volume) is dedicated to uncovering and critiquing the errors concealed within such arguments. It classifies thirteen types of fallacious reasoning.

From the perspective of our course, the most important aspect is the study of categorical syllogisms, which permeates the entire *Organon*. A categorical syllogism is defined at the beginning of *Topics* as "*a discourse (logos) in which, if something is given, something different from what is given follows necessarily from the fact that it is given.*" Even though, from today's perspective, Aristotle's theory of categorical syllogisms represents only a fragment of *first-order monadic logic* (i.e., the logic of unary predicates), the tendency, emerging from time to time since the early 20th century, to dismiss it as trivial is merely evidence of a failure to understand and appreciate the epochal significance of the step taken by this great ancient thinker and polymath as the very first.

Aristotle's work was continued by his followers, the members of the *Peripatetic school*, which he founded. They expanded his syllogistic logic, for example, by studying forms of *hypothetical reasoning*. The most notable among them were Theophrastus of Eresus on Lesbos (cca 372 – 287 BCE), considered the founder of botany and dendrology, and Eudemus of Rhodes (cca 370 – 300 BCE), known both as an editor of Aristotle's works and as a historian of Greek mathematics and astronomy.

Aristotle and the Peripatetics also influenced the *Stoic school*, which was active in Cyprus and whose most prominent figures were Zeno of Citium (cca 334 – 262 BCE) and Chrysippus of Soli (282 – 206 BCE). Although Stoicism is primarily a philosophy of moral principles and life attitudes (consider, for example, the proverbial Stoic calm), Stoic thinkers made a significant contribution to the development of logic as well. While Aristotelian logic is primarily concerned with analyzing the relationship of inference between subject-predicate statements, Stoic logic laid the foundations of propositional logic. A *proposition* is understood as a meaningful statement that is either true or false. New propositions can be formed from given ones using logical connectives, such as negation, conjunction, disjunction (understood in the exclusive sense), and implication. Namely *implication* played a fundamental role in Stoic logic and was understood (like in Philon's work) in the same way as today: the proposition $A \Rightarrow B$ is false only if its antecedent A is true and its consequent B is false. Additionally, Stoic logic also considered implication (inference) in the sense of a *valid argument*, i.e., an argument in which a conclusion is logically derived from true premises, thereby guaranteeing its truth. The Stoics discovered and systematized several inference rules, which they formulated using ordinal numerals in the role of propositional variables. For illustration, here are four examples, named as they were in the Middle Ages, written in modern notation for clarity:

from $A \Rightarrow B$ and A derive B	(modus ponens)
from $A \Rightarrow B$ and $\neg B$ derive $\neg A$	(modus tollens)
from $\neg(A \wedge B)$ and A derive $\neg B$	(modus ponendo tollens)
from $(A \wedge B) \Rightarrow C$, A and $\neg C$ derive $\neg B$	(antilogism rule)

Using their deductive system, the Stoics also analyzed well-known ancient paradoxes and discovered or created many others. Little has survived from the original writings of the Greek Stoic school; most of what we know about them comes from Diogenes Laertius' (cca 180 – 240 AD) work *Lives and Opinions of Eminent Greek Philosophers*, written in the 3rd century AD. However, the Stoic school found continuity in ancient Rome. Among its adherents were Lucius Annaeus Seneca (4 BC – 65 AD), Marcus

Aurelius (121 – 180 AD), and, to some extent, Marcus Tullius Cicero (106 – 43 BC). However, logic was no longer at the center of their interests.

Aristotle's syllogistic logic and the Stoic propositional logic were significantly enriched by the famous ancient physician and anatomist Galen (cca 130 – 200 AD). In his *Introduction to Logic*, Galen distinguished between exclusive disjunction (either ... or ...) and non-exclusive alternative (... or ...), as well as between implication and bi-implication (equivalence). He also studied the interchangeability (i.e., logical equivalence) of statements formulated in different ways, thereby anticipating the concept of tautology. Additionally, he introduced early elements of the logic of binary predicates (relations), examining the possibility of inversion and the symmetry property. In logical reasoning, he also allowed premises formed by a conjunction of several simple assumptions. He applied his logical knowledge, for example, in his analyses of anatomical descriptions and medical treatments.

However, logic in antiquity did not develop solely within the framework of ancient philosophy, from which it gradually became independent. An equally important contribution to logic came from Greek mathematics, where the art of argumentation was developed and refined in a fundamental way. The pinnacle of this is Euclid of Alexandria's thirteen-volume work *Elements* ($\Sigma\tau o\iota\chi\varepsilon\iota\alpha$), written around the turn of the 4$^{\text{th}}$ and 3$^{\text{rd}}$ centuries BC. This work systematically compiles and explains the mathematical knowledge of the time, covering plane and solid geometry as well as arithmetic. Especially the first six books, dedicated to plane geometry, became the model for construction of any deductive theory aspiring to exactness for many centuries.

Euclid begins with twenty-three *definitions* of basic concepts (point, line, straight line, angle, circle, etc.), relying on geometric intuition, sharpened by the view of the ideal geometric world. He then establishes five *postulates*. The first three postulates can be understood as instructions for solving basic constructive tasks (drawing a straight line connecting two points, extending a given segment indefinitely, and drawing a circle with a given center and radius). The other two postulates are more of axiomatic nature: all right angles are equal, and the famous fifth postulate states that two lines, when intersected by a third line, will meet on that side where the sum of their angles with the transversal is less than two right angles. He then presents nine *axioms*, which serve as self-evident principles or simple methodological rules (e.g., things equal to the same thing are equal to each other; if equals are added to equals, the results are equal; the whole is greater than the part, etc.). From these basic concepts and premises, Euclid rigorously logically derives further propositions. However, the logic of the *Elements* is not an independent discipline but rather a masterfully employed, albeit only implicitly specified, tool. This logic does not replace geometric intuition but rather builds upon, serves, supports, complements, and extends it.

It is widely accepted that Euclid was familiar with Aristotle's works, including his syllogistic logic, as well as Stoic logical writings. However, he does not explicitly refer to them in his *Elements*. On the other hand, the logic he employs is noticeably richer than both Aristotelian and Stoic logic: upon a retrospective view, it can be regarded as a substantial fragment of first-order logic, unrestricted to unary predicates.

Euclid's *Elements* profoundly influenced and inspired scholars such as Nicolaus Copernicus (1473 – 1543), Galileo Galilei (1564 – 1642), Johannes Kepler (1571 – 1630), René Descartes (1596 – 1650), and Isaac Newton (1643 – 1727). Euclid's axiomatic

method also served as a model for Baruch Spinoza (1632 – 1677) in his most significant philosophical work, *Ethica ordine geometrico demonstrata*, where he derives metaphysical, theological, and ethical conclusions from definitions and axioms stated in advance. After the *Bible*, *Elements* is the most translated and published book of all time.

Our gallery of ancient logicians concludes with the early Christian Roman philosopher Anicius Manlius Severinus Boethius (cca 480 – 524 AD). Boethius introduced medieval Europe to the intellectual achievements of ancient philosophy by translating works of Plato, Aristotle, and other Greek philosophers into Latin. He also wrote detailed commentaries on several of them, including Aristotle's logical writings. He contributed to logic through his own works, especially *On the Categorical Syllogism* and *On the Hypothetical Syllogism*, in which he attempted to synthesize Aristotle's categorical syllogistic with Stoic propositional logic. He also provided a detailed combinatorial classification of the formal structures of different types of reasoning. His logic, based on the structural analysis of the Latin language, significantly influenced medieval thought and laid the foundation for the impact of Aristotle's work on Scholastic philosophy and logic.

2.2 The Middle Ages

Logic in the Middle Ages was cultivated as one of the *seven liberal arts*. These included the so called *trivium*, i.e., the three verbal arts [grammar, rhetoric, and logic (associated with philosophy and also called dialectics)] and the so called *quadrivium*, i.e., the four numerical arts [arithmetic, geometry, astronomy (together with astrology), and music (including the study of proportions and harmony)]. From this classification, it is evident that medieval logic was separate from mathematics and did not become closely connected with it.

The development of medieval logic can be divided into three periods. The first period, also called *old logic* (*logica vetus*), extends from the end of antiquity to approximately the middle of the 12^{th} century. During this period, access to ancient sources was limited to two works by Aristotle (*Categories* and *On Interpretation*) along with the corresponding commentaries by Boethius. The most significant representative of "old logic" was the French scholastic philosopher, logician, theologian, poet, and composer Pierre Abélard (1079 – 1142).

The second period, known as *new logic* (*logica nova*), lasted approximately from the middle of the 12^{th} century to the end of the 13^{th} century. During this time, Aristotle's entire body of work, along with other ancient sources, was studied and commented upon. A combinatorial classification of categorical syllogisms and rules of propositional logic was completed. Considerable attention was devoted to issues of modal and temporal logic. Several logic textbooks appeared, leading to the establishment of something like their canonical form. Among the most significant representatives of "new logic" were the church scholars Albert the Great (cca 1193 – 1280), Peter of Spain (1205 – 1277), John Duns Scotus (1265/6 – 1308), and others.

The third period, called *modern logic* (*logica modernorum*), spans approximately the entire 14^{th} and 15^{th} centuries. It is characterized mainly by the further refinement

and systematization of the methods and results of the previous period and their application in the analysis of medieval Latin. Ancient paradoxes were also studied, and attempts were made to resolve them, while new paradoxes emerged on the way. Among the most significant representatives of modern logic were William of Ockham (1287 – 1347), Jean Buridan (cca 1300 – 1360), Albert of Saxony (1320 – 1390), Paolo Venetus (1369 – 1429), and their students.

Alongside its development as an independent discipline, logic in the Middle Ages was also refined in philosophical and theological debates and disputes. Typical questions addressed in these discussions included the relationship between predestination and free will, as well as related issues such as human responsibility for one's actions and the ability to influence one's own salvation through deeds. Another key issue was the logical compatibility of divine attributes (e.g., it was known that God's omniscience and omnipotence contradict each other), and the possibility of proving God's existence by logical means from purely rational principles. Further debated questions included the problem of how a benevolent, merciful, and omnipotent God could allow the existence of evil in the world He created, and where was God and what was He doing before the creation of the world. Logic also played a crucial role in the metaphysical dispute between *realism* and *nominalism* regarding the nature of *universals*. Realism held that universals (general concepts denoting ideas, properties, and relations, such as *goodness*, *evil*, *beauty*, *friendship*, *greatness*, etc.) have an independent, real existence. In contrast, nominalism asserted that universals are merely names or labels for certain abstract ideas, properties, and relations, without corresponding to independently existing real objects. It is hardly surprising that none of these problems were ever resolved to general satisfaction.

In contrast to the creative contributions of all-round educated medieval scholars to the development of logic and their brilliant disputations, the *teaching* of logic in the Middle Ages and for a long time afterward stood in stark opposition. It was largely reduced to the *memorization* of derivation rules and categorical syllogisms, thereby negating its very purpose, which should have been the cultivation of logical thinking. Paradoxically, the most famous and widespread results of medieval logic became various *mnemonic aids*: the *logical square*, also known as the *square of oppositions*, which connects four types of subject-predicate judgments, and the mnemonic codes for the nineteen syllogisms divided into four figures (I: Barbara, Celarent, Darii, Ferio; II: Cesare, Camestres, Festino, Baroco; III: Darapti, Disamis, Datisi, Felapton, Bocardo, Ferison; IV: Bramalip, Camenes, Dimaris, Fesapo, Fresison).

Exercise. Search online for information about the logical square, the four basic figures of categorical syllogisms, and the mnemonic codes for individual syllogisms. Based on the significance of the vowel codes a, e, i, o, reconstruct the verbal form of the syllogisms corresponding to the given 19 codes and write them using both medieval and modern notation. For example, the syllogism Ferio has the verbal expression:

no M is P, and some S are M, therefore some S are not P

And in medieval and modern notation, respectively, it appears as follows:

from $M\,e\,P$ and $S\,i\,M$ derive $S\,o\,P$

from $(\forall x)(M(x) \Rightarrow \neg P(x))$ and $(\exists x)(S(x) \wedge M(x))$ derive $(\exists x)(S(x) \wedge \neg P(x))$

Also, uncover the meaning of the initial letters B, C, D, F and the consonants m, p, r, s in the names of the individual codes (they contain instructions on how each code of the second to fourth figures can be derived from the corresponding code of the first figure).

2.3 Renaissance and Modern Era

Renaissance thought took a highly critical stance toward scholasticism and medieval Aristotelianism. Speculative philosophy and theology lost their prominence, while the focus shifted to the observation of nature and experimentation. A growing effort emerged, especially represented by Galileo Galilei, to explain natural phenomena and processes in a causal manner, uncover their laws — not only their qualitative but also their quantitative aspects — and describe them in the language of mathematics. In this intellectual atmosphere, medieval logic was, at least for a time, viewed with disdain as fruitless speculation and endless repetition of trivial rules.

A shift in perspective came with the modern era, particularly with René Descartes, who recognized the crucial methodological role of logic in rationalist philosophy. This is evident in his works *Discourse on the Method* (1637) and *Rules for the Direction of the Mind* (written in 1628 but published only in 1701). Descartes' *Discourse* and his *Rules*, along with the views of Blaise Pascal (1623 – 1662), significantly influenced Antoine Arnauld (1612 – 1694) and Pierre Nicole (1625 – 1695) in writing their textbook *Logic, or the Art of Thinking* (first published anonymously in Paris in 1662), better known as *The Port-Royal Logic*, named after the monastery Port-Royal de Champs, where its authors worked. Aristotelian logic, as it had developed in the late Middle Ages, is presented in the book in the spirit of Cartesian dualism. Accordingly, the material is presented verbally, without the use of symbolic notation. In semantics, the authors distinguish between the *content* (*intension*) and the *extension* of a concept. At the same time, they introduce a psychological perspective into logic, seeing its purpose not in itself but in its Enlightenment-inspired role of helping to promote reason and justice in the world. The book was translated into several languages within a relatively short time. Until the 19^{th} century, it remained the most published logic textbook in HOWEVER both the Old and New Worlds.

A special place in our story is occupied by the prominent polyhistor (mathematician, philosopher, and diplomat), Gottfried Wilhelm Leibniz (1646 – 1716). As a mathematician, Leibniz is known (alongside Newton) primarily as one of the creators of the infinitesimal calculus. His approach was based on the use of infinitely small (and infinitely large) quantities; the notation he introduced, $\frac{\mathrm{d}y}{\mathrm{d}x}$ for the derivative and $\int f(x)\,\mathrm{d}x$ for the integral, is still in use today. Leibniz also corresponded with the aforementioned Arnauld. His aptitude for working with symbols was evident in his design of a logical notation, which, however, was not published during his lifetime. By the time it was eventually made public at the turn of the 19^{th} and 20^{th} centuries, it had already been surpassed, so it neither influenced previous developments nor shaped further progress.

Nevertheless, Leibniz influenced the development of European thought, particularly in logic, through his vision, which built upon the ideas of the Catalan medieval

mystic Ramón Llull (1232 – 1316). Llull created a system of tables and diagrams known as *Ars Magna*, which aimed to gather universal foundational principles of cognition in symbolic form and, through mechanical combinations following logical rules, derive further (potentially all) knowledge. Within this framework, it was expected that even Christian theological dogmas could be proven beyond any doubt. Leibniz shared Llull's idea of discovering truths exhaustively by generating propositions through suitable combinations of concepts. His treatise *De arte combinatoria* (1666) outlines this "project". Similar projects were ruthlessly parodied by Jonathan Swift (1667 – 1745) in the third book of *Gulliver's Travels* (1726), in a scene depicting Gulliver's visit to the Grand Academy of Lagado.

Later on, Leibniz conceived the idea of developing a *universal characteristic language* (*lingua characteristica universalis*), in which fundamental concepts would be represented by numbers or other characteristic symbols, while more complex concepts would be represented by chains of symbols corresponding to the simpler concepts from which they were composed. The resulting graphical formations were intended to be easily comprehensible to readers regardless of the language they spoke. He drew inspiration from Egyptian hieroglyphs and Chinese pictographic writing, as well as from the algebraic notation introduced by the French mathematician François Viète (1540 – 1603). Furthermore, he aimed to develop the so-called *calculus ratiocinator*, a universal "rational calculus" that would allow calculations with concepts and propositions formulated using the symbols of the universal characteristic language. He also envisioned the *ars iudicandi*, a method by which it would be possible to determine the truth or falsehood of any proposition expressed in this symbolic system.

The goals pursued by Leibniz are anything but modest. The advancement of all sciences, including mathematics, would be just one, and not even the most important consequence. All controversies and quarrels between individuals, as well as conflicts and wars between nations would cease. Even if disagreements or differences of opinion arose, it would suffice to sit down together and calculate. This way, it would quickly become clear who is right, and this with undeniable clarity and persuasiveness recognized truth would naturally be accepted and respected by all. Humanity could gradually discover, and eventually exhaust, all truths. Among these, the most prominent would be the Truth of God's existence and the true religion, "which most closely aligns with reason, and from which people in the future would be as unlikely to stray as they are to turn away from arithmetic and geometry once they have learned them."

However naive and utopian Leibniz's hopes may seem to us today, the grandeur of his project cannot be denied. As for the methods, his genius was nearly three centuries ahead of his time and precisely anticipated the directions of further development in mathematical logic: propositional and predicate calculus, Gödel's arithmetization of metamathematics, as well as the symbolic representation of data and knowledge used in computer implementation and processing. And when we realize how far we still are — despite all scientific and technological progress — from fulfilling Leibniz's dream of eternal peace, it is hard to resist a feeling of futility, despair, and disappointment in the unteachable and incorrigible human race.

In the 18$^{\text{th}}$ century, several scholars engaged in the study of logic, whose scientific and philosophical interests, like those of Leibniz, had a significantly broader scope.

For all of them, we should mention at least the Alsace born German mathematician, astronomer, physicist, and philosopher Johann Heinrich Lambert (1728 – 1777). Lambert was the first to prove the irrationality of the number π, and with his study *Theory of Parallel Lines* (1766), he became one of the forerunners of non-Euclidean geometry. His works *Photometria* (1760) and *Pyrometria* (1779) represent important contributions to the theory of light and heat, respectively. His main philosophical work, *New Organon* (1764), provided a profound analysis of numerous issues concerning, among other things, formal logic, probability, and the methodology of science. Along with the English mathematician, astronomer, and scientific instrument maker Thomas Wright (1711 – 1786) and the German philosopher Immanuel Kant (1724 – 1804), with whom he corresponded, he was among the first to recognize that the spiral nebulae visible through astronomical telescopes in the night sky are star-formed galaxies of a disk-like shape, similar to our Milky Way.

On the other hand, it was Kant who, in his most influential philosophical work, *Critique of Pure Reason* (*Kritik der reinen Vernunft*, 1781), put forward the idea that logic was practically a completed discipline, to which nothing essential could be added. Considering how little progress logic had made until then since the time of Aristotle, one cannot wonder too much about that. However, within about half a century, it became clear that this was a fatal mistake. And it was not the only Kant's mistake, one that — supported by his unquestionably well-earned authority — temporarily slowed the natural development of science. Even a better-known example is Kant's thesis on the absolute character of space and time as pure *a priori* forms of our intuition, beyond which sensory perception and knowledge are impossible. According to Kant, this space is identical with the space of Euclidean geometry, and no other type of space is even conceivable. This thesis was likely one of the reasons why Carl Friedrich Gauss (1777 – 1855), one of the greatest mathematicians of all time, hesitated to publish his discoveries regarding non-Euclidean geometry. However, by then, the time had come for non-Euclidean geometry to see the light of day.

2.4 Non-Euclidean Geometry

Non-Euclidean geometry played a remarkable and not fully appreciated role in the development of logic. Its emergence is attributed to the already mentioned fifth postulate of Euclid, which is most commonly stated as follows:

> *Through a given point not on a given line, exactly one parallel can be drawn to the given line.*

This formulation was presented by the Neoplatonist philosopher Proclus of Lycia (cca 410 – 485 CE) in his *Commentary on Euclid*. The same wording is also found in the works of the English cleric and mathematician William Ludlam (1717 – 1788); however, it is named after the Scottish natural scientist and mathematician John Playfair (1748 – 1819), who included it in his work *Elements of Geometry* (1795). *Playfair's axiom* is entirely consistent with geometric intuition, yet it is formulated in a significantly more complex manner than the other axioms and postulates of Euclid. This complexity raises suspicions that the fifth postulate — whether in its original or modified form — should be derivable from the remaining axioms and postulates,

making it redundant. Indeed, many scholars attempted to prove it, and some were even convinced that they had succeeded. However, it was always later revealed that their "proof" relied not only on Euclid's other axioms and postulates but also on some additional statement so intuitively aligned with geometric insight that they took it for granted. This additional statement, however, was merely another, equivalent formulation of the fifth postulate.

If we decide to deny the fifth postulate while insisting on the homogeneity of space, meaning that no point or line in it has a privileged status, we are left with only two options. Either we must accept the postulate that no parallel can be drawn through a given point not on a given line, which implies that any two distinct lines must intersect at a single point. Alternatively, we must accept the postulate that at least two distinct parallels can be drawn through a given point not on a given line, which leads to the conclusion that there will actually be infinitely many such parallels.

In the first case, we can further deduce that all lines have a finite length, which contradicts Euclid's second postulate stating that any given line segment can be extended indefinitely. This rules out the first case as impossible, and we need not consider it further. [How Bernhard Riemann (1826 – 1866) addressed this case and developed what is now called *elliptic geometry* will not be discussed in our course.] Thus, only the second case remains. However, within this framework, we gradually arrive at seemingly bizarre conclusions that we might be inclined to consider absurd. For example, the sum of the interior angles in any triangle is less than two right angles and decreases as the area of the triangle increases. Specifically, if we construct an isosceles right triangle with its base lying on a given line p and its height on a perpendicular to this line, the angles formed by the triangle's legs with line p will be smaller than half a right angle, and their magnitude will decrease as the height increases. But not only that: there exists a certain threshold value for the height beyond which the triangle can no longer be constructed, as its intended legs will never intersect line p; in other words, the perpendicular lines extending triangle's legs will be parallel to line p. At this point, it becomes a matter of personal taste when we decide that we have had enough and declare some similar conclusion contradictory. However, this will not be a contradiction with reason (i.e., a logical contradiction), but merely a contradiction with our geometric intuition. And — as has been repeatedly demonstrated — the rejection of the supposedly "absurd" conclusion was *logically equivalent* to accepting the fifth postulate. On the other hand, these "absurd" consequences can be viewed as the historically first theorems of a new discipline: *hyperbolic geometry*.

The attitude of explorers toward the strange world that was subtly revealing itself to them, into which they cautiously peered, gradually changed. Due to its contradiction with natural geometric intuition, it was rejected by Islamic scholars such as Thabit Ibn Qurra (cca 830 – 901), Hassan Ibn al-Haytham (cca 965 – 1040), and Omar Khayyam (1048 – 1131), as well as by Italian scholars of the 17^{th} and early 18^{th} centuries, including Giovanni Alfonso Borelli (1608 – 1679), Giordano Vitale (1633 – 1711), and Giovanni Girolamo Saccheri (1667 – 1733). According to their own views, they considered this contradiction to be proof of the fifth postulate. For example, Saccheri, in his treatise *Euclides Vindicatus* (*Euclid Freed from Every Blemish*, 1733), published shortly before his death, concluded that in a plane where the fifth postulate did not hold, no square could exist. Thus, if we accept the "self-evident fact" that

at least one square exists in the plane, we can use this to prove the fifth postulate. Nevertheless, Saccheri evidently succumbed to the allure of the strange non-Euclidean world and hesitantly admitted it as logically possible, though contradicting the real world. The first to openly acknowledge his fascination with the world of non-Euclidean geometry and to wish for it to be at least logically possible alongside the Euclidean world was the aforementioned Lambert in 1766.

In 1807, the German lawyer and amateur mathematician Ferdinand Karl Schweikart (1780 – 1857) published a study with the telling title *The Theory of Parallel Lines Together with a Proposal for Its Expulsion from Geometry.* In 1818, he wrote to Gauss about a dual geometry: one being geometry in the narrower sense (by which he meant Euclidean geometry) and the other being a more general, so-called *astral geometry*, in which the sum of the interior angles of a triangle is less than two right angles. His ideas were further developed by his nephew, the mathematician Franz Adolph Taurinus (1794 – 1874). In 1825, Taurinus published *The Theory of Parallel Lines*, in which he attempted to prove that Euclidean geometry was the only possible one. However, just a year later, he published *Geometriae Prima Elementa*, in which he modeled astral geometry as "geometry on a sphere with an imaginary radius" and called it *logarithmic-spherical geometry.* In this work, Taurinus also admitted the possibility of a geometry where the sum of angles in a triangle is greater than two right angles, thereby anticipating Riemann's elliptic geometry. Such a geometry can be realized as the geometry on a spherical surface (after identifying pairs of antipodal points). Nevertheless, Taurinus still regarded Euclidean geometry as having a privileged status and considered it the only true geometry.

For Gauss, the world of non-Euclidean geometry began to open up around 1795, and by approximately 1800 he was already navigating it safely. He gradually realized that the "absurd" consequences of rejecting the fifth postulate did not represent a logical contradiction but were instead inherent properties of a new geometry, which he later called *non-Euclidean geometry.* The reasons why he never published any of his findings remain a matter of speculation. In addition to the already mentioned Kantian thesis, he may have feared the "uproar of the Boeotians," that is, the threatening scandal that could have jeopardized his reputation. At the time, Gauss himself adhered to the prevailing view that geometry was the science of the structural laws of *real space* rather than an investigation of various "space-like" structures admissible solely on the basis of logical consistency. This perspective made geometry an empirical natural science akin to physics.

In this context, Gauss also realized that the validity of the fifth postulate might be determined experimentally. While participating in the geodetic triangulation of the Kingdom of Hanover between 1821 and 1825, he carefully measured the angles of the largest triangle in the geodetic network under construction, formed by the peaks of Brocken, Hoher Hagen and Grosser Inselsberg, with sides measuring 69, 85 and 107 km. However, the deviation of the sum of these angles from 180° did not exceed the possible margin of measurement error. Much later, Gauss expressed the opinion that to detect a potential discrepancy between real space and Euclidean geometry, it would be necessary to measure the angles in a triangle whose sides were many times larger than Earth's radius. At the same time, he realized that such an experiment

could only disprove the fifth postulate; to confirm it with certainty, one would need to measure angles with absolute precision.

Whether Gauss's concerns were justified or not, and whatever his reasons may have been, the fact remains that he never published any of his discoveries or reflections on this new geometry. As a result, the credit for discovering non-Euclidean geometry rightfully belongs to the Russian mathematician Nikolai Ivanovich Lobachevsky (1792 – 1856) and the Hungarian officer of the Austrian army and amateur mathematician János Bolyai (1802 – 1860). Because of this, hyperbolic geometry is often referred to as Lobachevskian or Bolyai-Lobachevskian geometry.

János Bolyai was the son of Farkas Bolyai, a high school professor of mathematics, physics, and chemistry, and a friend of Gauss from their student days. Farkas himself had previously attempted, without much success, to address the problem of the fifth postulate, which led him to discourage his son from similar efforts. However, János was not deterred, and between 1820 and 1823, he claimed to have discovered "marvelous things that filled him with wonder, [... and ...] created a strange new world out of nothing." However, the manuscript he submitted for review to his teacher at the Imperial and Royal Military Academy in Vienna in 1826 was rejected. As a result, his discoveries were only published in 1832 as a brief appendix to his father's high school mathematics textbook. In it, the author introduced so-called *absolute geometry*, which encompassed both Euclidean geometry and (though not yet called as such) *hyperbolic geometry* as special cases. Farkas Bolyai sent his son's work to Gauss for evaluation. Gauss replied that he had reached similar conclusions long before and offered only very reserved praise. While Farkas felt honored by Gauss's response, the young János found it deeply discouraging.

Lobachevsky first publicly presented his discovery of non-Euclidean geometry, which he called *imaginary geometry*, in a lecture at Kazan University in 1826. Over the years 1829 to 1855, he published numerous works developing and explaining his theory. In Russian academic circles, particularly at the universities of St. Petersburg and Moscow, Lobachevsky's work was met with misunderstanding and rejection. He only gained recognition thanks to Gauss, who, after reading his *Geometrical Investigations on the Theory of Parallel Lines* in a German translation published in Berlin in 1840, began learning Russian to read more of Lobachevsky's work. On Gauss's recommendation, Lobachevsky was admitted as a member of the Göttingen Academy of Sciences in 1842.

Lobachevsky believed that studying different types of geometries was meaningful regardless of which one provided an accurate description of the structure of real space. This idea is reflected in his work *Pangeometry* (1855), in which Euclidean geometry is reconstructed within the framework of hyperbolic geometry. Nevertheless, he also attempted to determine the actual geometry of real space. Instead of measuring angles in Earth-sized triangles, he examined triangles formed by fixed stars. With the same goal, he analyzed astronomical data obtained by measuring the parallaxes of Sirius and other stars. However, even with these efforts, he was unable to detect a significant deviation that would confirm the non-Euclidean nature of real space.

Neither the insight, naturalness, nor even the virtuosity with which Gauss, Bolyai and Lobachevsky navigated the strange world of hyperbolic geometry guarantees that it is not a mere delusion, that might eventually collapse in a logical contradiction like a

house of cards. However, we must honestly admit that we do not have such certainty even in the case of the Euclidean geometric world. Yet, this does not bother us, as our geometric intuition, on which we base our belief in its consistency, is widely shared and reinforced by traditional education. Therefore, it would suffice if we could prove the consistency of hyperbolic geometry under the assumption of the consistency of Euclidean geometry. Notably, the consistency of Euclidean geometry under the assumption of consistency of hyperbolic geometry follows from the aforementioned results of Lobachevsky, though he himself did not frame the question in this way.

There exist several *models* of hyperbolic geometry within Euclidean geometry. It will be enough to briefly acquaint ourselves with at least one of them. The first such model was created in 1868 by Eugenio Beltrami (1835 – 1900). Based on an observation by Arthur Cayley (1821 – 1895), it was refined in 1871 by Felix Klein (1849 – 1925) into its final form, now known as the *Beltrami-Klein model*. It is surprisingly simple: in this model, the points of the hyperbolic plane correspond to the interior points of a circle in the Euclidean plane, while the lines are represented by the chords of the bounding circle, excluding their endpoints.

In 1899, David Hilbert (1862 – 1943), considered the most significant and versatile mathematician of his time, published the monograph *Foundations of Geometry* (*Grundlagen der Geometrie*). The book reached its final form in the second edition of 1902 and is regarded as the embodiment of the ideal of constructing a mathematical theory through the axiomatic method. In it, Hilbert systematically builds Euclidean plane and spatial geometry axiomatically, supplementing Euclid's original system with the mathematical formulation of certain intuitive principles. These principles, while commonly used unconsciously in the ancient version, were not explicitly stated as they were considered self-evident. Hilbert proves the *consistency* of his axiomatization of Euclidean geometry using the coordinate method, under the assumption of the consistency of the arithmetic of real numbers. As a result, the question of the consistency of hyperbolic geometry is also reduced to the consistency of real arithmetic. Furthermore, Hilbert provides a proof of the *independence* of his axioms, showing that none of them can be derived as a logical consequence of the others. The independence of the fifth postulate (Playfair's axiom of parallel lines) is demonstrated precisely through reference to non-Euclidean geometry. As later results by Alfred Tarski (1926) revealed, Hilbert's axiomatization is also *complete*: any statement expressible in the language of Euclidean geometry is either provable within the system or its negation is provable.

The idea of modeling (interpreting) one theory through another would play an invaluable role in mathematical logic and mathematics as a whole in the 20^{th} century. Two spectacular examples of the successful application of this method are Gödel's model of the *universe of constructible sets* (1940) and Cohen's *forcing* method (1963). Gödel's constructible universe is a model of Zermelo-Fraenkel set theory (ZF) extended by the axiom of choice and the generalized continuum hypothesis, within the universe of ZF theory.[1] Using forcing, Paul Cohen (1934 – 2007) constructed a model

[1] The *continuum hypothesis* states that any infinite subset of the set of all real numbers $\mathbb{R}$ is either countable or has the same cardinality as $\mathbb{R}$. According to the *generalized continuum hypothesis*, for any infinite set X, there is no set Y whose cardinality $|Y|$ satisfies the inequalities $|X| < |Y| < |\mathcal{P}(X)|$, where $\mathcal{P}(X)$ is the power set of X.

of the set theory ZF in which the axiom of choice is false, within the universe of ZF theory. He also constructed a model of the set theory ZFC (ZF with the axiom of choice) in which the continuum hypothesis is false within the universe of ZFC theory. Together, these results demonstrate the independence of the axiom of choice from the axioms of the set theory ZF, as well as the independence of the continuum hypothesis from the axioms of the theory ZFC.

2.5 Logic in the 19th Century

In the previous section, we already reached the 19th century and, by its end, even the 20th century. However, we only indirectly focused on the development of logic, through the dramatic circumstances surrounding the emergence of non-Euclidean geometry and their influence on it. In this century, however, logic itself also undergoes development. This occurs both within philosophy and through the formation of mathematical logic (as mathematically presented formal logic), alongside the emergence of set theory. Gradually, the idea matures that these two disciplines should serve as the foundation of mathematics, thereby providing the long-sensed internal unity of mathematics with an "institutionalized form".

A remarkable figure in the first, philosophical line of logic's development in the 19th century is the German-writing Prague-born Catholic priest Bernard Bolzano (1781 – 1848). Bolzano lectured on theology and philosophy at the Charles-Ferdinand University in Prague while also engaging in logic, mathematics, aesthetics and ethics, as well as state theory and social issues. Due to his liberal views on social matters and anti-militarist stance, he was dismissed from the university in 1819, prohibited from public speaking, and two of his works were placed on the index. He spent the rest of his life in seclusion but continued working and lecturing within a small circle of his students and friends.

Bolzano's mathematical work is characterized by the anticipation of future discoveries, or even by discoveries that remained unnoticed in his time and were rediscovered only later on. At a time when topology as a mathematical discipline did not yet exist, Bolzano arrived at a definition of dimension similar to the small inductive dimension, later hinted at by Henri Poincaré (1854 – 1912) in 1912 and introduced by Pavel Samuilovich Uryson (1898 – 1924) in 1922 and by Karl Menger (1902 – 1985) in 1923. He was the first — before Karl Weierstrass (1815 – 1897) — to construct an example of a continuous function that is nowhere differentiable. He also demonstrated that every bounded numerical sequence contains a convergent subsequence. Even before Augustin-Louis Cauchy (1789 – 1857), he formulated the convergence criterion now known as the Cauchy-Bolzano condition.[2] He also proved that a continuous function attaining values of opposite signs at the endpoints of an interval must attain the value 0 within that interval.

This also applies to Bolzano's treatise *Paradoxes of the Infinite* (*Paradoxien des Unendlichen*), published posthumously in 1851. In it, Bolzano constructs the first historical version of set theory (*Menge*), presents arguments in favor of accepting

[2] A sequence of real numbers a_n satisfies the *Cauchy-Bolzano condition* if, for every real number $\varepsilon > 0$, there exists a natural number m such that for every $n > m$, the inequality $|a_n - a_m| < \varepsilon$ holds.

actual infinity, provides a proof of the existence of an infinite set supported with theological argumnets (specifically by invoking God's omniscience), and demonstrates that an infinite set contains a proper subset equivalent to itself (that is, in a certain sense, of the same size, contradicting Euclid's axiom that "*the whole is greater than the part*"). Later on, in 1888, Richard Dedekind (1831 – 1916) will use this property as the *definition* of an infinite set. Bolzano's *Paradoxes of the Infinite* influenced in several aspects Georg Cantor, the creator and founder of set theory, who referred to it multiple times, particularly when defending his concept of actual infinity. A certain difference in approach to infinite sets, which Cantor regarded as Bolzano's error, later served as inspiration in the 1970s for the Czech mathematician and philosopher Petr Vopěnka (1935 – 2015) in developing his *Alternative Set Theory*.

However, Bolzano's most significant contribution to logic is his monumental four--volume work *Theory of Science, An Attempt at a Detailed and Largely New Exposition of Logic with Constant Regard to Its Previous Expositors* (*Wissenschaftslehre, Versuch einer ausführlichen und grösstenteils neuen Darstellung der Logik mit steter Rücksicht auf deren bisherige Bearbeiter*), published in 1837. The logic in *Theory of Science* is understood primarily as the *methodology of science* rather than formal logic, which remains in the background. Simply put, Bolzano's *Theory of Science* is dedicated to analyzing the philosophical and epistemological foundations for justifying scientific knowledge, as well as formulating logical and methodological principles to be followed when seeking truth, structuring scientific knowledge into distinct disciplines, and presenting it in the form of textbooks with the goal of achieving the highest possible degree of clarity and persuasiveness.

Bolzano is convinced that for every scientific discipline, a fundamental set of initial truths can be established from which all further knowledge within that field follows as consequences. In doing so, he distinguishes between the concepts of *consequence* (*Abfolge*) and formal *derivation* (*Ableitung*), thereby implicitly arriving at a distinction between the semantic and syntactic aspects of scientific theories. He emphasizes semantics, arguing, for example, that a good proof should not only be logically correct but should also include a *justification* (*Begründung*) that provides insight and understanding.

Bolzano's style of exposition and argumentation is logically refined, meticulously elaborated, and internally consistent, though not necessarily easy to read for a modern audience. He introduces original concepts such as *sentence in itself* (*Satz an sich*), *truth in itself* (*Wahrheit an sich*), and *idea in itself* (*Vorstellung an sich*). A *sentence in itself* is a statement that asserts something, whether true or false, regardless of whether it has been expressed in words or even merely thought in someone's mind. Each individual sentence in itself exists, so to speak, "in a single instance", and repeating it through speech or writing does not multiply it. In today's terminology, we might say that a sentence in itself represents the content, sense, or meaning of a statement, whether spoken, written, or thought, or even one that could potentially be spoken, written, or thought. A sentence in itself that states something as it really is (i.e., is true) is called by Bolzano a *truth in itself*. The components of sentence in themselves that do not themselves constitute a full proposition are termed *ideas in themselves*. These can be simple or complex. Today, what Bolzano called an idea in itself would likely be referred to as a concept, or depending on circumstances, the

content or meaning of a concept. If we look for examples in specific sciences, mathematical concepts, statements, and knowledge would likely correspond most faithfully to ideas, sentence and truths in themselves. This is no coincidence — mathematics, for Bolzano, serves as a model for building and exposition of other scientific disciplines.

Bolzano attributes to his propositions, truths and ideas in themselves a certain kind of existence outside of time and space, while cautioning that this is not existence in the real world. On the other hand, he acknowledges the real existence not only of verbally expressed representations and propositions but also of merely thought ideas in the mind of a particular person. This philosophical stance is commonly referred to as *modern* or *semantic Platonism*. At the same time, Bolzano's work establishes a clear separation of logic from any psychologizing tendencies and — despite his Kantian sounding "*an sich* terminology" — takes a critical stance against Kantian metaphysics. Bolzano was likely the first to recognize that the foundation of philosophy does not lie in metaphysical speculation but in the study of what and how we speak, and the laws governing our language. In his *Theory of Science*, he laid the groundwork for the semantic tradition in Western thought. In this spirit, he was followed much later by Austrian philosophers Franz Brentano (1838 – 1917), Alexius Meinong (1853 – 1920), and Edmund Husserl (1859 – 1938); by the founding figures of the Polish Logical School, Kazimierz Twardowski (1866 – 1938), Jan Łukasiewicz (1878 – 1956), and Stanisław Leśniewski (1886 – 1939), and most notably by Alfred Tarski (1901 – 1983) with his 1930s conception of logical consequence and truth in formal languages, which remains fundamental in mathematical logic to this day.

In the 19th century, algebra experienced significant development, along with an expansion of its traditional areas of application. Gradually, a perspective emerged in which algebraic expressions and manipulations were no longer necessarily viewed as representing only numbers and operations on numerical domains. Likewise, they did not have to satisfy all the identities governing numerical operations. Among the many examples of this shift, one notable discovery was that of *quaternions*, a four-dimensional analogue of complex numbers, made in 1843 by the Irish mathematician William Rowan Hamilton (1805 – 1865). Quaternion multiplication turned out to be non-commutative, challenging conventional mathematical assumptions. In this intellectual atmosphere, logic emerged as a mathematical discipline in the British Isles. In 1847, two influential works were published: *The Mathematical Analysis of Logic, Being an Essay Towards a Calculus of Deductive Reasoning* by George Boole, and *Formal Logic, or the Calculus of Inference, Necessary and Probable* by Augustus De Morgan.

George Boole (1815 – 1864) was a self-taught mathematician who gained recognition in the scientific community for his work on differential equations and invariants. One of his papers even earned him a gold medal from the Royal Society in 1844. In *The Mathematical Analysis of Logic*, Boole transformed logic — previously considered a branch of philosophy — into a mathematical discipline. He also contributed to probability theory. Boole realized that logical connectives can be interpreted as algebraic operations on statements, properties, or classes of objects. Logical laws can then be formulated as identities for these operations. This approach allowed deductive reasoning to be expressed as *calculations*, in line with Leibniz's vision of a *calculus*

ratiocinator. In probability theory, Boole sought to develop a general method for determining the probability of one event based on the probabilities of logically related events. In recognition of his contributions, Boole was appointed as the first professor of mathematics at Queen's College, Cork, in 1849, with De Morgan's support, despite not having a university degree himself. A more detailed systematic exposition of his approach and opinions appeared in his 1854 work: *An Investigation of the Laws of Thought, on Which Are Founded the Mathematical Theories of Logic and Probabilities*.

Boole used the symbol 1 for *truth* or the *universal class*, and 0 for *falsehood* or the *empty class*. He represented negation as a complement relative to the universal class $1 - x$; conjunction (*and*) as the logical product $x \cdot y$; disjunction (*or*) as the logical sum $x + y$, which he interpreted in an exclusive sense (x or y, but not both). Boolean operations followed conventional identities, such as the commutative laws $xy = yx$, $x + y = y + x$; the associative laws $(xy)z = x(yz)$, $(x + y) + z = x + (y + z)$; the distributive law $x(y + z) = xy + xz$; or the law of double negation (complement) $1 - (1 - x) = x$. However, there emerged also some "non-standard identities", such as $xx = x$, or $x + x = 0$. The last identity, arising from Boole's exclusive interpretation of the logical connective *or* $(+)$ puzzled even him, as he expected that $x + x = 0$ should imply $x = 0$, following numerical analogy.

In the 20^{th} and 21^{st} century, Boole's ideas led to applications their author even could not have dreamed of. Binary codes, Boolean gates and switches, logic circuits, organized into complex structures in chips became basic units for design and construction of computers and many other electronic devices. Moreover, Boolean algebras remain one of the fundamental mathematical structures in set theory, mathematical logic, measure theory, probability theory and related disciplines.

Augustus De Morgan (1806 – 1871) is likely best known to readers as the author of *De Morgan laws*, connecting the logical connectives of conjunction and disjunction (in the non-exclusive sense) through negation. In modern symbolic notation, they can be written, e.g., as the tautologies:

$$\neg(A \wedge B) \Leftrightarrow (\neg A \vee \neg B) \qquad \neg(A \vee B) \Leftrightarrow (\neg A \wedge \neg B)$$

These laws appeared in his 1846 work *On the Structure of the Syllogism* in a less transparent symbolic form. However, in verbal (i.e., non-symbolic) form they were already known to William of Ockham as early as the 14^{th} century.

De Morgan also coined the term *mathematical induction* for the well-known method of proving statements about natural numbers. He did so in his widely used textbook *Elements of Arithmetic* (1830). His most significant contribution to logic, however, was broadening its scope beyond the traditional subject-predicate reasoning, which had dominated since antiquity. Instead, he emphasized reasoning with statements involving two- or multi-place relationships between objects. This shift helped establish the view that relational reasoning plays a crucial role in mathematical deduction and scientific argumentation. Moreover, certain logical operations on relations — such as inversion, composition and projections — cannot be expressed using only logical connectives. Except for inversion, these operations require *quantification* over variables. Thus, De Morgan became one of the pioneers in the study of quantifiers in logic. This development took place in a series of articles titled *On the Syllogism* (1846 – 1868),

which included both of the previously mentioned works. De Morgan is also credited with an anecdotal example of a logically valid relational inference:

All horses are animals.
Therefore, *every horse's head is the head of an animal.*

While obviously correct, this inference cannot be formally justified within Aristotelian subject-predicate logic.

Under the influence of De Morgan's and especially Boole's work, logic became a fashionable subject among intellectuals in the English-speaking world in the second half of the 19th century. Among them, two figures of truly Renaissance stature stood out, working on opposite sides of the Atlantic.

William Stanley Jevons (1835 – 1882) is best known as a leading British economist who extensively applied statistical and mathematical methods. He discovered the so-called *Jevons paradox* concerning the relationship between production, demand and commodity prices, and was also the author of the theory of *marginal utility*. In logic, he built upon the work of Boole and De Morgan. He replaced Boole's logical sum (exclusive disjunction) with disjunction in a non-exclusive sense and systematically worked with variables and quantifiers. In his treatise *The Substitution of Similars* (1869), he highlighted the Leibnizian universal principle, according to which "*whatever is true of a thing is true of its equivalent*". This principle would later be expressed in the form of logical axioms of equality. His book *Elementary Lessons on Logic* (1870) soon became the most widely used logic textbook in the English language. In another work, *The Principles of Science* (1874), he examined the relationships between logic (deduction), generalization based on facts (induction), and probability. He illustrated his ideas with numerous relevant, meticulously elaborated examples from natural sciences and economics. He also foresaw the use of one-way functions and integer factorization in cryptography. In 1869, he constructed a mechanical calculating machine, called the *logical piano* due to its shape, which was the first of its kind to perform Boolean logical operations faster than a human.

Charles Sanders Peirce (1839 – 1914) is best known as one of the founders of the first philosophical movement born on American soil, known as *pragmatism*. However, he also engaged in physics, chemistry, astronomy, geodesy and cartography, as well as logic. He contributed to measuring fluctuations in the Earth's gravitational field, which helped refine the measurement of the ellipticity of Earth's globe. He is also the author of the so-called *quincuncial projection* — a cartographic representation that allows for periodic tiling of the plane with consecutive projections of the Earth's octants in the form of isosceles right triangles. In logic, he sought to connect Boole's algebraic approach with De Morgan's relational inferences. Like Jevons, he replaced Boole's logical sum with disjunction in a non-exclusive sense. He also realized that when working with classes, it is more advantageous to use the relation of inclusion instead of equality, as it corresponds to implication. Even before Sheffer, he discovered that all logical operators could be expressed using a single one — either *not* (... *or* ...) (later called *Peirce's dagger* or NOR) or *not* (... *and* ...) (later called *Sheffer's stroke* or NAND). He explicitly introduced the quantification of variables, interpreting the universal quantifier Π_i as an infinite logical product (conjunction) and the existential quantifier Σ_i as an infinite logical sum (disjunction in a non-exclusive sense). He was

likely the first to recognize the fundamental distinction between propositional calculus, logic with quantification of variables for individual objects (first-order logic), and logic that also quantifies properties of individual objects (second-order logic).

Peirce also anticipated the application of Boolean logic in the design of switches in electrical circuits and the subsequent possibility of constructing an electronic computing machine. His former student, Allan Marquand (1854 – 1924), inspired by Jevons's logical piano, built an improved, smaller, and portable version of it between 1881 and 1882. In 1887, at Peirce's suggestion, he designed an electronic version of a logic computer, though he never realized it.

With Peirce and Jevons, the era in which logic was primarily the domain of polymaths, i.e., scholars engaged in multiple fields of human knowledge, came to an end. For thinkers from Aristotle through Leibniz and Lambert to Jevons and Peirce, logic was just one of many areas of their interest and expertise. However, following the trend initiated by Boole and De Morgan, logic became a mathematical discipline, and its further development was shaped mainly by mathematicians — often with broad interests, but typically limited to several branches of mathematics and the related philosophical questions.

2.6 End of the 19th and Beginning of the 20th Century

The mathematization, or more precisely, the algebraization of logic was completed in the monumental three-volume work of the German mathematician Ernst Schröder (1841 – 1902), *Lectures on the Algebra of Logic* (*Vorlesungen über die Algebra der Logik*, 1890 – 1905). In this work, Schröder built upon Boole, De Morgan, Jevons, and especially Peirce. He systematically applied the duality between conjunction and disjunction as mediated by De Morgan's laws, worked with the relation of inclusion and quantifiers, and developed the algebra of relations with composition as multiplication. However, this was still mathematical logic only in the first of the two meanings we mentioned at the beginning: it was the *mathematization of logic* by algebraic means.

By the turn of the 19th and 20th centuries the time was ripe for the long-anticipated potential of logic for mathematics to begin fully manifesting itself. However, it was first necessary to recognize and formulate the aims and objectives that such mathematical logic should serve. From approximately the 1880s, two distinct lines of development can be observed in this process; for simplicity, we will refer to them as the logical line and the set-theoretic line. Although they are closely intertwined, their beginnings can be clearly distinguished, allowing us to follow them separately for a certain period. We will begin with the logical line.

A pioneer in the systematic use of logic in mathematics was the eminent and versatile Italian mathematician Giuseppe Peano (1858 – 1932). He worked in number theory, linear algebra, geometry, differential equations, and logic. He was the first to construct an example of a continuous curve filling a square. Peano was also the creator of the international language *Latino sine flexione*, based on Latin with a significantly simplified grammar. His linguistic interests, combined with the emergence of similar counterintuitive objects in mathematics, led him to the decision to elevate logic to a universal language capable of expressing all of mathematics with unprecedented pre-

cision and unambiguity. The first step in this direction was Peano's axiomatization of the arithmetic of natural numbers in 1889, using the *successor* as the sole fundamental operation. Based on it, the operations of addition and multiplication were then defined recursively. However, this was made possible only due to the *axiom of induction* (presented here in a modern formulation):

$$(\forall X \subseteq \mathbb{N})\big(0 \in X \wedge (\forall x \in X)(x+1 \in X) \Rightarrow X = \mathbb{N}\big)$$

In this statement, in addition to the variable x for natural numbers, the variable X for subsets of the set of all natural numbers $\mathbb{N}$ is also quantified. From a later perspective, Peano's original arithmetic was thus a *second-order theory*. From 1894 onward, Peano with several collaborators began publishing (in a somewhat utopian endeavor) the *Mathematical Formulary* (*Formulario Mathematico*), an encyclopedic project aimed at compiling all known mathematical formulas and theorems in a standardized mathematical and logical notation created for this purpose. Peano's logical and set-theoretic notation (with various modifications) is still widely used today; an example is the symbol ε introduced by Peano for element membership in a set, later modified to its present form $\in$.

An even more ambitious project was undertaken by the German mathematician, logician, and philosopher Gottlob Frege (1848 – 1925). In his two books, whose titles reflect the Leibnizian heritage *Concept-Script, a Formal Language of Pure Thought Modeled upon that of Arithmetic* (*Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*, 1879) and *The Foundations of Arithmetic: A Logico-Mathematical Investigation into the Concept of Number* (*Die Grundlagen der Arithmetik: eine logisch-mathematische Untersuchung über den Begriff der Zahl*, 1884) as well as in journal articles such as *On Sense and Reference* (*Über Sinn und Bedeutung*, 1892), *On Concept and Object* (*Über Begriff und Gegenstand*, 1892), and other works, Frege gradually developed his conception of logic. He expressed surprisingly modern views on its role in philosophy and laid the groundwork for what would later be called the *analytic turn in philosophy*. Frege also created his own unique logical symbolism, but due to its cumbrousness, it did not gain widespread acceptance, and namely his insistence on using it contributed to the lack of attention his work received from his contemporaries. At the same time, he became convinced that the concept of number and other arithmetic notions could be founded purely by logical means, thereby building arithmetic and mathematical analysis as branches of logic. Unlike Peano, who sought to use logic "only" as a universal language and method for presenting mathematics, Frege went much further: he aimed to interpret mathematical concepts as logical ones and to reconstruct all of calculative mathematics within the framework of logical forms. In doing so, he became the founder of a school of thought in the foundations of mathematics known as *logicism*. The central tenet of logicism is that mathematics can essentially be reduced to logic. Frege's effort was meant to culminate in the three-volume work *Basic Laws of Arithmetic* (*Grundgesätze der Arithmetik*), with the first volume published in 1893 and the second in 1903.

However, in 1902, while the second volume was still in press, Frege received a letter warning him of a contradiction in his proposed formal system. The system allowed for reasoning about a property (predicate) possessed by those properties (predicates) that do not apply to themselves, in other words, properties that do not possess the

property they designate. If we denote this property as $R(x)$, then any predicate $P(x)$ has the property $R(x)$ if and only if it does not have the property $P(x)$. Formally, $R(P) \Leftrightarrow \neg P(P)$. If we now ask whether the predicate $R(x)$ possesses the property $R(x)$, the result leads to a contradiction: $R(R) \Leftrightarrow \neg R(R)$. This paradox undermined Frege's project in its current form. However, it is worth noting that Frege successfully reconstructed at least the arithmetic of natural numbers in a logicist manner and proved Peano's axioms; this part of his work remained unaffected by the contradiction. Even after this discovery, Frege continued to defend the logicist thesis, but he abandoned work on the planned third volume of *Grundgesätze*.

The sender of the letter was the British logician, mathematician, philosopher, and political activist Bertrand Russell (1872 – 1970). The contradiction he pointed out to Frege became known as *Russell's paradox*. In 1900, Russell attended the First International Congress of Philosophy in Paris, where he met Peano and his collaborator Alessandro Padoa (1868 – 1937). Their *Formulario* made a deep impression on him, as it resonated with the goals of his own work, *The Principles of Mathematics*, which was published in 1903. As Russell stated in the preface, this work had two main objectives:

> *One of these, the proof that all pure mathematics deals exclusively with concepts that can be defined in terms of a very small number of fundamental concepts, and that all its propositions are deducible from a very small number of fundamental logical principles, ... will be established by rigorous symbolic reasoning... The other object of this work, ... the explanation of the fundamental concepts which mathematics accepts as indefinable, is a purely philosophical task.*

This is in full agreement with the logicist thesis; however, it can also be said that Russell set as his goal the interpretation of mathematics by the *axiomatic method*, utilizing symbolic logic.

Russell wrote his *Principles* before becoming acquainted with Frege's work and had planned to write a second volume. However, he was so captivated by Frege's ideas that he embraced the logicist thesis as his own and undertook to continue realizing Frege's vision. At the same time, confronted with his own paradox, he realized that preventing similar contradictions would require fundamental measures. For his new project, he managed to enlist his friend and former teacher, the distinguished British philosopher and mathematician Alfred North Whitehead (1861 – 1947). Whitehead even abandoned work on the second volume of his similarly oriented *A Treatise on Universal Algebra* (1898) to join the effort. Together, they wrote and, between 1910 and 1913, published the monumental three-volume work *Principia Mathematica*, whose title alluded to the great Newton. In it, they presented, in the spirit of logicism, not the entirety of contemporary mathematics but at least a considerably comprehensive synthesis of its fundamental concepts and results. Even so, they could not avoid relying on certain postulates that can hardly be justified purely logically (such as the axioms of infinity, choice, and reducibility). Moreover, as a safeguard against paradoxes, they introduced a sophisticated system of syntactic restrictions preventing *self-reference*—that is, among other things, the application of a predicate to itself (such as $P(P)$ or $\neg P(P)$) or the formulation of set membership statements where a set belongs to itself

(such as $x \in x$ or $x \notin x$). This system became known as the *ramified theory of types*, and we need not go into its further details here. It suffices to say that, in addition to banning paradoxical expressions, it also preemptively forbade many "harmless" ones, making logicist mathematics highly unintuitive and excessively complex.

The lasting contribution of *Principia Mathematica* is its logical symbolism, which became widely adopted. However, in this respect, Russell and Whitehead did not follow Frege but rather Peano. To Peano's existential quantifier $(\exists x)$, they added a "timid" notation for universal quantification (x). The modern notation $(\forall x)$ was introduced only in 1935 by Gerhard Gentzen (1909 – 1945) but did not gain definitive acceptance until the 1960s. Until then, a more "straightforward" notation, $(\mathrm{E}x)$ and $(\mathrm{A}x)$, was also commonly used in parallel.

2.7 Set Theory

Now is an appropriate time to step away from the logical trajectory we have been following and return to the origins of the set-theoretic approach. The theory of sets, in its initial informal form, was created by a single man — a mathematician with a strong visionary mindset — Georg Cantor (1845 – 1918). He began developing his theory in the 1870s, prompted by considerations about the uniqueness of function expansions into trigonometric Fourier series. Cantor's "first" sets were thus sets of points on a line. However, he gradually arrived at a much more general understanding of the concept. In his own words:

> *By a set, we understand any collection M of certain well-distinguished objects m of our perception or thought (called elements of the set M) into a single whole.*

The formation of a set thus presupposes that all its elements already exist, that is, they are either perceived or at least conceived as actually existing. In the case of infinite collections, such as the natural numbers or points on a line, Cantor's theory departs from the previously dominant conception of infinity as *potential* and replaces it with the notion of *actual infinity* by treating these collections as completed sets.

Furthermore, the elements of a set participate in it solely by their presence, regardless of their arrangement or other relations among them. Two sets are therefore equal if and only if they have the same elements. This principle would later be formally expressed as the *axiom of extensionality*. Another fundamental postulate of Cantor's theory is the general *comprehension principle*:

> *For every "meaningful" property $P(x)$, one can form the set $\{x \colon P(x)\}$ of all elements x that have this property.*

The two principles mentioned above result in set theory occupying a privileged position in relation to logic. It is not merely one of many theories that provide logic with a space for application. In a certain sense, it is logic itself, manifested in a form where its two closely related, largely subconscious guiding intentions have been taken to their extremes. The first of these intentions lies in the tendency to identify logically equivalent properties, in other words, to replace properties — regardless of their meaning and linguistic description — with their *extensions*, that is, with groupings of

objects that possess the given property. This intention began to be realized already during the algebraization of logic, for example, through Boolean algebras, and set theory is merely a more finely structured culmination of this process. However, it is important to recognize that this intention is highly problematic. A vivid illustration of this is a well-known example given by De Morgan. Suppose our domain of objects consists of the eight planets of the Solar System: Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus, and Neptune. The number of possible properties of these planets that one might consider is unimaginably vast: ranging from mass, state of matter, and surface temperature to rotational period around its axis, orbital period around the Sun, distance from the Sun, number of satellites (moons), presence of atmosphere or water, or even the existence of life, and so on. On the other hand, the strict application of the extensional viewpoint forces upon us the absurd conclusion that the total number of possible properties of the elements of this (or any other) eight-element set is merely $2^8 = 256$.

The second of the aforementioned guiding intentions is the interpretation of as many groupings of objects as possible as selfstanding objects. One of the consequences of this approach is the actualization of the domain of all subsets of a given set, that is, the ability to construct, for any set X, even an infinite one, the set $\mathcal{P}(X) = \{A\colon A \subseteq X\}$ of all its subsets. As we will see later, the realizability of these intentions has its limits. Their unrestrained application leads to paradoxes or even contradictions within Cantor's theory. Therefore, it is necessary to apply them with caution and to seek rules that would limit and regulate their application.

Cantor's fascinating and ground-breaking insight into the world of infinite sets did not emerge all at once but was gradually illuminated through a series of initially subtle steps. Questions about the convergence of sequences of points on a line and sequences of functions, as well as issues related to the classification of different types of discontinuities in functions, required a more detailed clarification of the structure of real numbers. In Cantor's approach, each real number is determined as the limit of a sequence of rational numbers satisfying the so-called *Cauchy-Bolzano condition*. Moreover, two such sequences $\{a_n\}$ and $\{b_n\}$ determine the same real number if and only if the sequence of their differences $\{a_n - b_n\}$ converges to zero. In this way, individual real numbers are effectively represented as certain actual infinite objects. Their actualization thus already presupposes the actualization of the domain of definition of the corresponding sequences, that is, the infinite set of all natural numbers. The actualization of the set of all real numbers, which we identify with the set of points on the line, is then a further logical, though by no means trivial, step in this process.

During his deeper study of trigonometric series and other questions of mathematical analysis of his time, Cantor began to systematically use the operation of *derivation of a point set* A, defined as the set A' of all accumulation points of A (i.e., points p in whose arbitrarily small neighborhood there exists some point a of the set A, distinct from p). This operation can be iterated any finite number of times according to the recursive scheme:

$$A^{(0)} = A \qquad\qquad A^{(n+1)} = \left(A^{(n)}\right)'$$

However, once all finite-order derivations $A^{(n)}$ are available, one can define the derivation of infinite order

$$A^{(\infty)} = A^{(1)} \cap A^{(2)} \cap \ldots \cap A^{(n)} \cap \ldots = \bigcap_{n=1}^{\infty} A^{(n)}$$

and continue further:

$$A^{(\infty+n+1)} = \left(A^{(\infty+n)}\right)' \ldots \quad A^{(2\infty)} = \bigcap_{n=1}^{\infty} A^{(\infty+n)} \ldots \quad A^{(\infty^2)} = \bigcap_{n=1}^{\infty} A^{(n\infty)} \ldots$$

$$A^{(\infty^3)} = \bigcap_{n=1}^{\infty} A^{(n\infty^2)} \ldots A^{(\infty^\infty)} = \bigcap_{n=1}^{\infty} A^{(\infty^n)} \ldots \quad A^{(\infty^{\infty^\infty})} = \bigcap_{n=1}^{\infty} A^{(\infty^{\infty^n})} \ldots$$

By focusing on the ordering of these successive steps, Cantor discovered infinite ordinal numbers, which extend the (finite) natural numbers. Moreover, the arithmetic of natural numbers, including addition, multiplication and exponentiation, can be naturally extended to ordinal numbers. However, this required a more specific notation for infinite ordinal numbers. For this reason, Cantor replaced the symbol ∞ with ω. In a slightlymore detailed view and an ever-accelerating moving forward, the beginning of the ordinal number series appears approximately as follows:

$$0,\ 1,\ 2,\ \ldots, n,\ n+1,\ \ldots\ \omega,\ \omega+1,\ \omega+2,\ \ldots,\ \omega+n,\ \ldots\ 2\omega,\ 2\omega+1,\ \ldots$$
$$\omega^2,\ \omega^2+1,\ \ldots\ \omega^2+\omega,\ \ldots\ 2\omega^2, \ldots\ n\omega^2,\ \ldots\ \omega^3,\ \ldots\ \omega^n,\ \ldots\ \omega^\omega,\ \ldots \omega^{\omega^\omega},\ \ldots$$

Another, distinct type of infinite numbers Cantor discovered when comparing the "sizes" of infinite sets, based on the comparison of finite sets by the number of their elements. This "generalized number of elements" of a set was called its *cardinal number* or *cardinality*. Two sets A and B have the same cardinal number, denoted $|A| = |B|$, if there exists a bijective mapping $f\colon A \to B$. We say that the cardinal number of set A is smaller than the cardinal number of set B, symbolically $|A| < |B|$, if there exists an injective mapping $f\colon A \to B$ and $|A| = |B|$ does not hold. Cardinal numbers extend natural numbers (which serve as the cardinal numbers of finite sets), and the arithmetic of natural numbers, including addition, multiplication and exponentiation, can also be naturally extended to them. However, ordinal and cardinal arithmetic differ fundamentally: while cardinal addition and multiplication are commutative, for ordinal numbers we have, for example:

$$1 + \omega = \omega < \omega + 1 \qquad\qquad \omega \cdot 2 = \omega < 2 \cdot \omega$$

For the cardinal number of the set $\mathbb{N}$ of all natural numbers, Cantor introduced the notation $\aleph_0 = |\mathbb{N}|$ ($\aleph$, pronounced *aleph*, is the first letter of the Hebrew alphabet), and he called sets of cardinality $\aleph_0$ *countable* (*abzählbar*). The next immediate cardinal number $\aleph_1 = |\Omega|$ corresponds to the set Ω of all ordinal numbers having at most countably many predecessors. Cantor called the cardinal number of the set $\mathbb{R}$ of all real numbers the *cardinality of the continuum* and denoted it by $\mathfrak{c} = |\mathbb{R}|$. He was able to show that $\mathfrak{c} = 2^{\aleph_0}$ where $2^{\aleph_0} = |\mathcal{P}(\mathbb{N})|$ is the cardinality of the power set of $\mathbb{N}$, as well as the general inequality $|X| < |\mathcal{P}(X)| = 2^{|X|}$ for any set X. For a long time, Cantor attempted to prove the equality $\mathfrak{c} = \aleph_1$ but without success. This

statement became known as the *continuum hypothesis* and was placed first on Hilbert's famous list of 23 problems, which he presented to the mathematical community at the Second International Congress of Mathematicians in Paris in 1900. We have already briefly mentioned how this problem was eventually resolved in the concluding section dedicated to non-Euclidean geometry.

However, Cantor succeeded in proving several surprising results. Any interval has the same cardinality as the entire set $\mathbb{R}$. The real line $\mathbb{R}$ and the segment (i.e., interval) $[0,1]$ have the same cardinality as the plane $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ or the square $[0,1]^2 = [0,1] \times [0,1]$. From this, it follows that the continuum of any finite dimension has the same cardinality (meaning it contains the same number of points) as the one--dimensional continuum. This — contrary to Cantor's original expectations — implies that the cardinal number is too coarse an invariant to capture the dimension of a point set. Furthermore, the set $\mathbb{Q}$ of all rational numbers and the set $\mathbb{A}$ of all (real) algebraic numbers[3] are countable, meaning they have cardinality $\aleph_0$. On the other hand, the set $\mathbb{R}$ of all real numbers has cardinality $\mathfrak{c} > \aleph_0$. Thus, Cantor provided a purely existential proof of the existence of *transcendental* (i.e., non-algebraic) real numbers without explicitly constructing a single one.

The characterization of the cardinalities of infinite sets using their cardinal numbers and the structure of so-called *well-ordered sets*[4] using their ordinal numbers, along with the ability to extend arithmetic operations on finite natural numbers to these new numbers, changed Cantor's perspective on the different kinds of infinity that they represented. On one hand, this helped him overcome the reluctance to interpret this infinity as actual. In a sense, this led to the devaluation of this type of infinity, reflected in new terms such as *transfinite sets* or *transfinite numbers*, which suggested that they formed just an intermediate stage between the finite and the "true" or "genuine" infinity represented, for example, by the totality of all sets or the totality of all cardinal or ordinal numbers. On the other hand, looking into the depths of these newly opened domains evoked a sense of vertigo, formally manifested in the impossibility of their actualization: if we assume that there exists a set of all sets or a set of all cardinal or ordinal numbers, we arrive at a contradiction. Cantor therefore referred to these collections as *inconsistent sets* and firmly believed that the apparent paradoxes could be resolved through deeper study of the principles of set formation. An exception was Russell's paradox, published in 1903, which applied equally to Frege's *Grundgesätze* and Cantor's set theory. Due to the simplicity of its formulation, this paradox significantly shook Cantor's confidence in his own theory.

Cantor was also well aware that his interpretation of infinity as actual challenged the paradigm that had predominantly prevailed in both mathematics and philosophy since ancient times. He found support in Christian theology, which holds that infinity exists actually, being a divine attribute in terms of God's omnipotence, omniscience, love and mercy. Yet, he was soon confronted with the objection that his theory undermined God's exclusive and unique claim to possess these attributes. Cantor, however, was convinced that his set theory was not heretical; rather, he believed it had been revealed to him from heaven as a new and profound truth, along with

[3] A real or complex number c is called *algebraic* if there exists a polynomial $f(x) = x^n + a_1 x^{n-1} + \ldots + a_{n-1}x + a_n$ with rational coefficients $a_1, \ldots, a_n \in \mathbb{Q}$ such that $f(c) = 0$.

[4] An ordered set $(X; <)$ is called *well-ordered* if every its nonempty subset has the least element.

the mission to spread it further. In this context, the paradoxes he had discovered played in his favor, as they delineated the "realm of the transfinite" — accessible to human intellect — while separating it from the realm of the true, that is, absolute and non-actualizable infinity, which eludes human reason and belongs solely to God's competence. To ensure that the actual infinity present in his infinite sets and infinite cardinal and ordinal numbers was not in conflict with Church doctrine, Cantor turned to two prominent Roman Catholic scholars of his time: the priest, theologian, and philosopher Konstantin Gutberlet, (1837 – 1928) and the Jesuit theologian and expert in Church dogmatics, Cardinal Johan Baptist Franzelin (1816 – 1886), with whom he exchanged several letters. Both theologians responded to Cantor's ideas with kindness and understanding. After a relatively brief discussion, they unanimously concluded that his set theory did not contradict Church teachings and acknowledged that the actualization of infinity in infinite sets and transfinite cardinal and ordinal numbers did not infringe upon God's majesty. On the contrary, they recognized that Cantor's discoveries contributed to the greater glory of God.

In 1908, the German mathematician Ernst Zermelo (1871 – 1953) published an axiomatic system of set theory in which he succeeded in preserving all the essential properties of Cantor's theory while simultaneously avoiding all known set-theoretic paradoxes. He achieved this by imposing certain syntactic restrictions on the properties to which the comprehension principle could be applied, as well as by requiring that new sets (with three exceptions) be delimited only as subsets of previously given sets. However, as he noted himself, he was unable to prove the consistency of his system (i.e., to demonstrate that no contradictions could arise within it). Zermelo's system was later extended by Abraham Fraenkel (1891 – 1965) and Zermelo himself. The Zermelo-Fraenkel axiomatic system with the axiom of choice, abbreviated as ZFC, remains the most widely used axiomatization of set theory to this day.

A different axiomatization of set theory, closer in spirit to logicism in the sense of Frege and Russell, was proposed in 1937 by the American philosopher and logician Willard Van Orman Quine (1908 – 2000). His system, called *New Foundations*, abbreviated NF, allows for a universal set (i.e., the set of all sets) while avoiding paradoxes by applying Cantor's comprehension principle only to properties whose syntactic form inherently excludes self-reference. However, New Foundations has several peculiar characteristics and has not gained widespread acceptance as foundations of mathematics. Even the question of the relative consistency of NF with respect to ZFC remains an open problem to this day.

In 1914, on the eve of World War I, Felix Hausdorff (1868 – 1942) published *Principles of Set Theory* (*Grundzüge der Mengenlehre*). After the war, this book became the most widely used textbook and reference monograph on set theory for a long time. A second, revised, and condensed edition was published in 1927 under the simpler title *Set Theory* (*Mengenlehre*). In addition to a systematic exposition of set theory itself, the book covers topics such as ordered sets, topology (i.e., point-set theory), the set-theoretic approach to measure theory and probability, as well as to mathematical analysis and function theory. In this work, Hausdorff introduced an approach that gradually gained dominance throughout 20^{th}-century mathematics: studying various mathematical theories through their set-theoretic models. Set theory not only

provides "sufficient space" to represent possible structures of different theories within the set-theoretic universe, but it also offers these theories a universal language, enabling their study in a unified manner while simultaneously revealing relationships between them. As a result, set theory became the guarantee of the long-sought for *internal unity of all mathematics*. According to Hilbert, it is a "paradise created for us by Cantor, from which no one can expel us". This made the task of proving the consistency of set theory all the more pressing — ideally, the consistency of its axiomatic system ZFC, but for the beginning it would be enough to prove the consistency of some its fragment sufficient for constructing arithmetic of natural and real numbers and of the foundations of mathematical analysis.

This is easier said than done. Namely, it is not clear at all how such a proof should proceed. Typically, the consistency of one theory is proved assuming the consistency of another. For example, assuming the consistency of Euclidean geometry, we can prove the consistency of hyperbolic geometry (and vice versa). However, for a theory meant to serve as the foundations of mathematics, a relative proof of consistency, relying on the consistency of some supposedly even "more fundamental" theory, is insufficient. What is needed is some "absolute" proof. Pursuing this goal, Hilbert, in the 1920s, proposed and developed a program of *formalization of mathematics* that allowed mathematical proofs to be represented as finite sequences of symbols with clearly recognizable syntactic structure. To prove the consistency of such a formalized theory, one would simply need to verify that, among the of proofs, there cannot occur a pair of certain precisely described form, responsible for a contradiction. However, as Kurt Gödel demonstrated in 1930, the goals of Hilbert's program are unattainable.

2.8 Farewell to History

It remains to be said that during the 1920s and early 1930s, the central role of *first--order logic* was recognized as the most important "layer" of mathematical logic — at least from the perspective of mathematics. This recognition was largely due to the work of the Norwegian mathematician and logician Thoralf Skolem (1881 – 1963) and Gödel's proof of the *completeness theorem* for first-order logic in 1929. This was further confirmed by subsequent developments in the works of Polish-American mathematician and logician Alfred Tarski and his students in pre-war Poland and later in the United States, the Soviet mathematician Anatoly Ivanovich Maľcev (1909 – 1967), the Israeli-American logician and applied mathematician Abraham Robinson (1918 – 1974), and many others who opened up a wide range of applications for logic in mathematics. Gradually, they established a new mathematical discipline, later called *model theory*. To some extent, model theory can be regarded as the "metamathematics of algebra", as it provides a general framework and unifying perspective on various algebraic disciplines, allowing for the discovery of their often hidden interconnections. However, as demonstrated by Robinson's *nonstandard analysis*, its applications extend far beyond algebra.

The rapid development of mathematical logic, which began in the 1920s and 1930s, accelerated even further after World War II and continued throughout the remainder of the 20^{th} century up to the present day. It is simply impossible to outline its full course within the limited scope of this textbook. Therefore, we will change our

approach and no longer follow the development of mathematical logic from a historical perspective. Instead, we will focus on presenting its core components: propositional calculus, first-order logic, and a brief introduction to model theory. Between the second and third of these topics, we will include a chapter on the celebrated *Gödel's incompleteness theorems* and a short reflection on their philosophical implications.

3 Propositional Calculus

Propositional Calculus, also called *Sentential Calculus*, *Propositional Logic*, or *Zeroth--Order Logic*, examines the structure of certain fragment common both to natural languages as well as to languages of scientific theories, namely the abstract patterns or forms of propositions or statements occurring in them as well as the relations between them which can be inferred from their form. This will make possible a complete description of valid arguments in terms of structural relations between the forms of their constituents.

We will take advantage of the fact that our readers have already some acquaintance with Propositional Calculus, including the logical connectives and their meaning, truth tables, tautologies, etc. This allows us to focus on some general "philosophical" features of Propositional Calculus which are not always part of the usual courses and use this familiar and rather elementary material as a platform for the exposition of some topics and issues which will reappear later on within the Predicate Calculus in a more advanced form.

3.1 Propositions and Propositional Forms

A *propostion* or a *statement* is an affirmative grammatical sentence making a meaningful announcement which is either true or false, no matter whether we or whoever else are able to decide its verity. We say that the *truth value* of a proposition is 1 if it is true and it is 0 if it is false.

Given some propositions we can form new ones combining them by means of the unary logical connective *not* and the binary logical connectives *and*, *or*, *if ... then*, *if and only if*, *either ... or*, *neither ... nor*, etc. Propositional Calculus is based on the following fundamental observation:

> *If A is a proposition formed of some simpler propositions $p_1,\ldots,p_n$ by means of logical connectives in a certain way then the truth value of A can be determined just out of the truth values of the propositions $p_1,\ldots,p_n$ and the way how A is formed, regardless of the meaning and content of the propositions $p_1,\ldots,p_n$.*

In other words, the truth value of A can be *computed* from the truth values of its components $p_1,\ldots,p_n$ and the abstract pattern or the *form* of A.

As a consequence, the subject of Propositional Calculus is not primarily propositions themselves but the forms propositions can take on, according to the way how they are composed from simpler propositions by means of logical connectives. These abstract forms we call *propositional forms*; they are expressions (words) of some formal language to be introduced below. In order to describe the *syntax* of the language

of Propositional Calculus we will codify its symbols and describe the way how its words are generated.

The *language of Propositional Calculus* has the following symbols divided into three groups:

- Propositional variables: p, q, r, p_0, p_1, p_2, $\ldots$, q', q'', $\ldots$
- Logical connectives: $\neg$ (*not*), $\wedge$ (*and*), $\vee$ (*or*), $\Rightarrow$ (*if ... then* or *implies*), $\Leftrightarrow$ (*if and only if*) (two would suffice)
- Auxiliary symbols: (,) (*parentheses*) (they could be avoided)

We denote by P the set of all propositional variables. We assume that the set P is countably infinite at least in the potential sense, i.e., whenever we have any finite list of propositional variables $p_1, \ldots, p_n$, we are able to find some new propositional variable q not included in that list and, at the same time, all the propositional variables $p \in P$ can be set into a one-to-one correspondence with the natural numbers $n \in \mathbb{N}$.

Propositional forms are certain finite strings, i.e., words, consisting of the above quoted symbols. The set $\mathrm{VF}(P)$ of all propositional forms over the set of propositional variables P is defined recursively as the smallest set containing all the propositional variables and closed with respect to the application of logical connectives, i.e., the smallest set satisfying the following two conditions:

1° $P \subseteq \mathrm{VF}(P)$ (every propositional variable $p \in P$ is a propositional form over the set P)

2° If $A, B \in \mathrm{VF}(P)$ then $\neg A, (A \wedge B), (A \vee B), (A \Rightarrow B), (A \Leftrightarrow B) \in \mathrm{VF}(P)$ (if the strings A, B are propositional forms over the set P then so are the strings $\neg A$, $(A \wedge B)$, $(A \vee B)$, $(A \Rightarrow B)$ and $(A \Leftrightarrow B)$)

According to 1°, propositional variables are sometimes referred to as *atomic propositional forms*. As a consequence of the fact that the set P of all propositional variables is countable, the set $\mathrm{VF}(P)$ of all propositional forms over P is countable, as well.

The set $\mathrm{VF}(Q)$ of all propositional forms over any nonempty set of propositional variables $Q \subseteq P$ can be defined in an analogous way. In particular, for a finite set $Q = \{p_1, \ldots, p_n\}$, we denote

$$\mathrm{VF}(Q) = \mathrm{VF}(p_1, \ldots, p_n)$$

Since every propositional form $A \in \mathrm{VF}(P)$ is composed from atomic propositional forms by applying the rule 2° just finitely many times, there always is a finite number of propositional variables $p_1, \ldots, p_n \in P$ such that $A \in \mathrm{VF}(p_1, \ldots, p_n)$.

If A, B are propositional forms then the propositional form $\neg A$ is called the *negation* of A, and the propositional forms $(A \wedge B)$, $(A \vee B)$, $(A \Rightarrow B)$ and $(A \Leftrightarrow B)$ are called the *conjunction*, the *disjunction* or the *alternative*, the *implication* and the *equivalence* of A and B, respectively.

3.1.1 Remark. (a) According to the just stated definition not all finite strings of symbols of the language of Propositional Calculus are propositional forms. For instance, the expressions p, q, r, $\neg p$, $(p \wedge q)$, $(\neg p \Rightarrow r)$, $((p \wedge q) \vee (\neg p \Rightarrow r))$ can easily be recognized as propositional forms, while the expressions like $(p \neg q)$, $\neg pp \Rightarrow)(r \neg$ obviously fail to be propositional forms. Less obvious is the finding that neither the

expressions $p \wedge q$, $\neg p \Rightarrow r$, $(p \wedge q) \vee (\neg p \Rightarrow r)$ are propositional forms although we are inclined to recognize them to be. In order to reconcile the above definition with our intuition and the usual practice, we accept the convention of omitting the outermost parentheses (which clearly are superfluous) in any propositional form. Thus we consider the expressions like $p \wedge q$, $\neg p \Rightarrow r$, $(p \wedge q) \vee (\neg p \Rightarrow r)$ as denoting the propositional forms $(p \wedge q)$, $(\neg p \Rightarrow r)$, $((p \wedge q) \vee (\neg p \Rightarrow r))$, respectively.

(b) We could completely manage without the parentheses using the *Polish notation*. In that case point 2° of the above definition would be modified as follows:

2* If $A, B \in \mathrm{VF}(P)$ then $\neg A$, $\wedge AB$, $\vee AB$, $\Rightarrow AB$, $\Leftrightarrow AB \in \mathrm{VF}(P)$ (if the strings A, B are propositional forms over the set P then so are the strings $\neg A$, $\wedge AB$, $\vee AB$, $\Rightarrow AB$ and $\Leftrightarrow AB$)

For instance, in Polish notation the propositional form $(p \wedge q) \vee (\neg p \Rightarrow r)$ would be written as

$$\vee \wedge pq \Rightarrow \neg pr$$

However cumbersome and hardly legible this expression may appear to us, it should be realized that from the point of view of a computer assisted processing this aspect is of almost no importance.

(c) In spite of the names we have attached to the logical connectives pointing to their intended role, they should be regarded as mere graphical symbols deprived of any meaning for the moment. They will only acquire their usual meaning later on, when we develop the semantics of Propositional Calculus.

(d) It should be noted that the signs A, B, C used to denote arbitrary propositional forms, the sign P and the expression $\mathrm{VF}(P)$ denoting the set of all propositional variables and the set of all propositional forms, respectively, etc., do not belong to the language of Propositional Calculus — they are symbols or expressions of certain metalanguage we use in the study of Propositional Calculus.

Let us turn reader's attention to the point that $\mathrm{VF}(P)$ is the *smallest* set satisfying conditions 1° and 2°. This inconspicuous requirement endows us with a powerful tool for proving facts about propositional forms, namely with the proof method by *induction on complexity*: In order to establish that all propositional forms have certain property it is enough to show that the set of all propositional forms having this property satisfies the above conditions 1° and 2°.

3.1.2 Theorem. *Let $M \subseteq \mathrm{VF}(P)$ be any set of propositional forms satisfying the following two conditions:*

1° $P \subseteq M$
(every propositional variable $p \in P$ belongs to the set M)

2° *If $A, B \in M$ then $\neg A$, $A \wedge B$, $A \vee B$, $A \Rightarrow B$, $A \Leftrightarrow B \in M$ (M is closed with respect to the formation of propositional forms by means of logical connectives)*

Then $M = \mathrm{VF}(P)$, i.e., every propositional form over P belongs to M.

The reader should compare the induction on complexity with the usual method of induction, used in proving that certain property holds for all natural numbers: Since the set $\mathbb{N}$ of all natural numbers is the smallest set containing 0 and closed with

respect to the successor operation $n \mapsto n+1$, in order to show that certain set $M \subseteq \mathbb{N}$ contains all natural numbers, i.e., $M = \mathbb{N}$, it is enough to show that $0 \in M$ and, for every $n \in M$ also $n+1 \in M$. In the induction on complexity the role of the number $0 \in \mathbb{N}$ is played by the propositional variables $p \in P$, and the role of the successor operation is played by the logical connectives. Already in this moment it could be anticipated that for the sake of induction proofs it would be desirable to minimize the number of logical connectives for which the condition 2° has to be verified. We will return to this point in the next paragraph.

3.2 Interpretations, Truth Tables and Logical Equivalence

Next to syntax we will develop the *semantics* of Propositional Calculus. Let us recall that in Logic we take no account of the content of propositions, and the propositional forms are indeed deprived of any content. Nevertheless, we can still examine the situations under which they become true or false. These situations will be called interpretations or truth evaluation and they will represent the way of assigning however limited but still certain meaning to propositional forms.

We start by introducing the boolean algebraic operations on the two element set $\{0,1\}$ of the truth values 0 (*false*) and 1 (*true*), corresponding to the logical connectives and denoted by the same symbols. They are given by the following tables:

$\neg$	0	1
	1	0

$\wedge$	0	1
0	0	0
1	0	1

$\vee$	0	1
0	0	1
1	1	1

$\Rightarrow$	0	1
0	1	1
1	0	1

$\Leftrightarrow$	0	1
0	1	0
1	0	1

In Propositional Calculus an *interpretation* or a *truth evaluation* is any mapping $I\colon P \to \{0,1\}$, i.e., any assignment of truth values 0 or 1 to the propositional variables. Intuitively, such an interpretation represents a possible situation described in terms of the assignment of truth values to the propositional variables.

Every interpretation $I\colon P \to \{0,1\}$ will be extended to a mapping $I\colon \mathrm{VF}(P) \to \{0,1\}$, denoted by the same symbol and still called an interpretation or a truth evaluation, by means of the following recursive definition:

$$I(\neg A) = \neg I(A) \qquad I(A \wedge B) = I(A) \wedge I(B) \qquad I(A \Rightarrow B) = I(A) \Rightarrow I(B)$$
$$I(A \vee B) = I(A) \vee I(B) \qquad I(A \Leftrightarrow B) = I(A) \Leftrightarrow I(B)$$

for any $A, B \in \mathrm{VF}(P)$, assuming that the values $I(A)$ and $I(B)$ have already been defined. Instead of $I(A) = 1$ we say that A is true or satisfied in the interpretation I; $I(A) = 0$ means that A is false in the interpretation I.

The reader should realize the following two facts:

- In each of the above equalities the signs $\neg$, $\wedge$, $\vee$, $\Rightarrow$, $\Leftrightarrow$ denote the logical connectives on the left side while on the right side they denote the corresponding boolean operations on the set $\{0,1\}$.

- The equality symbol $=$ and the signs I, J, denoting arbitrary truth evaluations, belong to our metalanguage and not to the language of Propositional Calculus itself.

Just from this moment on, and by the virtue of the tables of the operations $\neg$, $\wedge$, $\vee$, $\Rightarrow$ and $\Leftrightarrow$ on the set $\{0,1\}$ of the truth values, the corresponding logical connectives can rightfully bear their names of *negation*, *conjunction*, *alternative* or *disjunction* (in nonexclusive sense), *implication* and *equivalence*, respectively.

It can be easily realized that the above recursive definition is redundant in some sense. It would be enough to describe the extension of the mapping $I\colon P \to \{0,1\}$ according to the negation $\neg$ and one (and anyone) of the binary connectives $\wedge$, $\vee$, $\Rightarrow$; then the remaining equalities would be satisfied, as well. In other words, a mapping $I\colon \mathrm{VF}(P) \to \{0,1\}$ is (an extension of) an interpretation if and only if it satisfies the equality $I(\neg A) = \neg I(A)$, and one (and anyone) of the equalities $I(A \wedge B) = I(A) \wedge I(B)$, $I(A \vee B) = I(A) \vee I(B)$, $I(A \Rightarrow B) = I(A) \Rightarrow I(B)$ for all $A, B \in \mathrm{VF}(P)$. Then it automatically satisfies the remaining equalities, as well. As a consequence, the notion of an interpretation (truth evaluation) could be defined in a more elegant way, as a mapping $I\colon \mathrm{VF}(P) \to \{0,1\}$ preserving the operations of the algebras $(\mathrm{VF}(P); \wedge, \vee, \neg)$, $(\{0,1\}; \wedge, \vee, \neg)$, i.e., as a *homomorphism*

$$I\colon (\mathrm{VF}(P); \wedge, \vee, \neg) \to (\{0,1\}; \wedge, \vee, \neg)$$

Let us make the just discussed point more precise. We call two propositional forms $A, B \in \mathrm{VF}(P)$ *logically equivalent* if $I(A) = I(B)$ for every interpretation $I\colon P \to \{0,1\}$; in that case we write $A \equiv B$. (It should be realized that the sign $\equiv$, similarly as the signs A, B, C, P, I, J or the expression $\mathrm{VF}(P)$, etc., does not belong to the symbols of the language of Propositional Calculus — it is a symbol of our metalanguage, again.) The reader is asked to verify that the relation of logical equivalence $\equiv$ is reflexive, symmetric and transitive, hence it is indeed an equivalence relation on the set $\mathrm{VF}(P)$.

It is known that any of the pairs $(\neg, \wedge)$, $(\neg, \vee)$, $(\neg, \Rightarrow)$ forms a *complete list of logical connectives*, i.e., any propositional form $A \in \mathrm{VF}(P)$ is logically equivalent to some propositional form A' containing the same propositional variables as A and involving just the logical connectives from one (and anyone) of the three pairs above.

Choosing the connectives $\neg$, $\wedge$ as the primitive ones, the remaining connectives could be introduced as abbreviations for the propositional forms on the right:

$$\begin{aligned} A \vee B &\equiv \neg(\neg A \wedge \neg B) \\ A \Rightarrow B &\equiv \neg(A \wedge \neg B) \\ A \Leftrightarrow B &\equiv \neg(A \wedge \neg B) \wedge \neg(\neg A \wedge B) \end{aligned}$$

Choosing $\neg$ and $\vee$ as primitive connectives we would have

$$\begin{aligned} A \wedge B &\equiv \neg(\neg A \vee \neg B) \\ A \Rightarrow B &\equiv \neg A \vee B \\ A \Leftrightarrow B &\equiv \neg(\neg(A \vee \neg B) \vee \neg(\neg A \vee B)) \end{aligned}$$

Finally, if our primitive connectives were $\neg$ and $\Rightarrow$, we would have

$$A \wedge B \equiv \neg(A \Rightarrow \neg B)$$
$$A \vee B \equiv \neg A \Rightarrow B$$
$$A \Leftrightarrow B \equiv (A \Rightarrow \neg B) \Rightarrow \neg(\neg A \Rightarrow B)$$

It follows that we could have used just the unary connective $\neg$ and just one (and anyone) from among the three binary connectives $\wedge$, $\vee$, $\Rightarrow$ in the recursive definition of the notion of propositional form in the previous paragraph; the forth binary connective $\Leftrightarrow$ becomes superfluous in any case.

Additionally, we will make use of the logical equivalences of associativity of the connectives $\wedge$ and $\vee$

$$(A \wedge B) \wedge C \equiv A \wedge (B \wedge C) \qquad \text{and} \qquad (A \vee B) \vee C \equiv A \vee (B \vee C)$$

for any propositional forms $A, B, C \in \mathrm{VF}(P)$. This allows us to omit the superfluous parenthesis in conjunctions and alternatives of any finite number of propositional forms and write simply $A \wedge B \wedge C$, $A \vee B \vee C$, $A_1 \wedge \ldots \wedge A_m$, $B_1 \vee \ldots \vee B_n$, etc.

It is worthwhile to notice that we could manage with a single binary connective, namely the *Sheffer stroke* $|$ (the NAND operator) which can be expressed by means of $\neg$ and $\wedge$, or by means of $\neg$ and $\vee$ as follows:

$$A|B \ \equiv \neg(A \wedge B) \equiv \neg A \vee \neg B$$

Conversely, the standard logical connectives $\neg$, $\wedge$ and $\vee$ can be expressed in terms of the Sheffer stroke as follows:

$$\neg A \equiv A|A$$
$$A \wedge B \equiv (A|B)|(A|B)$$
$$A \vee B \equiv (A|A)|(B|B)$$

The task to find the corresponding expressions for $A \Rightarrow B$ and $A \Leftrightarrow B$ is left to the reader.

Another single logical connective capable to generate all the remaining ones is the NOR operator $\dagger$, also known as the *Peirce arrow* or *Quine dagger*, which is dual to the Sheffer stroke. In terms of $\neg$ and $\wedge$, or $\neg$ and $\vee$, respectively, it can be expressed as follows:

$$A \dagger B \equiv \neg A \wedge \neg B \equiv \neg(A \vee B)$$

The reader is asked to express the usual logical connectives $\neg$, $\wedge$, $\vee$, $\Rightarrow$ and $\Leftrightarrow$ in terms of the Quine dagger $\dagger$, and, at the same time to find the expressions for the Sheffer stroke in terms of the Quine dagger and vice versa.

3.3 Tautologies and Other Classes of Propositional Forms

Using the concept of interpretation we can single out several important classes of propositional forms. A propositional form $A \in \mathrm{VF}(P)$ is called

- a *tautology* if $I(A) = 1$ for every interpretation $I\colon P \to \{0,1\}$
- a *contradiction* if $I(A) = 0$ for every interpretation $I\colon P \to \{0,1\}$
- *satisfiable* if $I(A) = 1$ for at least one interpretation $I\colon P \to \{0,1\}$
- *refutable* if $I(A) = 0$ for at least one interpretation $I\colon P \to \{0,1\}$

There is a twofold duality between the four notions just introduced:
the *inner duality*

- A is a tautology if and only if $\neg A$ is a contradiction
- A is satisfiable if and only if $\neg A$ is refutable

and the *outer duality*

- A is a tautology if and only if A is not refutable
- A is a contradiction if and only if A is not satisfiable

It can be easily seen that, for any propositional forms A, B, we have $A \equiv B$ if and only if the propositional form $A \Leftrightarrow B$ is a tautology.

The question whether a given propositional form A belongs to any of the four classes defined above can be decided algorithmically using the method of *truth value tables*, evaluating the truth values $I(A)$ for all the interpretations $I\colon P \to \{0,1\}$. In view of the fact that, for an infinite set P, there are infinitely many such interpretations, it is important that to that end it is enough to deal just with finitely many of them.

3.3.1 Theorem. *Let $A \in \mathrm{VF}(P)$ be any propositional form such that all the propositional variables occurring in A are included in the list $p_1, \ldots, p_n$. Then $I(A) = J(A)$ for any truth evaluations $I, J\colon P \to \{0,1\}$ such that $I(p_k) = J(p_k)$ for each $k = 1, \ldots, n$.*

In other words, the value $I(A)$ of a truth evaluation I on a propositional form A depends on the values of I on the finite set of propositional variables occurring in A, only. However obvious and intuitively clear this fact may appear, we nonetheless prove it, mainly in order to illustrate the proof method by induction on complexity.

Let us remark that the term *proof* will be, starting from the next paragraph, used exclusively for formal proofs within the Propositional Calculus. They will form one constituent of the subject of our study. For the sake of distinction, proofs of results about Propositional Calculus (and later on about Predicate Calculus), led in our metalanguage, will be referred to as *demonstrations*.

Demonstration. Denoting $Q = \{p_1, \ldots, p_n\}$ and

$$M = \{A \in \mathrm{VF}(Q)\colon I(A) = J(A)\}$$

we are to show that $M = \mathrm{VF}(Q)$. Since I and J coincide on the set Q, we have $Q \subseteq M$, which is the initial induction step 1°. In order to verify the induction step 2°, assume that $A, B \in M$, i.e., $A, B \in \mathrm{VF}(Q)$, and $I(A) = J(A)$ as well as $I(B) = J(B)$. Then, as both I, J preserve the logical connectives,

$$I(\neg A) = \neg I(A) = \neg J(A) = J(\neg A)$$
$$I(A \wedge B) = I(A) \wedge I(B) = J(A) \wedge J(B) = J(A \wedge B)$$

hence both $\neg A, A \wedge B \in M$. Similarly, we could show that $A \vee B, A \Rightarrow B, A \Leftrightarrow B \in M$, too. However, in view of our previous accounts, it is clear that the induction step 2° for the connectives $\vee$, $\Rightarrow$ and $\Leftrightarrow$ is not necessary to perform.

3.3.2 Example. Using the truth value table method, it can be easily shown that the following propositional form

$$(p \Rightarrow (q \Rightarrow r)) \Leftrightarrow ((p \wedge q) \Rightarrow r)$$

is a tautology. Denoting by L the propositional form $p \Rightarrow (q \Rightarrow r)$ and by R the propositional form $(p \wedge q) \Rightarrow r$, we have

p	q	r	$q \Rightarrow r$	L	$p \wedge q$	R	$L \Leftrightarrow R$
1	1	1	1	1	1	1	1
1	1	0	0	0	1	0	1
1	0	1	1	1	0	1	1
0	1	1	1	1	0	1	1
1	0	0	1	1	0	1	1
0	1	0	0	1	0	1	1
0	0	1	1	1	0	1	1
0	0	0	1	1	0	1	1

As a consequence,

$$A \Rightarrow (B \Rightarrow C) \equiv (A \wedge B) \Rightarrow C$$

for any propositional forms A, B, C.

More important and interesting than filling in mechanically the above truth table it is to realize what kind of logical law, called the *Law of Exportation*, is expressed by this tautology or by the above logical equivalence. The left hand form $A \Rightarrow (B \Rightarrow C)$ states that "if A, then B implies C". The right hand form $(A \wedge B) \Rightarrow C$ states that "A and B jointly imply C". These two forms of statements are always equivalent: going from the left to the right it is possible to join the two assumptions A, B to a single assumption $A \wedge B$; going from the right to the left it is possible to divide the assumption $A \wedge B$ into its constituents A and B and apply them consecutively one after the other.

3.3.3 Exercise. (Normal forms) A propositional form is called an *elementary conjunction* if it has the shape $B_1 \wedge \ldots \wedge B_m$ where each of the forms B_i is either a propositional variable or a negation of some propositional variable. A propositional form is called a *disjunctive normal form* if it has the shape $C_1 \vee \ldots \vee C_k$ where each of the forms C_j is an elementary conjunction. Show that every propositional form $A \in \mathrm{VF}(p_1, \ldots, p_n)$ is logically equivalent to some disjunctive normal form $A' \in \mathrm{VF}(p_1, \ldots, p_n)$. To this end design an algorithmic method how to obtain the disjunctive normal form $A' \equiv A$ from the truth table of the form A.

Similarly, define the dual notions of an *elementary disjunction* and of a *conjunctive normal form* and show that every propositional form is logically equivalent to some conjunctive normal form.

3.3.4 Exercise. (Boolean functions) Let $A \in \mathrm{VF}(p_1, \ldots, p_n)$ be a propositional form in propositional variables $p_1, \ldots, p_n$. Then A can be turned into a *boolean function* $F_A \colon \{0,1\}^n \to \{0,1\}$, i.e., into an n-ary operation on the two-element set $\{0,1\}$ given by $F_A(e_1, \ldots, e_n) = I(A)$ for any $e_1, \ldots, e_n \in \{0,1\}$, where $I \colon \{p_1, \ldots, p_n\} \to \{0,1\}$ is the interpretation on the set $\{p_1, \ldots, p_n\}$ such that $I(p_k) = e_k$ for $k = 1, \ldots, n$.

(a) Show that for any propositional forms $A, B \in \mathrm{VF}(p_1, \ldots, p_n)$ we have $F_A = F_B$ if and only if $A \equiv B$.

(b) Prove that there are exactly 2^{2^n} boolean functions $\{0,1\}^n \to \{0,1\}$ for each $n \geq 1$. What about $n = 0$?

(c) Show that for every boolean function $F \colon \{0,1\}^n \to \{0,1\}$ there infinitely many propositional forms $A \in \mathrm{VF}(p_1, \ldots, p_n)$ such that $F = F_A$.

3.3.5 Exercise. Let $A \in \mathrm{VF}(p_1, \ldots, p_n)$ be a propositional form in propositional variables $p_1, \ldots, p_n$ and $B_1, \ldots, B_n \in \mathrm{VF}(P)$ be arbitrary propositional forms. We denote by $A(B_1, \ldots, B_n)$ the propositional form obtained by substituting the forms $B_1, \ldots, B_n$ into the form A in places of the variables $p_1, \ldots, p_n$, respectively. For instance, if A is the form $(p \wedge \neg q) \Rightarrow (q \vee r)$ in propositional variables p, q, r and B, C, D are the propositional forms $r \vee s$, $p \Rightarrow \neg r$, q, respectively, then $A(B, C, D)$ denotes the form

$$((r \vee s) \wedge \neg(p \Rightarrow \neg r)) \Rightarrow ((p \Rightarrow \neg r) \vee q)$$

(a) Demonstrate that if A is a tautology (contradiction) then $A(B_1, \ldots, B_n)$ is also a tautology (contradiction) for any $B_1, \ldots, B_n$.

(b) Give examples of a satisfiable (refutable) form A and of forms $B_1, \ldots, B_n$ such that $A(B_1, \ldots, B_n)$ is not satisfiable (refutable).

3.4 Theories in Propositional Calculus

In common language the word *theory* usually refers to some interconnected system of knowledge, consisting of statements about certain topic and including also a methodology of obtaining and verifying or refuting these statements. The statements or propositions forming the "body of knowledge" of the theory could have been obtained in various ways: some of them may express certain empirical facts established by observation or experiments, some of them may be a part of common beliefs, tradition or cultural heritage, some of them may be mere hypotheses to be verified or refuted in the future, and, finally, some of them may be derived from any of the previously mentioned ones as their logical consequences.

Following the leading intention of logic, we will ignore the content, methodology and the overall character of a theory, we will neither distinguish which of its postulates are true or false, which are firmly established and which are mere hypotheses, nor take care of the way how all that happened. We will bring to the focus just a single aspect of all such theories, namely the structure of logical inference, i.e., the way new statements necessarily follow or can be derived from those made into the departing postulates or axioms of the particular theory.

Accordingly, a *propositional theory* or simply a *theory* is any set $T \subseteq \mathrm{VF}(P)$ of propositional forms; its elements $A \in T$ are called the *specific axioms* or just the *axioms* of T. We warn the readers not to take this definition word for word, not even within the framework of Propositional Calculus, let alone when speaking about a broader perspective. It should rather be understood as stating that, within Propositional Calculus, a theory *is given* or *uniquely determined* by the set of its specific axioms. Propositional Calculus will take care of the rest, i.e., of the structure of logical inference, which is the same for all the theories.

An interpretation $I\colon \mathrm{VF}(P) \to \{0,1\}$ is called an *interpretation of the theory* T if $I(A) = 1$ for each $A \in T$, i.e., if all the axioms of T are true in the interpretation I. Intuitively, an interpretation of the theory T represents a situation in which all the axioms of the theory T, hence T itself, are satisfied.

A propositional form B is a *logical consequence* of the axioms of a theory T or just a *logical consequence* of T if $I(B) = 1$ for every interpretation I of the theory T. Alternatively we say that B is *true valid* or *satisfied* in T, or that T *entails* B. In symbols we write $T \vDash B$. Intuitively, $T \vDash B$ means that, in every possible situation in which all the axioms of the theory T are satisfied, B is satisfied as well.

Instead of $\emptyset \vDash B$ we write just $\vDash B$; it means that B is true under every interpretation $I\colon P \to \{0,1\}$, in other words, B is a tautology.

As it follows from the theorem below, the question whether $T \vDash B$ can be algorithmically decided using truth value tables, for any theory T with just finitely many specific axioms and each propositional form $B \in \mathrm{VF}(P)$.

3.4.1 Theorem. *Let $T = \{A_1, \ldots, A_n\}$ be a theory with finitely many specific axioms and $B \in \mathrm{VF}(P)$. Then $T \vDash B$ if and only if the propositional form $(A_1 \wedge \ldots \wedge A_n) \Rightarrow B$ is a tautology.*

Demonstration. Assume that $T \vDash B$. Let $I\colon \mathrm{VF}(P) \to \{0,1\}$ be any interpretation. Then either $I(A_k) = 0$ for at least one $k = 1, \ldots, n$, or $I(A_k) = 1$ for each $k = 1, \ldots, n$. In the first case $I(A_1 \wedge \ldots \wedge A_n) = 0$, therefore,

$$I((A_1 \wedge \ldots \wedge A_n) \Rightarrow B) = 1$$

In the second case I is an interpretation of the theory T, hence $I(B) = 1$ since $T \vDash B$. Then

$$I((A_1 \wedge \ldots \wedge A_n) \Rightarrow B) = 1$$

again. Thus $(A_1 \wedge \ldots \wedge A_n) \Rightarrow B$ is indeed a tautology.

Conversely, assume that $(A_1 \wedge \ldots \wedge A_n) \Rightarrow B$ is a tautology, i.e., it is true in every interpretation I. If I is an interpretation of T, then $I(A_1 \wedge \ldots \wedge A_n) = 1$. Thus

$$I((A_1 \wedge \ldots \wedge A_n) \Rightarrow B) = 1$$

can happen only if $I(B) = 1$, too. It follows that $T \vDash B$.

In general, however, T may have infinitely many specific axioms. Even in that case, in order to show that B is not a logical consequence of T, i.e., $T \nvDash B$, it is enough to find a single interpretation I of T such that $I(B) = 0$. Is it the case, then we say that (the validity of) B in T was *refuted by the counterexample* I. However, in order to confirm that $T \vDash B$, the definition requires of us to determine the truth value $I(B)$

for all interpretations of T, and there are possibly infinitely many of them. Thus we are seemingly facing an unrealizable task.

In mathematics, however, the usual way how to establish the validity of some statement within some theory is by *proving it from the axioms* of the theory and not by examining all the possible situations in which all these axioms are true and checking the validity of the statement in each of those situations. Also in Propositional Calculus we will develop the syntactic concepts of proof and provability with the aim to get in grasp with the semantic concept of validity or truth by means of them.

3.5 Axiomatization of Propositional Calculus

In order to have a brief and concise axiomatization of Propositional Calculus we will proceed as if the set $\mathrm{VF}(P)$ of all propositional forms were built of the propositional variables by means of the logical connectives $\neg$ and $\Rightarrow$, only. Thus the remaining logical connectives are considered as certain abbreviations displayed in the previous paragraph. An alternative axiomatization using the logical connectives $\neg$, $\wedge$, $\vee$ and $\Rightarrow$ can be found in the final Section 3.9.

3.5.1 Logical axioms. (4 axiom schemes)
For any propositional forms A, B, C, the following propositional forms are *logical axioms*:

(LAx 1) $A \Rightarrow (B \Rightarrow A)$

(LAx 2) $(A \Rightarrow (B \Rightarrow C)) \Rightarrow ((A \Rightarrow B) \Rightarrow (A \Rightarrow C))$

(LAx 3) $(A \Rightarrow B) \Rightarrow ((A \Rightarrow \neg B) \Rightarrow \neg A)$

(LAx 4) $\neg\neg A \Rightarrow A$

Additionally, we have a single deduction rule or rule of inference:

3.5.2 Deduction rule: Modus Ponens

(MP) $\dfrac{A,\ A \Rightarrow B}{B}$ (from A and $A \Rightarrow B$ infer B)

3.5.3 Exercise. (a) Show that all the logical axioms are tautologies and explain their intuitive meaning.
(b) Show that the deduction rule Modus Ponens is correct in the following sense:
If $I\colon \mathrm{VF}(P) \to \{0,1\}$ is any interpretation and $A, B \in \mathrm{VF}(P)$ are propositional forms such that $I(A) = I(A \Rightarrow B) = 1$, then $I(B) = 1$, as well.

A *proof* in the theory $T \subseteq \mathrm{VF}(P)$ is a finite sequence $A_0, A_1, \ldots, A_n$ of propositional forms such that every item A_k is either a logical axiom, or a specific axiom of the theory T (i.e., $A_k \in T$), or it follows from the previous items by the rule Modus Ponens (i.e., there are $i, j < k$ such that A_j has the form $A_i \Rightarrow A_k$).

A propositional form B is *provable* in a theory T if there is a proof $A_0, A_1, \ldots, A_n$ in T such that its last item A_n coincides with B. In symbols, $T \vdash B$. Instead of $\emptyset \vdash B$ we write just $\vdash B$; it means that B is provable from the logical axioms, only.

3.5.4 Remark. The above axiomatization of Propositional Calculus is by far not the only possible one. As already mentioned, an alternative axiomatization can be found in Section 3.9. Both these axiomatizations contain infinitely many axioms (listed in form of finitely many axiom schemes) and a single rule of inference. Such axiomatizations are referred to as *Hilbert style axiomatizations* featured by "many" logical axioms and "few" deduction rules. On the other hand, *Gentzen style axiomatizations* contain "many" rules of inference and just "few" logical axioms (or even none, replacing a logical axiom A by the deduction rule $\frac{}{A}$ with meaning *derive A out of nothing*). In general, Hilbert style axiomatizations are better suited for the description, study and analysis of the formal logical system itself, while Gentzen style axiomatizations are more effective in applications like logical programming or automatic theorem proving. However, as far as they serve as axiomatizations of the classical Propositional Calculus, they are all equivalent in the sense that they produce the same family of provable forms.

3.5.5 Exercise. Show that, for any propositional forms A, B, the following propositional forms are tautologies, and that they all are provable just from the logical axioms:

(a) $A \Rightarrow A$
(b) $A \Rightarrow \neg\neg A$
(c) $\neg A \Rightarrow (A \Rightarrow B)$
(d) $(\neg B \Rightarrow \neg A) \Rightarrow (A \Rightarrow B)$
(e) $(A \Rightarrow B) \Rightarrow (\neg B \Rightarrow \neg A)$
(f) $(A \Rightarrow (\neg B \Rightarrow \neg(A \Rightarrow B)))$
(g) $(A \Rightarrow B) \Rightarrow ((\neg A \Rightarrow B) \Rightarrow B)$
(h) $(\neg A \Rightarrow A) \Rightarrow A$

As an example (a rather deterring one) we show that for every propositional form A the form in (a) is provable from the logical axioms.

1. $(A \Rightarrow ((A \Rightarrow A) \Rightarrow A)) \Rightarrow ((A \Rightarrow (A \Rightarrow A)) \Rightarrow (A \Rightarrow A))$
   (LAx 2), taking A for both A and C and $A \Rightarrow A$ for B
2. $A \Rightarrow ((A \Rightarrow A) \Rightarrow A)$
   (LAx 1), taking A for A and $A \Rightarrow A$ for B
3. $(A \Rightarrow (A \Rightarrow A)) \Rightarrow (A \Rightarrow A)$
   follows from 1 and 2 by (MP)
4. $A \Rightarrow (A \Rightarrow A)$
   (LAx 1), taking A for both A and B
5. $A \Rightarrow A$
   follows from 3 and 4 by (MP)

3.5.6 Exercise. Sow that the axiom schemes (LAx 3) and (LAx 4) can be replaced by a single axiom scheme

(LAx 5) $(\neg A \Rightarrow B) \Rightarrow ((\neg A \Rightarrow \neg B) \Rightarrow A)$

To this end show that every instance of the scheme (LAx 5) is provable from some instances of the schemes (LAx 1), (LAx 2), (LAx 3), (LAx 4), and vice versa, all instances of the schemes (LAx 3), (LAx 4) are provable from some instances of the schemes (LAx 1), (LAx 2), (LAx 5).

3.6 The Soundness Theorem

Having introduced the axiomatization of Propositional Calculus we are facing the task to establish that it is *sound* or *correct* in the following sense: For every theory $T \subseteq \mathrm{VF}(P)$, all the propositional forms provable in T are satisfied in T. Otherwise it could happen that for some propositional form B provable in T it would be possible to find an interpretation I of T such that $I(B) = 0$. Such an I would represent a situation in which all the axioms of T were satisfied, nevertheless, B were false. Thus we could be able to prove, from the axioms of T, some conclusions contradicting these axioms, which would be a disaster witnessing a collapse of our axiomatization. Therefore it is of crucial importance that we have the following result:

3.6.1 Soundness Theorem. *Let $T \subseteq \mathrm{VF}(P)$ be a theory. Then, for every propositional form $B \in \mathrm{VF}(P)$, if $T \vdash B$ then $T \vDash B$.*

Demonstration. Let $T \vdash B$ and $A_0, A_1, \ldots, A_n$ be a proof of B in T. We will show that $I(A_k) = 1$ for every interpretation I of the theory T and *each* $k \leq n$. Then, of course, $I(B) = 1$, since B is A_n. Each A_k is either a logical axiom, in which case $I(A_k) = 1$ for every interpretation I, or a specific axiom of T, in which case $I(A_k) = 1$ as I is an interpretation of T, or A_k follows from some previous items A_i, A_j by (MP). Assuming that we already have proved that $I(A_i) = I(A_j) = 1$, we can conclude $I(A_k) = 1$, too, since, as we already have noted in Exercise 3.5.3 (b), the rule (MP) is correct.

3.6.2 Remark. Let us turn the reader's attention to the fact that — however simple and transparent the above argument might appear — it contains a kind of vicious circle. In the demonstration of the Soundness Theorem we have been using logical deduction and inference within the natural language extended by some fairly simple mathematical notation. Thus we have used in an informal way the same logical means the soundness of which we wanted to establish within the formalized Propositional Calculus. Strictly speaking, the formal counterpart of the informal logical means we have been using goes even beyond Propositional Calculus: since our arguments contained some quantification, they interfered already with the Predicate Calculus. It is important to realize that we are unable to prove the Soundness Theorem out of nothing, without assuming some minimal logical fragment of natural language as granted. Thus what we have achieved is nothing more and nothing less than the understanding and realization that our formal axiomatization of Propositional Calculus is in good accord with the logical structure of deduction and inference within the natural language.

Later on we will also establish the converse of the Soundness Theorem.

3.6.3 Completeness Theorem. *Let $T \subseteq \mathrm{VF}(P)$ be a theory. Then, for every propositional form $B \in \mathrm{VF}(P)$, if $T \vDash B$ then $T \vdash B$.*

3.6.4 Remark. It is illuminating to compare the status of the Completeness Theorem with that of the Soundness Theorem. As we have seen, the demonstration of the Soundness Theorem was fairly simple. On the other hand, as we shall see later on, the demonstration of the Completeness Theorem will be considerably more involved. While the failure of the Soundness Theorem would cause a collapse of our axiomatization of Propositional Calculus, the consequences of a possible failure of the Completeness Theorem would be, at least at a glance, less dramatic: It would just mean that our axiomatization of Propositional Calculus is not powerful enough and we should look for some additional logical axioms and/or deduction rules extending our original list in order to achieve its completeness. Then, however, we would have to face a more delicate question: Is it at all possible to achieve completeness in our axiomatization without destroying its soundness? Namely, the Soundness Theorem and the Completeness Theorem together answer this question affirmatively and guarantee that the relation between the syntax and semantics of Propositional Calculus is carefully balanced.

Later on, when dealing with an analogous issue for Predicate Calculus, we will quote an example of certain its fairly natural fragment not admitting any axiomatization satisfying both the Soundness and the Completeness Theorem.

3.7 The Deduction Theorem and Its Corollaries

On the way to the demonstration of the Completeness Theorem we are going to state a handful of results which are of independent interest in their own right. In their demonstrations we will use the notation $A \approx B$, expressing that the characters A and B denote the same propositional form. The symbol $\approx$ belongs to our metalanguage and not to the language of Propositional Calculus itself, similarly as the symbols A, B, P, VF, I, $\equiv$, etc.

3.7.1 Deduction Theorem. *Let $T \subseteq \mathrm{VF}(P)$ be a theory and $A, B \in \mathrm{VF}(P)$ be propositional forms. Then $T \vdash A \Rightarrow B$ if and only if $T \cup \{A\} \vdash B$.*

Demonstration. Let $T \vdash A \Rightarrow B$. Then the more $T \cup \{A\} \vdash A \Rightarrow B$. Obviously, $T \cup \{A\} \vdash A$, from which we get $T \cup \{A\} \vdash B$ by (MP). Namely, if $C_0, C_1, \ldots, C_n$ is a proof of $A \Rightarrow B$ in $T \cup \{A\}$, then $C_0, C_1, \ldots, C_n, A, B$ is a proof of B in $T \cup \{A\}$.

Conversely, let $T \cup \{A\} \vdash B$. First we take care of the following two trivial cases:

(a) B is a logical axiom or $B \in T$. Then B, $B \Rightarrow (A \Rightarrow B)$ (LAx 1), $A \Rightarrow B$ is a proof of $A \Rightarrow B$ in T.

(b) $B \approx A$. Then $\vdash A \Rightarrow A$ (Exercise 3.5.5 (a)), hence the more $T \vdash A \Rightarrow A$.

Otherwise there must be a proof $B_0, B_1, \ldots, B_n$ of B in the theory $T \cup \{A\}$ such that $n \geq 2$ and B_n (i.e., B) follows from some previous items of this sequence by (MP). We will proceed by induction according to n. To this end we assume that the

needed conclusion is valid for all proofs $C_0, C_1, \ldots, C_m$ in $T \cup \{A\}$ where $m < n$. Let $j, k < n$ be such that $B_j \approx (B_k \Rightarrow B_n)$. Then both $B_0, \ldots, B_j$ and $B_0, \ldots, B_k$ are proofs in $T \cup \{A\}$. By the induction assumption we have $T \vdash A \Rightarrow B_j$, i.e., $T \vdash A \Rightarrow (B_k \Rightarrow B_n)$, as well as $T \vdash A \Rightarrow B_k$. Then

$$(A \Rightarrow (B_k \Rightarrow B_n)) \Rightarrow ((A \Rightarrow B_k) \Rightarrow (A \Rightarrow B_n))$$

is (LAx 2), and by (MP) we consecutively get

$$T \vdash (A \Rightarrow B_k) \Rightarrow (A \Rightarrow B_n)$$
$$T \vdash A \Rightarrow B_n$$

i.e., $T \vdash A \Rightarrow B$.

The reader should notice that it is the "harder" implication

$$\text{if } T \cup \{A\} \vdash B \text{ then } T \vdash A \Rightarrow B$$

which is frequently used in mathematical proofs as well as in many deductive arguments elsewhere. A typical direct proof of the implication $A \Rightarrow B$ out of a list (theory) T of assumptions (axioms) starts with the "ritual" formulation: *"Let A"*, or *"Assume that A"*. This is nothing else than extending the axiom list T by a new axiom A. We continue by a sequence of statements $C_1, \ldots, C_n$ formed according to some deductive rules and finish once we succeed to arrive at the final term B. However, strictly speaking, what we have produced that way is a proof of B within the theory $T \cup \{A\}$ and not a proof of the implication $A \Rightarrow B$ in T as we claim. The Deduction Theorem shows that this natural method of argumentation is legitimate within Propositional Calculus, justifying our claim.

Another way of proving a statement out of some list of assumptions is the *proof by contradiction*. Instead of proving A in T directly, we produce a contradiction with the axioms of T out of the negation of A. Also this method is legitimate in Propositional Calculus.

A theory T is called *contradictory* or *inconsistent* if there exists some propositional form A such that both $T \vdash A$ and $T \vdash \neg A$. Otherwise, T is called *consistent*. From Exercise 3.5.5 (c) it follows that every propositional form B is provable in an inconsistent theory T.

3.7.2 Corollary on Proof by Contradiction. *Let $T \subseteq \mathrm{VF}(P)$ be a theory and $A \in \mathrm{VF}(P)$ be a propositional form. Then $T \vdash A$ if and only if the theory $T \cup \{\neg A\}$ is contradictory (inconsistent).*

Demonstration. Let $T \vdash A$. The more $T \cup \{\neg A\} \vdash A$. Since, clearly, $T \cup \{\neg A\} \vdash \neg A$, the theory $T \cup \{\neg A\}$ is contradictory.

Conversely, let the theory $T \cup \{\neg A\}$ be contradictory. Then every propositional form is provable in this theory; in particular, $T \cup \{\neg A\} \vdash A$. Then $T \vdash \neg A \Rightarrow A$ by the Deduction Theorem. According to Exercise 3.5.5 (h), $\vdash (\neg A \Rightarrow A) \Rightarrow A$, and the more $T \vdash (\neg A \Rightarrow A) \Rightarrow A$. Using (MP) we get $T \vdash A$.

Sometimes we are unable to find a proof of a statement B in a theory T, however, we are able to prove B under some additional assumption A in one way, and in another way under the opposite assumption $\neg A$. Then, all the same, it follows that B is provable in T. This way of argumentation is legitimate in Propositional Calculus, as well.

3.7.3 Corollary on Proof by Distinct Cases. *Let $T \subseteq \mathrm{VF}(P)$ be a theory and $A, B \in \mathrm{VF}(P)$ be propositional forms. Then $T \cup \{A\} \vdash B$ and $T \cup \{\neg A\} \vdash B$ if and only if $T \vdash B$.*

Demonstration. Assume that $T \cup \{A\} \vdash B$ and $T \cup \{\neg A\} \vdash B$. According to the Deduction Theorem it follows $T \vdash A \Rightarrow B$ and $T \vdash \neg A \Rightarrow B$. By Exercise 3.5.5 (g) we have

$$\vdash (A \Rightarrow B) \Rightarrow ((\neg A \Rightarrow B) \Rightarrow B)$$

and applying (MP) twice we get $T \vdash B$.

Conversely, let $T \vdash B$. Then, trivially, $T \cup \{A\} \vdash B$, as well as $T \cup \{\neg A\} \vdash B$.

3.7.4 Exercise. *Let $T \subseteq \mathrm{VF}(P)$ be a theory and $A_1, \ldots, A_n, B \in \mathrm{VF}(P)$ be propositional forms such that $T \vdash A_1 \vee \ldots \vee A_n$. Show that $T \vdash B$ if and only if $T \cup \{A_i\} \vdash B$ for each $i = 1, \ldots, n$.*

3.8 The Completeness Theorem

We start with a technical lemma. Given any interpretation $I\colon \mathrm{VF}(P) \to \{0, 1\}$ and a propositional form $A \in \mathrm{VF}(P)$ we denote

$$A^I \approx \begin{cases} A & \text{if } I(A) = 1 \\ \neg A & \text{if } I(A) = 0 \end{cases}$$

In other words, A^I is namely that member of the couple A, $\neg A$ which is true in I, i.e., $I(A^I) = 1$.

3.8.1 Interpretation Lemma. [A. Church] *Let $Q = \{p_1, \ldots, p_n\} \subseteq P$ be a finite set of propositional variables and $A \in \mathrm{VF}(Q)$. Then for any interpretation $I\colon \mathrm{VF}(P) \to \{0, 1\}$ we have*

$$\{p_1^I, \ldots, p_n^I\} \vdash A^I$$

Demonstration. By induction on complexity of A:

(a) If $A \approx p \in P$, then the statement means that $\{p\} \vdash p$, if $I(p) = 1$, or $\{\neg p\} \vdash \neg p$, if $I(p) = 0$. In both cases we get the needed conclusion.

(b) Let $A \approx \neg B$ and our conclusion be true for B. Then $B \in \mathrm{VF}(p_1, \ldots, p_n)$.

If $I(A) = 1$, then $I(B) = 0$ and $A^I \approx A \approx \neg B \approx B^I$. By the assumption, $\{p_1^I, \ldots, p_n^I\} \vdash B^I$, i.e., $\{p_1^I, \ldots, p_n^I\} \vdash A^I$.

If $I(A) = 0$, then $I(B) = 1$, $B^I \approx B$ and $A^I \approx \neg A \approx \neg\neg B$. By the assumption $\{p_1^I, \ldots, p_n^I\} \vdash B^I$, i.e., $\{p_1^I, \ldots, p_n^I\} \vdash B$. According to Exercise 3.5.5 (b) we have $\vdash B \Rightarrow \neg\neg B$, and by (MP) we get $\{p_1^I, \ldots, p_n^I\} \vdash \neg\neg B$, i.e., $\{p_1^I, \ldots, p_n^I\} \vdash A^I$.

(c) Let $A \approx (B \Rightarrow C)$ and for B, C the conclusion is true. Then $B, C \in \mathrm{VF}(p_1, \ldots, p_n)$. We distinguish three cases:

1. $I(B) = 0$. Then $I(A) = I(B \Rightarrow C) = 1$, i.e., $A^I \approx A$. Further, $B^I \approx \neg B$, hence, by the induction assumption, $\{p_1^I, \ldots, p_n^I\} \vdash \neg B$. According to Exercise 3.5.5 (c) we have $\vdash \neg B \Rightarrow (B \Rightarrow C)$ and by (MP) we get $\{p_1^I, \ldots, p_n^I\} \vdash B \Rightarrow C$, i.e., $\{p_1^I, \ldots, p_n^I\} \vdash A^I$.

2. $I(C) = 1$. Then $C^I \approx C$ and $I(A) = I(B \Rightarrow C) = 1$, hence $A^I \approx A$. By the induction assumption, $\{p_1^I, \ldots, p_n^I\} \vdash C$. (LAx 1) gives $\vdash C \Rightarrow (B \Rightarrow C)$, and by (MP) we get $\{p_1^I, \ldots, p_n^I\} \vdash B \Rightarrow C$, i.e., $\{p_1^I, \ldots, p_n^I\} \vdash A^I$.

3. $I(B) = 1$, $I(C) = 0$. Then $B^I \approx B$, $C^I \approx \neg C$ and $I(A) = I(B \Rightarrow C) = 0$, hence $A^I \approx \neg A$. By the induction assumption, $\{p_1^I, \ldots, p_n^I\} \vdash B$ and $\{p_1^I, \ldots, p_n^I\} \vdash \neg C$. Exercise 3.5.5 (f) gives $\vdash B \Rightarrow (\neg C \Rightarrow \neg(B \Rightarrow C))$. Using (MP) twice we get $\{p_1^I, \ldots, p_n^I\} \vdash \neg(B \Rightarrow C)$, i.e., $\{p_1^I, \ldots, p_n^I\} \vdash A^I$.

3.8.2 Exercise. Let $Q = \{p_1, \ldots, p_n\} \subseteq P$ be a finite set of propositional variables and $A \in \mathrm{VF}(Q)$. Let

$$\mathrm{TE}(A) = \big\{I \colon Q \to \{0,1\} \colon I(A) = 1\big\} = \{I_1, \ldots, I_m\}$$

denote the set of all truth evaluations I on the set of propositional variables Q such that A is true in I. Obviously, $m \leq 2^n$. For each $I \in \mathrm{TE}(A)$ we denote by

$$C_I = p_1^I \wedge \ldots \wedge p_n^I$$

the elementary conjunction corresponding to I. Finally, we put

$$A' = C_{I_1} \vee \ldots \vee C_{I_m}$$

Give reasons for the claim that $A' \in \mathrm{VF}(Q)$ is a disjunctive normal form logically equivalent to A (see Exercise 3.3.3).

A special case of the Completeness Theorem, due to Emil Post, deals with the provability of tautologies.

3.8.3 Post's Completeness Theorem. *For every propositional form $A \in \mathrm{VF}(P)$, we have $\vDash A$ if and only if $\vdash A$; in other words, A is a tautology if and only if A is provable just from the logical axioms.*

Demonstration. We just show that every tautology is provable from the logical axioms; the converse follows from the Soundness Theorem.

Let $A \in \mathrm{VF}(p_1, \ldots, p_n)$. Since A is a tautology, $I(A) = 1$ and $A^I \approx A$ for every truth evaluation $I \colon \{p_1, \ldots, p_n\} \to \{0,1\}$. By the Interpretation Lemma 3.8.1,

$$\{p_1^I, \ldots, p_n^I\} \vdash A$$

For any truth evaluation $J \colon \{p_1, \ldots, p_{n-1}\} \to \{0,1\}$, both possibilities $I_1(p_n) = 1$, $I_2(p_n) = 0$ jointly with the condition $I_1(p_k) = I_2(p_k) = J(p_k)$, for $k < n$, produce interpretations $I_1, I_2 \colon \{p_1, \ldots, p_n\} \to \{0,1\}$. Therefore, both

$$\{p_1^J \ldots, p_{n-1}^J, p_n\} \vdash A \qquad \text{and} \qquad \{p_1^J \ldots, p_{n-1}^J, \neg p_n\} \vdash A$$

According to Corollary 3.7.3 on Proof by Distinct Cases this implies

$$\{p_1^J, \dots, p_{n-1}^J\} \vdash A$$

Repeating this procedure we finally get $\vdash A$.

A theory $T \subseteq \mathrm{VF}(P)$ is called *complete* if it is consistent and for every propositional form $A \in \mathrm{VF}(P)$ we have $T \vdash A$ or $T \vdash \neg A$. In other words, T is complete if and only if for every propositional form A exactly one of the two possibilities $T \vdash A$, $T \vdash \neg A$ takes place.

Next we show an alternative version of the Completeness Theorem.

3.8.4 Completeness Theorem. [Alternative version] *Every consistent theory $T \subseteq \mathrm{VF}(P)$ has an interpretation.*

The reader is asked to realize that also the other way round, if a theory has an interpretation then it is necessarily consistent; in other words, a contradictory theory has no interpretation. (This is the alternative version of the Soundness Theorem.)

Demonstration. Any interpretation I of a consistent theory T has to satisfy

$$I(A) = \begin{cases} 1 & \text{if } T \vdash A \\ 0 & \text{if } T \vdash \neg A \end{cases}$$

Since T is consistent, $T \vdash A$ and $T \vdash \neg A$ cannot happen at once for any $A \in \mathrm{VF}(P)$. On the other hand, unless T is complete, we cannot guarantee that we always have either $T \vdash A$ or $T \vdash \neg A$, i.e., the value $I(A)$ need not be defined for every $A \in \mathrm{VF}(P)$. However, if T is *complete* then the above casework defines an interpretation of T, indeed. In other words, a complete theory T has exactly one interpretation.

In the general case, since the set $\mathrm{VF}(P)$ of all propositional forms is countable, it allows for some enumeration $\mathrm{VF}(P) = \{A_0, A_1, \dots, A_n, \dots\}$. Now we define a sequence of theories $T_0, T_1, \dots, T_n, \dots$ recursively:

$$T_0 = T \qquad \text{and} \qquad T_{n+1} = \begin{cases} T_n \cup \{A_n\} & \text{if } T_n \cup \{A_n\} \text{ is consistent} \\ T_n \cup \{\neg A_n\} & \text{if } T_n \cup \{A_n\} \text{ is contradictory} \end{cases}$$

Obviously, $T_n \subseteq T_{n+1}$ for each n. Let us show by induction on n that every T_n is a consistent theory. $T_0 = T$ is consistent by the initial assumption. Assuming that T_n is consistent, T_{n+1} could be inconsistent only in case that both the theories $T_n \cup \{A_n\}$, $T_n \cup \{\neg A_n\}$ were contradictory. By the Corollary on Proof by Contradiction this would mean that both $T_n \vdash \neg A_n$ and $T_n \vdash A_n$. However, this is impossible, as T_n is consistent.

Next we show that $\widehat{T} = \bigcup_{n \in \mathbb{N}} T_n$ is a complete theory. It is easy to realize that $\widehat{T}$ is consistent. Indeed, if $\widehat{T}$ were inconsistent then already some of the theories T_n would inconsistent as well (this is left to the reader — see also the demnostration of the Compactness Theorem 3.8.7). It remains to show that, for each n, either $\widehat{T} \vdash A_n$ or $\widehat{T} \vdash \neg A_n$. This is equivalent to showing that $\widehat{T} \nvdash \neg A_n$ implies $\widehat{T} \vdash A_n$. If $\widehat{T} \nvdash \neg A_n$

then $\widehat{T} \cup \{A_n\}$ is consistent, and $T_n \cup \{A_n\} \subseteq \widehat{T} \cup \{A_n\}$ is consistent, as well. Then $A_n \in T_{n+1}$, hence $T_{n+1} \vdash A_n$, and, since $T_{n+1} \subseteq \widehat{T}$, also $\widehat{T} \vdash A_n$.

Thus the unique interpretation I of the complete theory $\widehat{T}$ is an interpretation of T, as well.

3.8.5 Remark. The reader should notice that the above casework is not necessarily the only way how the sequence of theories $(T_n)_{n\in\mathbb{N}}$ extending T, leading to a complete theory $\widehat{T} = \bigcup T_n$, and the interpretation I could be defined. In each step when neither $T_n \vdash A_n$ nor $T_n \vdash \neg A_n$, we are free to choose either $T_{n+1} = T_n \cup \{A_n\}$ or $T_{n+1} = T_n \cup \{\neg A_n\}$.

3.8.6 Exercise. Let $I\colon P \to \{0,1\}$ be any truth evaluation. Let us denote

$$\mathrm{Th}(I) = \{p^I : p \in P\} = \{p \in P\colon I(p) = 1\} \cup \{\neg p\colon p \in P,\ I(p) = 0\}$$

the *theory of* I. Demonstrate the following facts:

(a) $\mathrm{Th}(I)$ is a complete propositional theory.

(b) For any propositional form $A \in \mathrm{VF}(P)$ the following conditions are equivalent:
   (i) $I(A) = 1$
   (ii) $\mathrm{Th}(I) \vdash A$
   (iii) $\mathrm{Th}(I) \vDash A$

Now, we can prove the original form of the Completeness Theorem. We state it in a way comprising the Soundness Theorem, as well.

3.8.7 Completeness Theorem. *Let $T \subseteq \mathrm{VF}(P)$ be a theory. Then, for every propositional form $B \in \mathrm{VF}(P)$, $T \vDash B$ if and only if $T \vdash B$.*

Demonstration. If $T \vdash B$ then $T \vDash B$ by the Soundness Theorem. To show the converse, assume that $T \vDash B$, nevertheless $T \nvdash B$. By the Theorem on Proof by Contradiction, this means that the theory $T \cup \{\neg B\}$ is consistent. Then, according to the Alternative Version of the Completeness Theorem, $T \cup \{\neg B\}$ has an interpretation I. Then I is an interpretation of the theory T such that $I(\neg B) = 1$, i.e., $I(B) = 0$. However, since $T \vDash B$, we have $J(B) = 1$ for every interpretation J of T; in particular, $I(B) = 1$. This contradiction proves that $T \vdash B$.

Finally, let us record the following consequence of the Completeness Theorem.

3.8.8 Compactness Theorem. *Let $T \subseteq \mathrm{VF}(P)$ be a theory. Then T has an interpretation if and only if every finite subtheory T_0 of T has an interpretation.*

Demonstration. By the Completeness Theorem, T has an interpretation if and only if T is consistent. Similarly, every finite subtheory $T_0 \subseteq T$ has an interpretation if and only if every finite subtheory $T_0 \subseteq T$ is consistent. Thus it is enough to realize that T is consistent if and only if every finite subtheory T_0 of T is consistent. Obviously, if T is consistent then so are all its subtheories (and not just the finite ones). The other way round, if T is inconsistent, then any proofs of some couple

of contradicting propositional forms B, $\neg B$ in T involve just finitely many specific axioms of T. Putting them together we obtain a finite subtheory $T_0 \subseteq T$ which is already contradictory.

We have formulated and proved the Compactness Theorem in Propositional Calculus mainly with the aim to prepare the ground for the Compactness Theorem in Predicate Calculus to come later on. However, the Propositional Calculus version of the Compactness Theorem lacks the importance and the plentitude of consequences of its Predicate Calculus version.

3.9 Axiomatization of Propositional Calculus Using Four Logical Connectives

For completeness' sake we include the axiomatization of Propositional Calculus using all the usual logical connectives $\neg$, $\wedge$, $\vee$ and $\Rightarrow$; the remaining connective $\Leftrightarrow$ is introduced via the logical equivalence

$$A \Leftrightarrow B \equiv (A \Rightarrow B) \wedge (B \Rightarrow A)$$

i.e., the left hand expression serves as the abbreviation for the right hand one. The corresponding list of logical axioms consists of ten axiom schemes. The only inference rule is Modus Ponens, again.

3.9.1 Logical axioms. (10 axiom schemes)
For any propositional forms A, B, C, the following propositional forms are logical axioms:

(LAx 1) $A \Rightarrow (B \Rightarrow A)$

(LAx 2) $(A \Rightarrow (B \Rightarrow C)) \Rightarrow ((A \Rightarrow B) \Rightarrow (A \Rightarrow C))$

(LAx 3) $(A \wedge B) \Rightarrow A$

(LAx 4) $(A \wedge B) \Rightarrow B$

(LAx 5) $A \Rightarrow (B \Rightarrow (A \wedge B))$

(LAx 6) $A \Rightarrow (A \vee B)$

(LAx 7) $B \Rightarrow (A \vee B)$

(LAx 8) $((A \Rightarrow C) \wedge (B \Rightarrow C)) \Rightarrow ((A \vee B) \Rightarrow C))$

(LAx 9) $((A \Rightarrow B) \wedge (A \Rightarrow \neg B)) \Rightarrow \neg A)$

(LAx 10) $A \vee \neg A$

3.9.2 Deduction rule: Modus Ponens

(MP) $\dfrac{A,\ A \Rightarrow B}{B}$ (from A and $A \Rightarrow B$ infer B)

3.9.3 Exercise. Show that all the above logical axioms are tautologies and explain their intuitive meaning.

4 First-Order Logic

Compared to Propositional Calculus, in First-Order Logic we relieve to some extent the requirement to abstract from the content of particular statements and arguments and to concentrate upon their form, only. Namely, we admit that all the statements and arguments deal with some *objects*, that these objects have some *properties* or enter some *relations*, that some objects are explicitly mentioned by their *names* and, finally, that from given objects some new objects can be produced by means of certain *operations*. Though this shift of focus makes sense equally for natural languages as well as for languages of scientific theories, we will restrict our attention to the languages of *mathematical theories* exclusively, where such an approach is rather natural and its fruitfulness can be demonstrated in a convincing way. On the other hand, the reduction of the logical structure of natural languages as well as of the languages of most scientific theories to its fragment fitting within the framework set by the First-Order Logic would turn out rather artificial.

First-Order Logic, also called (Lower) *Predicate Calculus*, examines the structure of arguments and proofs used in mathematics, more precisely in mathematical theories describing classes of mathematical structures formed by sets of objects endowed with various finitary operations and relations, singled out by certain axioms. That way First-Order Logic is mathematical logic both by its methods as well as by its subject.

We intend to develop and present the First-Order Logic in a way parallel, as much as possible, to our previous development and presentation of Propositional Calculus. Consequently, the reader should see clearly both the similarities as well as the differences between these two branches of mathematical logic.

4.1 First-Order Languages and First-Order Structures

A typical *mathematical structure* consists of a nonempty base set A of objects, equipped with some finitary operations, some distinguished elements and some finitary relations.

4.1.1 Example. Number systems with operations of addition $+$, multiplication $\cdot$, distinguished elements 0 and 1 and (with the exception of complex numbers) the ordering relation $<$ form mathematical structures commonly denoted as follows:

Natural numbers $(\mathbb{N}; +, \cdot, 0, 1, <)$
Integers $(\mathbb{Z}; +, \cdot, 0, 1, <)$
Rational numbers $(\mathbb{Q}; +, \cdot, 0, 1, <)$
Real numbers $(\mathbb{R}; +, \cdot, 0, 1, <)$
Complex numbers $(\mathbb{C}; +, \cdot, 0, 1)$

A *first-order language* $L = (F, C, R, \nu)$ is given by some (maybe void) sets F, C, R of functional (operation) symbols, constant symbols and relational (predicate) symbols, respectively, together with an *arity function* (*signature*) $\nu\colon F \cup R \to \mathbb{N}$, assigning to any symbol $s \in F \cup R$ its *arity* $\nu(s) \geq 1$. These are the *specific symbols* of L. Constant symbols are sometimes considered as nullary functional symbols, i.e., as elements f of the set F subject to $\nu(f) = 0$ (in that case the set C does not explicitly occur in the description of L).

All first-order languages contain common *logical symbols*:

- Object variables (or just variables): x, y, z, u, v, w, x_0, x_1, x_2, $\ldots$, y', y'', $\ldots$
- Logical connectives: $\neg$ (*not*), $\wedge$ (*and*), $\vee$ (*or*), $\Rightarrow$ (*if ... then* or *implies*), $\Leftrightarrow$ (*if and only if*) (two would suffice)
- Quantifiers: $\forall$ (*universal quantifier*), $\exists$ (*existential quantifier*) (one would suffice)
- Equality sign: $=$
- Auxiliary symbols: (,) (*parentheses*), , (*comma*) (they could be avoided)

4.1.2 Remark. At a glance it could seem that we should require the sets F, C, R to be at most countable, since it makes no sense to admit that the language L contains uncountably many specific symbols. Such a restriction, however, would bring us no technical advantage. More important, as we shall see later on, e.g., when dealing with various *diagrams* of structures, the methods and results of the study of uncountable languages have applications even for structures of countable first-order languages.

A *first-order structure*, i.e., a structure of some first-order language $L = (F, C, R, \nu)$, briefly, an *L-structure*, $\mathcal{A} = (A; I)$ consists of a nonempty set A, called *base set* or *underlying set* or *carrier* of $\mathcal{A}$, and an *interpretation* I of the specific symbols of L in A:

- for $f \in F$, such that $\nu(f) = n$, $f^I\colon A^n \to A$
  (each n-ary operation symbol is interpreted as an n-ary operation on A)
- for $c \in C$, $c^I \in A$
  (each constant symbol is interpreted as some distinguished element of A)
- for $r \in R$, such that $\nu(r) = n$, $r^I \subseteq A^n$
  (each n-ary relation symbol is interpreted as an n-ary relation on A)

Instead of s^I we frequently write $s^{\mathcal{A}}$ or just s for any specific symbol s.

4.2 Terms and Formulas

Terms of a first-order language L, briefly *L-terms*, are composed of variables, constant symbols and functional symbols of L. The set $\mathrm{Term}(L)$ of all L-terms is the smallest set such that

1° $x \in \mathrm{Term}(L)$ for each variable x (every variable is a term);

2° $c \in \mathrm{Term}(L)$ for each $c \in C$ (every constant symbol is a term);

3° if $f \in F$, $\nu(f) = n$, and $t_1, \ldots, t_n \in \mathrm{Term}(L)$, then $f(t_1, \ldots, t_n) \in \mathrm{Term}(L)$
(if f is an n-ary functional symbol and $t_1, \ldots, t_n$ are terms, then $f(t_1, \ldots, t_n)$ is a term, as well).

This is a *recursive definition*. Terms in 1°, 2°, i.e., variables and constant symbols, are called *atomic terms*. In order to prove that some set of L-terms contains all the L-terms, it suffices to show that it fulfills all the three conditions 1°, 2° and 3°.

For a binary operation symbol f we usually write $t_1 f t_2$ instead of $f(t_1, t_2)$ in 3°.

Constant terms are terms containing no variables. If all the variables occurring in a term t are contained in the list $x_1, \ldots, x_k$, we write $t(x_1, \ldots, x_k)$. There is one exception: writing t doesn't necessarily mean that t is a constant term.

Given an L-structure $\mathcal{A} = (A; I)$ and a term $t(x_1, \ldots, x_k)$, the interpretation of t in $\mathcal{A}$ is a k-ary operation $t^I \colon A^k \to A$ defined on any k-tuple $(a_1, \ldots, a_k) \in A^k$ as follows:

1° if t is the variable x_i where $i \leq k$, then $t^I(a_1, \ldots, a_k) = a_i$;

2° if t is a constant symbol $c \in C$, then $t^I(a_1, \ldots, a_k) = c^I$;

3° if t is of the form $f(t_1, \ldots, t_n)$ where $f \in F$, $\nu(f) = n$, and the interpretations t_j^I of the terms $t_1(x_1, \ldots, x_k)$, …, $t_n(x_1, \ldots, x_k)$ are already defined, then

$$t^I(a_1, \ldots, a_k) = f^I\big(t_1^I(a_1, \ldots, a_k), \ldots, t_n^I(a_1, \ldots, a_k)\big)$$

In particular, the interpretation of a constant term t is always an element $t^I \in A$. Instead of t^I we frequently write $t^{\mathcal{A}}$ or just t for any term t.

Formulas of a first-order language L, briefly *L-formulas* are built of atomic formulas by means of logical connectives and quantifiers. *Atomic formulas* of the language L are expressions of the form $t_1 = t_2$ where t_1, t_2 are arbitrary L-terms, and $r(t_1, \ldots, t_n)$ where $r \in R$ is a relational symbol, $\nu(r) = n$, and $t_1, \ldots, t_n$ are arbitrary L-terms (instead of $r(t_1, t_2)$ we frequently write $t_1 \, r \, t_2$). The set $\mathrm{Form}(L)$ of all L-formulas is the smallest set such that

1° if φ is an atomic formula then $\varphi \in \mathrm{Form}(L)$ (every atomic formula is a formula);

2° if $\varphi, \psi \in \mathrm{Form}(L)$ then $\neg\varphi$, $(\varphi \wedge \psi)$, $(\varphi \vee \psi)$, $(\varphi \Rightarrow \psi)$, $(\varphi \Leftrightarrow \psi) \in \mathrm{Form}(L)$
(if φ, ψ are L-formulas then so are $\neg\varphi$, $(\varphi \wedge \psi)$, $(\varphi \vee \psi)$, $(\varphi \Rightarrow \psi)$, $(\varphi \Leftrightarrow \psi)$);

3° if $\varphi \in \mathrm{Form}(L)$ and x is a variable, then $(\forall x)\varphi$, $(\exists x)\varphi \in \mathrm{Form}(L)$
(for any formula φ and variable x, the expressions $(\forall x)\varphi$, $(\exists x)\varphi$ are formulas).

This is a *recursive definition*, again. In order to prove that some set of L-formulas contains all the L-formulas, it suffices to show that it fulfills all the three conditions 1°, 2° and 3°.

Similarly as in Propositional Calculus, we tend to omit unnecessary parentheses. On the other hand, in order to promote readability, we sometimes use parentheses not required by the above definition. For instance, the modified expressions $\neg(\varphi)$, $(\varphi) \wedge (\psi)$, etc., could sometimes be better legible than the "rigorous" formulas $\neg\varphi$, $\varphi \wedge \psi$, etc., respectively. Atomic formulas of the form $t_1 = t_2$ are called *identities*. The conjunction of two identities $(t_1 = t_2) \wedge (t_2 = t_3)$ is frequently abbreviated to $t_1 = t_2 = t_3$. Instead of $\neg(t_1 = t_2)$ we usually write $t_1 \neq t_2$; $\neg(t_1 \, r \, t_2)$ is sometimes abbreviated to $t_1 \not r \, t_2$.

Consecutive quantifications with the same quantifier $(\mathrm{Q}\,x_1)\ldots(\mathrm{Q}\,x_n)$ will be abbreviated to $(\mathrm{Q}\,x_1,\ldots,x_n)$. For instance, we will write $(\forall\,x_1,\ldots,x_n)(\exists\,u,v)(\forall\,z)\varphi$ instead of $(\forall\,x_1)\ldots(\forall\,x_n)(\exists\,u)(\exists\,v)(\forall\,z)\varphi$.

An *occurrence of a variable* x in a formula φ is simply any occurrence of the symbol x in some of the atomic formulas out of which φ is built. Such an occurrence is called *bounded* if it falls under the range of some quantifier, otherwise it is called *free*.

4.2.1 Example. In the formula φ

$$(\forall\,x)(x+y=y+x)\wedge(\exists\,y)(\forall\,u)(x\le y+z)$$

of the first-order language containing a binary operation symbol $+$ and a binary predicate symbol $\le$ both the occurrences of x in the atomic formula $x+y=y+x$ are bounded while its occurrence in $x\le y+z$ is free, both the occurrences of y in the atomic formula $x+y=y+x$ are free, while its occurrence in $x\le y+z$ is bounded, the occurrence of z in the atomic formula $x\le y+z$ is free, finally, the variable u has no occurrence in φ.

Sentences or *closed formulas*, sometimes also referred to as *statements*, are L-formulas containing no free variables. If all the variables occurring *freely* in a formula φ are contained in the list $x_1,\ldots,x_n$, we write $\varphi(x_1,\ldots,x_n)$. Exception: writing φ doesn't necessarily mean that φ is a closed formula.

The following definition of the *satisfaction relation* is due to Alfred Tarski. The satisfaction of an L-formula $\varphi(x_1,\ldots,x_n)$ by the elements $a_1,\ldots,a_n\in A$ of some L-structure $\mathcal{A}=(A;\,I)$ is denoted by $\mathcal{A}\vDash\varphi(a_1,\ldots,a_n)$ and read as $\varphi(a_1,\ldots,a_n)$ *is satisfied* or *true in* $\mathcal{A}$. It is defined recursively:

- if φ is $t_1=t_2$, then $\mathcal{A}\vDash\varphi(a_1,\ldots,a_n)$ if and only if $t_1^I(a_1,\ldots,a_n)=t_2^I(a_1,\ldots,a_n)$
- if φ is $r(t_1,\ldots,t_m)$, then $\mathcal{A}\vDash\varphi(a_1,\ldots,a_n)$ if and only if $\big(t_1^I(a_1,\ldots,a_n),\ldots,t_m^I(a_1,\ldots,a_n)\big)\in r^I$
- if φ is $\neg\psi$, then $\mathcal{A}\vDash\varphi(a_1,\ldots,a_n)$ if and only if it is not true that $\mathcal{A}\vDash\psi(a_1,\ldots,a_n)$
- if φ is $\psi\wedge\chi$, then $\mathcal{A}\vDash\varphi(a_1,\ldots,a_n)$ if and only if both $\mathcal{A}\vDash\psi(a_1,\ldots,a_n)$ and $\mathcal{A}\vDash\chi(a_1,\ldots,a_n)$
- if φ is $\psi\vee\chi$, then $\mathcal{A}\vDash\varphi(a_1,\ldots,a_n)$ if and only if $\mathcal{A}\vDash\psi(a_1,\ldots,a_n)$ or $\mathcal{A}\vDash\chi(a_1,\ldots,a_n)$ (in the nonexclusive meaning)
- if φ is $\psi\Rightarrow\chi$, then $\mathcal{A}\vDash\varphi(a_1,\ldots,a_n)$ if and only if from $\mathcal{A}\vDash\psi(a_1,\ldots,a_n)$ it follows that $\mathcal{A}\vDash\chi(a_1,\ldots,a_n)$
- if φ is $\psi\Leftrightarrow\chi$, then $\mathcal{A}\vDash\varphi(a_1,\ldots,a_n)$ if and only if the conditions $\mathcal{A}\vDash\psi(a_1,\ldots,a_n)$ and $\mathcal{A}\vDash\chi(a_1,\ldots,a_n)$ are equivalent
- if φ is $(\forall\,x)\psi$, then $\mathcal{A}\vDash\varphi(a_1,\ldots,a_n)$ if and only if $\mathcal{A}\vDash\psi(a,a_1,\ldots,a_n)$ for every $a\in A$
- if φ is $(\exists\,x)\psi$, then $\mathcal{A}\vDash\varphi(a_1,\ldots,a_n)$ if and only if $\mathcal{A}\vDash\psi(a,a_1,\ldots,a_n)$ for some $a\in A$

The above list can be given an easier to survey and remember form rewriting it using the same symbols for logical connectives, quantifiers and the relation of equality, both in the first-order language L we are dealing with, as well as in the common English (which is a part of our metalanguage). In order to avoid the impending confusion, such an attitude is usually introduced with the phrase "by abuse of notation". The reader should carefully inspect the rewritten version and identify the corresponding role of every particular occurrence of the logical symbols in the new list below:

$$\begin{aligned}
\mathcal{A} \vDash (t_1 = t_2)(a_1, \dots, a_n) &\Leftrightarrow \mathcal{A} \vDash t_1(a_1, \dots, a_n) = t_2(a_1, \dots, a_n) \\
\mathcal{A} \vDash r(t_1, \dots, t_m)(a_1, \dots, a_n) &\Leftrightarrow \mathcal{A} \vDash r(t_1(a_1, \dots, a_n), \dots, t_m(a_1, \dots, a_n)) \\
\mathcal{A} \vDash \neg\varphi(a_1, \dots, a_n) &\Leftrightarrow \neg(\mathcal{A} \vDash \varphi(a_1, \dots, a_n)) \Leftrightarrow \mathcal{A} \nvDash \varphi(a_1, \dots, a_n) \\
\mathcal{A} \vDash (\varphi \wedge \psi)(a_1, \dots, a_n) &\Leftrightarrow \big(\mathcal{A} \vDash \varphi(a_1, \dots, a_n) \wedge \mathcal{A} \vDash \psi(a_1, \dots, a_n)\big) \\
\mathcal{A} \vDash (\varphi \vee \psi)(a_1, \dots, a_n) &\Leftrightarrow \big(\mathcal{A} \vDash \varphi(a_1, \dots, a_n) \vee \mathcal{A} \vDash \psi(a_1, \dots, a_n)\big) \\
\mathcal{A} \vDash (\varphi \Rightarrow \psi)(a_1, \dots, a_n) &\Leftrightarrow \big(\mathcal{A} \vDash \varphi(a_1, \dots, a_n) \Rightarrow \mathcal{A} \vDash \psi(a_1, \dots, a_n)\big) \\
\mathcal{A} \vDash (\varphi \Leftrightarrow \psi)(a_1, \dots, a_n) &\Leftrightarrow \big(\mathcal{A} \vDash \varphi(a_1, \dots, a_n) \Leftrightarrow \mathcal{A} \vDash \psi(a_1, \dots, a_n)\big) \\
\mathcal{A} \vDash (\forall x)\varphi(x, a_1, \dots, a_n) &\Leftrightarrow (\forall a \in A)\big(\mathcal{A} \vDash \varphi(a, a_1, \dots, a_n)\big) \\
\mathcal{A} \vDash (\exists x)\varphi(x, a_1, \dots, a_n) &\Leftrightarrow (\exists a \in A)\big(\mathcal{A} \vDash \varphi(a, a_1, \dots, a_n)\big)
\end{aligned}$$

For $\varphi(x_1, \dots, x_n, y_1, \dots, y_m)$ and $a_1, \dots, a_n \in A$ we write $\mathcal{A} \vDash \varphi(a_1, \dots, a_n, y_1, \dots, y_m)$ if and only if $\mathcal{A} \vDash \varphi(a_1, \dots, a_n, b_1, \dots, b_m)$ for all $b_1, \dots, b_m \in A$, i.e., if and only if $\mathcal{A} \vDash (\forall y_1, \dots, y_m)\varphi(a_1, \dots, a_n, y_1, \dots, y_m)$. In particular, $\mathcal{A} \vDash \varphi(x_1, \dots, x_n)$ means that $\mathcal{A} \vDash \varphi(a_1, \dots, a_n)$ for all $a_1, \dots, a_n \in A$, i.e., $\mathcal{A} \vDash (\forall x_1, \dots, x_n)\varphi(x_1, \dots, x_n)$. Thus the expression $\mathcal{A} \vDash \varphi$ has clear meaning even if the formula φ is not closed.

4.3 First-Order Theories and Models

A *first-order theory* T, i.e., a theory in a first-order language L, is represented by and identified with the set of its *specific axioms* $T \subseteq \mathrm{Form}(L)$. From now on, as a rule, a first-order theory will be referred to simply as a *theory*. An L-structure $\mathcal{A}$ is said to *satisfy* the theory T or to be a *model* of T if $\mathcal{A}$ satisfies all the axioms of T. We write $\mathcal{A} \vDash T$; thus we have

$$\mathcal{A} \vDash T \quad \text{if and only if} \quad \mathcal{A} \vDash \varphi \text{ for every } \varphi \in T$$

We say that a formula ψ is a *logical consequence* (of the axioms) of the first-order theory T, or that ψ is *true* in T if ψ is true in every model $\mathcal{A}$ of the theory T. In that case we write $T \vDash \psi$. Thus we have

$$T \vDash \psi \quad \text{if and only} \quad \mathcal{A} \vDash \psi \text{ for every model } \mathcal{A} \vDash T$$

Our goal is to describe the semantic notion of truth or satisfaction or logical consequence in terms of the syntactic notion of *provability*. This will take part in the next section. However, before turning our attention to that point, it is desirable to get some acquaintance with a handful of important examples of first-order theories and their models.

Preliminarily, we can divide first-order theories into the following two categories:

(a) First-order theories describing a variety of different structures sharing the same first-order language and singled out by some common properties formulated as axioms of the corresponding theory. Some subclasses of the class of all models of that theory can be described in terms of some additional axioms, as well as by some properties not formulated in terms of that first-order language. From the theories listed bellow the *Theory of Groups*, various *Theories of Rings* and *Fields*, *Vector Spaces, Theories of Order, Boolean Algebras* and the *Theories of Ordered Rings* and *Fields* belong to this family.

(b) First-order theories attempting to describe a single mathematical structure "as completely as possible". As we shall see later on, such attempts are unattainable, except for some trivial cases. Our first example of this kind is furnished by *Peano Arithmetic*, aiming to fully describe the structure of all natural numbers with the addition and multiplication. The second example includes the *Zermelo-Fraenkel Set* Theory with the Axiom of Choice (in a not quite precisely delineated version) which should possibly faithfully grasp the Universe of Sets.

4.3.1 Groups. The *Theory of Groups* or *Group Theory* has several alternative axiomatizations in slightly differing first-order languages. A *group* is simply a model of Group Theory.

(a) In the first-order language containing a binary operation symbol $\cdot$ (multiplication), a constant symbol e (unit or neutral element) and a unary operation symbol $^{-1}$ (taking inverses), the axioms of the Theory of Groups are formed by the following identities:

$$x \cdot (y \cdot z) = (x \cdot y) \cdot z$$
$$x \cdot e = x = e \cdot x$$
$$x \cdot x^{-1} = e = x^{-1} \cdot x$$

expressing the associativity of the multiplication and the facts that e is its unit element and that x^{-1} is the inverse element of x. Then groups are structures $\left(G; \cdot, e, {}^{-1}\right)$ satisfying the just stated axioms.

A group is called *commutative* or *abelian* if it satisfies the commutative law $x \cdot y = y \cdot x$. Group Theory is sometimes, especially in the abelian case, formulated in the language using the binary operation symbol $+$ (addition), the constant symbol 0 (zero) and the unary operation symbol $-$ (minus, the inverse element $-x$ is called the opposite element to x).

(b) Omitting the unary operation symbol $^{-1}$ from the language of Group Theory, the last axiom, expressing the existence of inverse elements, has to be formulated in a slightly more complicated way

$$(\forall\, x)(\exists\, y)(x \cdot y = e = y \cdot x)$$

Then groups are considered as structures $(G; \cdot, e)$ satisfying the corresponding three axioms.

(c) In the language containing just the symbol of multiplication, the axioms expressing the existence of the unit element and of the inverse elements are merged into a single more complex axiom

$$(\exists\, u)(\forall\, x)\big(x \cdot u = x = u \cdot x \wedge (\exists\, y)(x \cdot y = u = y \cdot x)\big)$$

Another possibility is represented by the axiom

$$(\forall\, x)(\exists\, y)(\forall\, z)\big(z \cdot (x \cdot y) = z = (y \cdot x) \cdot z\big)$$

Then a group is a structure $(G;\, \cdot)$ satisfying the associative law and one (hence both) of the last two axioms.

It is clear that a group $\big(G;\, \cdot, e, {}^{-1}\big)$ in the sense of (a) can be made a group in the sense of (b) or (c) by omitting the interpretations of the superfluous symbols. The other way round, for Group Theory in the sense of (c) one can extend its language by the missing symbols and define the unit element and the inverse element operation by

$$\begin{aligned} u = e \;&\Leftrightarrow\; (\forall\, x)(x \cdot u = x = u \cdot x) \\ y = x^{-1} \;&\Leftrightarrow\; x \cdot y = e = y \cdot x \end{aligned}$$

respectively. Then a group $(G;\, \cdot)$ in the sense of (c) becomes a group in the sense of (b) or (a).

Some examples of commutative groups (in the language with a single binary operation) are the additive groups $(\mathbb{Z};\, +)$ of integers, $(\mathbb{Z}_n;\, +)$ of remainders modulo $n \geq 2$, of rationals $(\mathbb{Q};\, +)$, etc. Some examples of noncommutative groups are provided by the structures $(S(X);\, \circ)$ of all bijective maps (permutations) of any set X with more than two elements into itself and the operation of composition, or by the structures $(\mathrm{GL}(n, \mathbb{R});\, \cdot)$ of all invertible real $n \times n$ matrices, for $n \geq 2$, with the operation of matrix multiplication.

4.3.2 Rings and Fields. (a) The *Theory of Rings* or *Ring Theory* is formulated in the language with two binary operation symbols $+$ (addition), $\cdot$ (multiplication) and a constant symbol 0 (zero). Then a ring is a structure $(A;\, +, \cdot, 0)$ satisfying the axioms of Ring Theory. The axioms express that $(A;\, +, 0)$ is an abelian group, the associative law for multiplication and two distributive laws

$$x \cdot (y + z) = (x \cdot y) + (x \cdot z) \qquad (x + y) \cdot z = (x \cdot z) + (y \cdot z)$$

usually written simply as $x(y + z) = xy + xz$ and $(x + y)z = xz + yz$. Models of Ring Theory are called *rings*.

A ring is called *commutative* if it satisfies the commutative law for multiplication $x \cdot y = y \cdot x$. A commutative ring is called an *integral domain* if satisfies the axiom

$$xy = 0 \;\Rightarrow\; (x = 0 \vee y = 0)$$

(b) The *Theory of Rings with Unit* or the *Theory of Unitary Rings* is formulated in the language obtained by extending the language of Ring Theory by a new constant

symbol 1, denoting the unit of multiplication and adding the identities $x \cdot 1 = x = 1 \cdot x$ to the axioms of Ring Theory.

(c) The *Theory of Fields* is obtained from the Theory of Commutative Rings with Unit by adding to it the axiom $0 \neq 1$ and the axiom

$$(\forall\, x)(x \neq 0 \;\Rightarrow\; (\exists\, y)(x \cdot y = 1))$$

requiring the existence of multiplicative inverses for all nonzero elements. A *field* is simply a model of the Theory of Fields.

(d) The *Theory of Division Rings* is obtained from the Theory of Unitary Rings by extending it by the condition $0 \neq 1$ and the axiom of inverses

$$(\forall\, x)(x \neq 0 \;\Rightarrow\; (\exists\, y)(x \cdot y = 1 = y \cdot x))$$

A *division ring* is simply a model of this theory.

Even integers form a commutative ring $(2\mathbb{Z};\, +, \cdot, 0)$ without unit. The integers $(\mathbb{Z};\, +, \cdot, 0, 1)$ and the remainders $(\mathbb{Z}_n;\, +, \cdot, 0, 1)$ modulo $n \geq 2$ form commutative rings with unit. Examples of noncommutative rings with unit are provided by the structures $\big(\mathbb{R}^{n\times n};\, +, \cdot, 0, I\big)$ of all real $n \times n$ matrices, for $n \geq 2$, with operations of addition and multiplication of matrices, the zero matrix 0 and the unit matrix I. Examples of fields are the structures $(\mathbb{Q};\, +, \cdot, 0, 1)$, $(\mathbb{R};\, +, \cdot, 0, 1)$, $(\mathbb{C};\, +, \cdot, 0, 1)$ of rational, real and complex numbers, respectively, as well as the structures $(\mathbb{Z}_p;\, +, \cdot, 0, 1)$ of remainders modulo any prime number p. Clearly, every field is an integral domain; however, the integers $(\mathbb{Z};\, +, \cdot, 0, 1)$ form an integral domain which is not a field. An example of a non commutative division ring, i.e., a division ring which is not a field, is provided by the *quaternions* $(\mathbb{H};\, +, \cdot, 0, 1)$. Quaternions represent a four dimensional version of complex numbers, i.e., they are numbers of the form $q_0 + q_1\,\mathrm{i} + q_1\,\mathrm{j} + q_3\,\mathrm{k}$, where $q_0, q_1, q_2, q_3 \in \mathbb{R}$ and i, j, k are three imaginary units, satisfying $\mathrm{i}^2 = \mathrm{j}^2 = \mathrm{k}^2 = \mathrm{i\,j\,k} = -1$. The equality and addition of quaternions are defined componentwise, while their multiplication is defined in the only possible way extending the above relations between the generators i, j, k enforced by the axioms of unitary rings.

(e) A field $(F;\, +, \cdot, 0, 1)$ is called *algebraically closed* if it satisfies the infinite list of axioms

$$(\forall\, u_1, \ldots, u_n)(\exists\, x)\big(x^n + u_1 x^{n-1} + \cdots + u_{n-1}x + u_n = 0\big)$$

postulating the existence of roots of all polynomials of any degree $n \geq 2$ with coefficients from F. The field $(\mathbb{C};\, +, \cdot, 0, 1)$ of all complex numbers and the field $(\mathbb{A};\, +, \cdot, 0, 1)$ of all algebraic numbers (i.e., the roots of polynomials with rational coefficients) are examples of algebraically closed fields.

(f) A field $(F;\, +, \cdot, 0, 1)$ is called *formally real* if it satisfies the infinite list of axioms

$$x_1^2 + \cdots + x_n^2 = 0 \;\Rightarrow\; x_1 = \ldots = x_n = 0$$

for every positive $n \in \mathbb{N}$, requiring that a sum of squares of nonzero elements is never 0. The *Theory of Real Closed Fields* is the extension of the *Theory of Formally Real Fields* by the axiom

$$(\forall\, x)(\exists\, y)\big(y^2 = x \;\vee\; y^2 = -x\big)$$

postulating the existence of the square root of either x or $-x$ for any x, as well as the infinite list of axioms

$$(\forall\, u_1, \ldots, u_n)(\exists\, x)\big(x^n + u_1 x^{n-1} + \cdots + u_{n-1}x + u_n = 0\big)$$

guaranteeing the existence of at least one root for every polynomial of any *odd* degree $n \geq 3$ with coefficients from F. The field $(\mathbb{R};\, +, \cdot, 0, 1)$ of all real numbers and the field $(\mathbb{R} \cap \mathbb{A};\, +, \cdot, 0, 1)$ of all real algebraic numbers are examples of real closed fields. The field $(\mathbb{Q};\, +, \cdot, 0, 1)$ of all rational numbers is formally real but not real closed.

The following, though rather familiar example is worthwhile to notice since it shows that the sets of functional or relational symbols of a first-order language may themselves carry their own first-order structure.

4.3.3 Vector spaces over a field. For a fixed field $(F;\, +, \cdot, 0, 1)$ we introduce the first-order language $L(F)$ which has no relational symbols, a single constant symbol $\mathbf{0}$, a binary operation symbol $+$ and the set F of unary operation symbols. A typical structure of the language $L(F)$ is denoted as $\mathcal{V} = (V; F, +, \mathbf{0})$; the elements of the set V are called *vectors*. The elements $f \in F$ in role of unary operation symbols are referred to as *scalars*; instead of $f(x)$ we usually write just fx and this result is referred as the *scalar multiple* of x by f. *Vector spaces* over the field F are structures $\mathcal{V} = (V; F, +, \mathbf{0})$ of the language $L(F)$ satisfying the axioms expressing that $(V; +, \mathbf{0})$ is an abelian group, as well as the axioms

$$\begin{gathered} 1x = x \\ f(x + y) = fx + fy \\ (fg)x = f(gx) \\ (f + g)x = fx + gx \end{gathered}$$

for any scalars $f, g \in F$. The reader should realize that the last three equalities are in fact *axiom schemes* and not single axioms.

4.3.4 Theories of Order. The *Theory of Partial Order* is formulated alternatively in the first-order language with a single binary relational symbol $<$ (strict partial order) or $\leq$ (non-strict partial order). The strict version is given by the axioms

$$\neg(x < x) \qquad \neg(x < y \wedge y < x) \qquad (x < y \wedge y < z) \Rightarrow x < z$$

The non-strict (and more frequently used) version has the axioms

$$x \leq x \qquad (x \leq y \wedge y \leq x) \Rightarrow x = y \qquad (x \leq y \wedge y \leq z) \Rightarrow x \leq z$$

A *partially ordered set* (*poset*) is simply a model $(P;\, <)$ or $(P;\, \leq)$ of the corresponding version of the theory.

Obviously, a poset $(P;\, <)$ can be converted into a poset $(P;\, \leq)$ defining the non-strict partial order by

$$x \leq y \Leftrightarrow x < y \vee x = y$$

Vice versa, a poset $(P; \leq)$ can be made to a poset $(P; <)$ defining the strict partial order by

$$x < y \Leftrightarrow x \leq y \wedge x \neq y$$

The reader should be able to switch between the two versions anytime.

The *Theory of Ordered Sets*, sometimes called also the *Theory of Linear Order*, or the *Theory of Total Order*is obtained by adding the *trichotomy axiom*

$$x < y \vee x = y \vee y < x$$

or the *dichotomy axiom*

$$x \leq y \vee y \leq x$$

respectively, to the corresponding version of the Theory of Partial Order.

4.3.5 Boolean Algebras. The language of the *Theory Boolean Algebras* has two binary operation symbols $\wedge$ (meet), $\vee$ (join), one unary operation symbol $'$ (complement) and two constant symbols 0 and 1. This notation violates the implicit convention that the specific symbols of any first-order language should be clearly distinguished from its logical symbols. However, since all the axioms of the Theory of Boolean Algebras are identities and do not contain any logical connectives, there is no danger of confusion. On the other hand, this notation points to the familiar connection between Boolean algebras and Propositional Calculus. A *Boolean algebra* $\mathcal{B} = (B; \wedge, \vee, ', 0, 1)$ is simply a model of the theory with the following axioms:

$x \wedge y = y \wedge x$	$x \vee y = y \vee x$	(commutative laws)
$x \wedge (y \wedge z) = (x \wedge y) \wedge z$	$x \vee (y \vee z) = (x \vee y) \vee z$	(associative laws)
$x \wedge x = x$	$x \vee x = x$	(idempotent laws)
$x \wedge (x \vee y) = x$	$x \vee (x \wedge y) = x$	(absorbtion laws)
$x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$	$x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z)$	(distributive laws)
$(x \wedge y)' = x' \vee y'$	$(x \vee y)' = x' \wedge y'$	(De Morgan laws)
$x \wedge 0 = 0$	$x \vee 0 = x$	(laws of 0)
$x \wedge 1 = x$	$x \vee 1 = 1$	(laws of 1)
$x \wedge x' = 0$	$x \vee x' = 1$	(laws of complement)

This axiom list is redundant: some of its items can be derived as consequences of the remaining ones. A partial order on every Boolean algebra can be introduced via any of the following two equivalent conditions:

$$x \leq y \Leftrightarrow x = x \wedge y \Leftrightarrow x \vee y = y$$

Then 0 and 1 become the smallest and the biggest element, respectively, with respect to this partial order in $\mathcal{B}$.

The typical examples of Boolean algebras are formed by the power sets of any sets. More precisely, for every set I its power set $\mathcal{P}(I) = \{X \colon X \subseteq I\}$ with the operations of set-theoretical intersection, union and complement with respect to the

set I, the empty set $\emptyset$ as 0 and the whole set I as 1, gives rise to the Boolean algebra $(\mathcal{P}(I); \cap, \cup, ', \emptyset, I)$. The partial order $X \leq Y$ on $\mathcal{P}(I)$ coincides with the set-theoretical inclusion $X \subseteq Y$.

Another example of a Boolean algebra can be obtained from the set $\mathrm{VF}(P)$ of all propositional forms over any set of propositional variables $P \neq \emptyset$ interpreting the equality relation as the relation of logical equivalence $A \equiv B$, with logical connectives $\wedge$, $\vee$ and $\neg$ in the role of the Boolean operations of meet, join and complement, respectively, and with any tautology in the role of 1 and any contradiction in the role of 0.

4.3.6 Exercise. (a) Show that both the conditions defining the relation of partial order $\leq$ on Boolean algebras are indeed equivalent, and that $\leq$ defined that way is indeed a (non-strict) partial order.

(b) Show that the meet $x \wedge y$ is the infimum and the join $x \vee y$ is the supremum of the set $\{x, y\}$ with respect to the partial order $\leq$, respectively. This is to say that for any z we have $z \leq x$ and $z \leq y$ if and only if $z \leq x \wedge y$, and, similarly, $x \leq z$ and $y \leq z$ if and only if $x \vee y \leq z$.

(c) Prove the *law of double complement* $x'' = x$, as well as the identities $0' = 1$, $1' = 0$ from the axioms of Boolean algebras.

4.3.7 Ordered Rings and Fields. The *Theory of Ordered Rings* is obtained by extending the language of Ring Theory by the binary relational symbol $<$ and adding to its axioms the strict version of the axioms of (total) order, as well as the axioms

$$x < y \;\Rightarrow\; x + z < y + z$$
$$(x < y \;\wedge\; 0 < z) \;\Rightarrow\; (xz < yz \;\wedge\; zx < zy)$$

expressing that the operations of the addition of any element z as well as the multiplication by any positive element z are increasing.

The same procedure applies to the Theory of Unitary Rings, the Theory of Commutative Rings, the Theory of Fields, etc., yielding the *Theory Ordered Unitary Rings*, the *Theory of Ordered Commutative Rings*, the *Theory of Ordered Fields*, etc., respectively. The integers $(\mathbb{Z}; +, \cdot, 0, 1, <)$ provide a representative (and minimal) example of an ordered commutative ring with unit. The rationals $(\mathbb{Q}; +, \cdot, 0, 1, <)$ and the reals $(\mathbb{R}; +, \cdot, 0, 1, <)$ form ordered fields. It can be proved that neither the field of complex numbers $(\mathbb{C}; +, \cdot, 0, 1)$ nor any finite field, e.g., the fields $(\mathbb{Z}_p; +, \cdot, 0, 1)$ of remainders modulo a prime p, can be turned into an ordered field by any ordering relation $<$.

An ordered field $(F; +, \cdot, 0, 1, <)$ is called *real closed* if it is a formally real field satisfying the condition

$$x \geq 0 \;\Rightarrow\; (\exists\, z)\big(x = z^2\big)$$

and the infinite list of conditions stating that every polynomial of *odd* degree $n \geq 3$ with coefficients from F has at least one root in F, similarly as in the case of (unordered) real closed fields. The *Theory of Ordered Real Closed Fields* is obtained as the extension of the *Theory of Ordered Fields* by these axioms. It can be easily

verified that every real closed field $(F;+,\cdot,0,1)$ can be turned into an ordered field $(F;+,\cdot,0,1,<)$ by defining the (non-strict) order relation on it as follows:

$$x \le y \;\Leftrightarrow\; (\exists\, z)\big(y = x + z^2\big)$$

The other way round, for every ordered real closed field $(F;+,\cdot,0,1,<)$, its reduct $(F;+,\cdot,0,1)$, obtained by omitting the order relation $<$, is a real closed field.

4.3.8 Peano Arithmetic. *Peano Arithmetic* PA is the most common first-order theory describing the structure of natural numbers. In logical texts it is usually formulated in the first-order language with a unary operation symbol S (successor operation, i.e., adding 1), two binary operations of addition $+$ and multiplication $\cdot$, and a constant symbol 0. However, we find it more convenient to replace the successor symbol S by the constant symbol 1; that way our formulation of PA will use the familiar language of the Theory of Unitary Rings (then the successor operation can be defined by $S(x) = x + 1$).

The axioms of PA can be divided into four groups. The first group consists of three axioms for the successor operation (the left column), the second group is in fact the recursive definition of addition in terms of successor (the middle column), and the third group is the recursive definition of multiplication in terms of addition (the right column):

$$\begin{array}{lll} 0 + 1 = 1 & x + 0 = x & x \cdot 0 = 0 \\ 0 \ne x + 1 & x + (y + 1) = (x + y) + 1 & x \cdot (y + 1) = (x \cdot y) + x \\ x + 1 = y + 1 \;\Rightarrow\; x = y & & \end{array}$$

The fourth group consists of infinitely many axioms comprised in the *Scheme of Induction*

$$\big(\varphi(0,\vec{u}) \wedge (\forall\, x)(\varphi(x,\vec{u}) \;\Rightarrow\; \varphi(x+1,\vec{u}))\big) \;\Rightarrow\; (\forall\, x)\varphi(x,\vec{u})$$

where $\varphi(x,u_1,\ldots,u_n)$ is any formula in the language of PA, abbreviated to $\varphi(x,\vec{u})$.

The language of PA is usually extended by the ordering symbol $\le$ defined by

$$x \le y \;\Leftrightarrow\; (\exists\, z)(y = x + z)$$

and the axioms of PA imply that this is a (non-strict) linear order with the least element 0. Then the Scheme of Induction can be equivalently expressed in form of the *Well Ordering Principle*

$$(\exists\, x)\psi(x,\vec{u}) \;\Rightarrow\; (\exists\, x)\big(\psi(x,\vec{u}) \wedge (\forall\, y)(\psi(y,\vec{u}) \;\Rightarrow\; x \le y)\big)$$

for any formula $\psi(x,\vec{u})$ in the language of PA. Informally, this principle expresses the condition that every nonempty set of the form $\{x\colon \psi(x,\vec{u})\}$ has the least element.

The "usual" natural numbers form the so called *standard model* $(\mathbb{N};+,\cdot,0,1)$ of PA. Every element $n \in \mathbb{N}$ is the interpretation of some constant term in the language of PA. The canonical representatives of particular natural numbers are defined recursively as follows:

1° The natural number 0 coincides with the constant symbol 0.

2° If the natural number n coincides with the constant term t, then the natural number $n+1$ coincides with the constant term $t+1$.

By abuse of notation we can write $0 = 0$, $1 = 0{+}1$, $2 = (0{+}1){+}1$, $3 = ((0{+}1){+}1){+}1$, $\ldots$, $n = (\ldots((0+1)+1)\cdots+1)+1$ (with n instances of 1), etc. Thus the natural number n is represented by the constant term obtained as the n^{th} iterate of the successor operation applied to the constant symbol 0.

Later on we shall see that PA has some nonstandard models, as well.

4.3.9 Set Theory. Most versions of *Set Theory* are formulated in the first-order language with a single binary relational symbol $\in$ denoting the membership relation. The formula $x \in X$ means that x *is an element (a member) of* X or that x *belongs to the set* X. The common core of these versions consists of the following four axioms:

$$\begin{aligned}
&X = Y \;\Leftrightarrow\; (\forall z)(z \in X \;\Leftrightarrow\; z \in Y)\\
&(\forall x, y)(\exists Z)(\forall z)(z \in Z \;\Leftrightarrow\; (z = x \vee z = y))\\
&(\forall X)(\exists U)(\forall u)(u \in U \;\Leftrightarrow\; (\exists x \in X)(u \in x))\\
&(\forall X)(\exists Y)(\forall y)(y \in Y \;\Leftrightarrow\; (\forall x)(x \in y \;\Rightarrow\; x \in X))
\end{aligned}$$

called the *Axiom of Extensionality*, the *Axiom of Pair*, the *Axiom of Union* and the *Power Set Axiom*, respectively, and of the following infinite list of axioms

$$(\forall X)(\exists Y)(\forall x)\big(x \in Y \;\Leftrightarrow\; (x \in X \wedge \varphi(x, \vec{u}))\big)$$

for any set-theoretical formula $\varphi(x, u_1, \ldots, u_n)$, called the *Scheme of Comprehension*.

The Axiom of Extensionality states that two sets X and Y are equal if and only if they contain the same elements. The Axiom of Pair postulates the existence of the set

$$Z = \{x, y\} = \{z\colon z = x \vee z = y\}$$

for any pair of elements x, y. The Axiom of Union guarantees the existence of the union

$$U = \bigcup X = \{u\colon (\exists x \in X)(u \in x)\}$$

of all the sets x from a given set X. The Power Set Axiom postulates the existence of the power set (i.e., the set of all subsets)

$$Y = \mathcal{P}(X) = \{y\colon y \subseteq X\} = \{y\colon (\forall x)(x \in y \;\Rightarrow\; x \in X)\}$$

of any set X. Finally, the Scheme of Comprehension guarantees the possibility to single out every subset of the form

$$Y = \{x \in X\colon \varphi(x, \vec{u})\} = \{x\colon x \in X \wedge \varphi(x, \vec{u})\}$$

from a given set X by means of any set-theoretical formula $\varphi(x, u_1, \ldots, u_n)$.

The *Zermelo-Fraenkel Set Theory* ZF is obtained by adding to this list the *Axiom of Foundation*, the *Axiom of Infinity* and the *Scheme of Replacement*. The *Zermelo-Fraenkel Set Theory with Choice* ZFC, which is the most common version of Set

Theory used in modern mathematics, is obtained from ZF by adding to it the *Axiom of Choice* (AC). We do not include the formulation of these higher axioms of Set Theory into our elementary text. Let us just confine to the following four informal formulations. The Axiom of Foundation states that every set of sets contains as an element a set disjoint from this set. The Axiom of Infinity postulates the existence of an infinite set and, as a consequence, it makes possible to prove the existence of the set of all natural numbers. The Scheme of Replacement generalizes the Scheme of Comprehension by enabling to form sets not just by singling them out from a given set but also as images of subsets of a given set defined by set-theoretical formulas. Finally, the Axiom of Choice guarantees, for every set X of pairwise disjoint nonempty sets, the existence of a set containing exactly one element from each of the sets $x \in X$.

4.4 Axiomatization of First-Order Logic and the Soundness Theorem

Similarly as in Propositional Calculus, we prefer to have a brief and concise axiomatization of First-Order Logic. Therefore we will proceed as if the set $\mathrm{Form}(L)$ of all formulas of any first-order language L were built of the atomic formulas by means of the logical connectives $\neg$ and $\Rightarrow$, and the universal quantifier $\forall$, only. By $\varphi \approx \psi$ we express that the characters φ and ψ denote the same formula. Again, the symbol $\approx$ does not belong to our first-order language, similarly as the symbols φ, ψ, χ, L, etc. They are symbols of our *metalanguage* by means of which we describe Predicate Calculus.

The remaining logical connectives and the existential quantifier can be introduced as the abbreviations

$$\begin{aligned}(\varphi \wedge \psi) &\approx \neg(\varphi \Rightarrow \neg\psi),\\ (\varphi \vee \psi) &\approx (\neg\varphi \Rightarrow \psi),\\ (\varphi \Leftrightarrow \psi) &\approx (\varphi \Rightarrow \psi) \wedge (\psi \Rightarrow \varphi),\\ (\exists x)\varphi &\approx \neg(\forall x)\neg\varphi.\end{aligned}$$

Logical axioms of Predicate Calculus are divided into three groups: *propositional axioms*, *quantifier axioms* and *axioms of equality*.

4.4.1 Propositional axioms. (4 axiom schemes)
For any formulas φ, ψ, χ the following formulas are axioms:

(PrAx 1) $\varphi \Rightarrow (\psi \Rightarrow \varphi)$
(PrAx 2) $(\varphi \Rightarrow (\psi \Rightarrow \chi)) \Rightarrow ((\varphi \Rightarrow \psi) \Rightarrow (\varphi \Rightarrow \chi))$
(PrAx 3) $(\varphi \Rightarrow \psi) \Rightarrow ((\varphi \Rightarrow \neg\psi) \Rightarrow \neg\varphi)$
(PrAx 4) $\neg\neg\varphi \Rightarrow \varphi$

If φ is a formula and t is a term, then $\varphi(t/x)$ denotes the formula obtained by the *substitution* of the term t for the variable x in the formula φ. It means that every *free* occurrence of the variable x in φ is replaced by t. Similarly we can introduce multiple substitutions $\varphi(t_1/x_1, \ldots, t_n/x_n)$. If there's no danger of confusion then we write just $\varphi(t)$ and $\varphi(t_1, \ldots, t_n)$.

The substitution of t for x in φ is *admissible* if no variable of the term t falls under the range of some quantifier in φ after substituting t for x in φ. Informally this means that "$\varphi(t/x)$ is telling of t the same thing as φ is telling of x".

4.4.2 Example. Within the integers the formula $\varphi(x) \approx (\exists\, y)(x = y+y)$ tells that x is an even number. If t is the term $u+x$ then $\varphi(t/x)$ is the formula $(\exists\, y)(u+x = y+y)$, telling that $u + x$ is even; this is an admissible substitution. If t is the term y then $\varphi(t/x)$ is the sentence $(\exists\, y)(y = y + y)$ expressing the (true) fact that the equation $y = y + y$ has some solution; this substitution is not admissible.

4.4.3 Quantifier axioms. (2 axiom schemes)
For any formulas φ, ψ and any term t the following formulas are axioms:

(QAx 1) $(\forall\, x)(\varphi \Rightarrow \psi) \Rightarrow (\varphi \Rightarrow (\forall\, x)\psi)$
(whenever the variable x has no free occurrence in φ)

(QAx 2) $(\forall\, x)\varphi \Rightarrow \varphi(t/x)$
(whenever the substitution of t for x in φ is admissible)

4.4.4 Axioms of equality. (3 axioms + 2 axiom schemes)
For any n-ary functional symbol f and any n-ary relational symbol r the following formulas are axioms:

(EAx 1) $x = x$

(EAx 2) $x = y \Rightarrow y = x$

(EAx 3) $x = y \Rightarrow (y = z \Rightarrow x = z)$

(EAx 4) $x_1 = y_1 \Rightarrow (\dots \Rightarrow (x_n = y_n \Rightarrow f(x_1, \dots, x_n) = f(y_1, \dots, y_n)) \dots)$

(EAx 5) $x_1 = y_1 \Rightarrow (\dots \Rightarrow (x_n = y_n \Rightarrow r(x_1, \dots, x_n) \Rightarrow r(y_1, \dots, y_n)) \dots)$

4.4.5 Deduction rules: Modus Ponens (MP) and
Rule of Generalization (Gen)

(MP) $\dfrac{\varphi,\ \varphi \Rightarrow \psi}{\psi}$ (from φ and $\varphi \Rightarrow \psi$ infer ψ)

(Gen) $\dfrac{\varphi}{(\forall\, x)\varphi}$ (from φ infer $(\forall\, x)\varphi$)

4.4.6 Exercise. (a) Show that all the logical axioms are satisfied in every L-structure $\mathcal{A}$.

(b) Show that the deduction rule Modus Ponens is correct in the following sense: For every L-structure $\mathcal{A}$ and any L-formulas φ, ψ, if $\mathcal{A} \vDash \varphi$ and $\mathcal{A} \vDash \varphi \Rightarrow \psi$ then $\mathcal{A} \vDash \psi$.

(c) Show that the Rule of Generalization is correct in the following sense: For every L-structure $\mathcal{A}$ and any L-formula φ, if $\mathcal{A} \vDash \varphi$ then $\mathcal{A} \vDash (\forall\, x)\varphi$ for any variable x, no matter whether x is free in φ or not.

A *proof* in a first-order theory T is a finite sequence $\varphi_0, \varphi_1, \ldots, \varphi_n$ of formulas such that each item φ_k is either a logical axiom or a specific axiom of the theory T (i.e., $\varphi_k \in T$), or it follows from the previous items by Modus Ponens (MP) (i.e., there are $i, j < k$ such that φ_j has the form $\varphi_i \Rightarrow \varphi_k$) or by the Rule of generalization (Gen) (i.e., there is a $j < k$ such that φ_k has the form $(\forall x)\varphi_j$ for some variable x).

A formula ψ is *provable* in a theory T if there is a proof $\varphi_0, \varphi_1, \ldots, \varphi_n$ in T such that its last item φ_n coincides with ψ. In symbols, $T \vdash \psi$. Instead of $\emptyset \vdash \psi$ we write just $\vdash \psi$; it means that ψ is provable from the logical axioms, only.

4.4.7 Exercise. Show that the following first-order schemes are provable just from the logical axioms:

(a) all propositional tautologies

(b) $\varphi(t/x) \Rightarrow (\exists x)\varphi$ (if the substitution t/x in φ is admissible)

(c) $(x_1 = y_1 \wedge \ldots \wedge x_n = y_n) \Rightarrow t(x_1, \ldots, x_n) = t(y_1, \ldots, y_n)$ (for any term t)

(d) $(x_1 = y_1 \wedge \ldots \wedge x_n = y_n) \Rightarrow (\varphi(x_1, \ldots, x_n) \Leftrightarrow \varphi(y_1, \ldots, y_n))$
(for any formula φ such that all the substitutions y_i/x_i in φ are admissible)

4.4.8 Soundness Theorem. *Let T be a theory in a first-order language L. Then, for every L-formula ψ, if $T \vdash \psi$ then $T \vDash \psi$.*

Demonstration. Let $T \vdash \psi$ and $\varphi_0, \varphi_1, \ldots, \varphi_n$ be a proof of ψ in T. We will show that $\mathcal{A} \vDash \varphi_k$, for any model $\mathcal{A} \vDash T$ of the theory T and *each* $k \leq n$. Then, of course, $\mathcal{A} \vDash \psi$, since ψ is φ_n. Each φ_k is either a logical axiom, in which case $\mathcal{A} \vDash \varphi_k$ for every L-structure $\mathcal{A}$, or a specific axiom of T, in which case $\mathcal{A} \vDash \varphi_k$ as $\mathcal{A} \vDash T$, or φ_k follows from some previous proof items by (MP) or by (Gen). In the (MP) case there are $i, j < k$ such that φ_j has the form $\varphi_i \Rightarrow \varphi_k$. Now, for any L-structure $\mathcal{A}$, assuming that we already have $\mathcal{A} \vDash \varphi_i$ and $\mathcal{A} \vDash \varphi_j$, i.e., $\mathcal{A} \vDash \varphi_i \Rightarrow \varphi_k$, we can conclude $\mathcal{A} \vDash \varphi_k$. In the (Gen) case there is a $j < k$ such that φ_k has the form $(\forall x)\varphi_j$. Again, for any L-structure $\mathcal{A}$, assuming that we already have $\mathcal{A} \vDash \varphi_j$, we can conclude $\mathcal{A} \vDash (\forall x)\varphi_j$, i.e., $\mathcal{A} \vDash \varphi_k$. (Cf. Exercise 4.4.6.)

At this moment, the reader should return to Remark 3.6.2 following the demonstration of the Soundness Theorem 3.6.1 in Propositional Calculus and realize that the accounts stated there equally apply to its first-order version.

Later on we will also establish the converse of the Soundness Theorem.

4.4.9 Completeness Theorem. *Let T be a theory in a first-order language L. Then, for every L-formula ψ, if $T \vDash \psi$ then $T \vdash \psi$.*

As it follows from the following example, the fine balance between the syntax and semantics which we have both in the Propositional as well as in the First-Order Logic is no way self-evident or automatic.

4.4.10 Example. (Finite Model Semantics) Let L be a first-order language. For any theory T in L and any L-formula φ we define the *finite satisfaction relation* $T \vDash_{\text{fin}} \varphi$ if and only if $\mathcal{A} \vDash \varphi$ for every *finite* model $\mathcal{A}$ of the theory T, i.e., if and only if φ is satisfied in every *finite* model of the theory T. By methods going beyond the scope

of our elementary course it can be shown that this *Finite Model Semantics* cannot be axiomatized in a way enabling to establish both the corresponding versions of the Soundness Theorem and of the Completeness Theorem. More precisely, for any sound axiomatization, consisting of finitely many axiom schemes and finitely many rules of inference, the resulting provability relation $T \vdash_{\text{fin}} \varphi$ does not exhaust the finite satisfaction relation $T \vDash_{\text{fin}} \varphi$. This is to say that it is possible to find a theory T and a sentence φ such that $T \vDash_{\text{fin}} \varphi$, nevertheless $T \nvdash_{\text{fin}} \varphi$.

To convey to the reader at least some feeling of the issue, let us mention the deep *Wedderburn's Theorem*, stating that every finite division ring is a field. In other words, the commutative law $xy = yx$ is finitely satisfied in the theory of division rings. On the other hand, the infinite division ring of all quaternions $(\mathbb{H}; +, \cdot, 0, 1)$ is non commutative; therefore, the commutative law for multiplication is not a first-order consequence of the axioms for division rings.

4.4.11 Exercise. We say that an L-formula φ is *logically valid* if it is satisfied in every L-structure $\mathcal{A}$. Two L-formulas φ, ψ are called *logically equivalent*, notation $\varphi \equiv \psi$, if the formula $\varphi \Leftrightarrow \psi$ is logically valid. A formula φ is said to be in *prenex normal form* if it has the shape $(\mathrm{Q}_1 x_1) \dots (\mathrm{Q}_n x_n)\psi$ where $\mathrm{Q}_1, \dots, \mathrm{Q}_n$ are arbitrary quantifiers and ψ is a quantifier-free formula.

(a) Show that, for any formulas φ, ψ, the following pairs of formulas are logically equivalent:

$$\begin{aligned} \neg(\forall x)\varphi &\equiv (\exists x)\neg\varphi & \qquad \neg(\exists x)\varphi &\equiv (\forall x)\neg\varphi \\ (\forall x)\varphi \wedge \psi &\equiv (\forall x)(\varphi \wedge \psi) & \qquad (\exists x)\varphi \vee \psi &\equiv (\exists x)(\varphi \vee \psi) \\ (\forall x)\varphi \Rightarrow \psi &\equiv (\exists x)(\varphi \Rightarrow \psi) & \qquad \psi \Rightarrow (\exists x)\varphi &\equiv (\exists x)(\psi \Rightarrow \varphi) \end{aligned}$$

(b) Show that, for any formulas φ, ψ such that the variable x has no free occurrence in ψ, the following pairs of formulas are logically equivalent:

$$\begin{aligned} (\forall x)\varphi \vee \psi &\equiv (\forall x)(\varphi \vee \psi) & \qquad (\exists x)\varphi \wedge \psi &\equiv (\exists x)(\varphi \wedge \psi) \\ (\exists x)\varphi \Rightarrow \psi &\equiv (\forall x)(\varphi \Rightarrow \psi) & \qquad \psi \Rightarrow (\forall x)\varphi &\equiv (\forall x)(\psi \Rightarrow \varphi) \end{aligned}$$

(c) Using (a) and (b) show that every L-formula is logically equivalent to some formula in prenex normal form.

(d) Replace in any case in (a) and (b) the logical equivalence $\theta \equiv \chi$ by the formula $\theta \Leftrightarrow \chi$ and show that all the formulas thus obtained are provable from the logical axioms (in fact just from the propositional axioms and the quantifier axioms). In each case in (b) decide which of the implications $\theta \Rightarrow \chi$, $\chi \Rightarrow \theta$ remain logically valid even without the assumption that x has no free occurrence in ψ, and give examples showing that in the remaining cases this assumption cannot be omitted.

4.5 The Deduction Theorem and its Corollaries

On the way to the demonstration of the Completeness Theorem we are going to state several results which are of independent interest in their own right. The first group consists of the *Deduction Theorem* and its two corollaries, namely the *Corollary*

on Proof by Contradiction and the *Corollary on Proof by Distinct Cases*, which are analogous to their propositional counterparts.

4.5.1 Deduction Theorem. *Let T be a theory in a first-order language L and φ, ψ be any L-formulas. If φ is closed then $T \vdash \varphi \Rightarrow \psi$ if and only if $T \cup \{\varphi\} \vdash \psi$.*

The easy fact that from $T \vdash \varphi \Rightarrow \psi$ there follows $T \cup \{\varphi\} \vdash \psi$ can be established in exactly the same way as in the demonstration of the Deduction Theorem 3.7.1 in Propositional Calculus, even without the assumption that φ is closed. It is the converse which is needed for the justification of the usual way of argumentation when instead of proving the implication $\varphi \Rightarrow \psi$ in T we prove ψ in $T \cup \{\varphi\}$. This can be established in a similar, just a little bit more complicated way, as the demonstration of the corresponding statement in Propositional Calculus, again. One just has to deal additionally with the case when ψ follows from some preceding item of its proof in $T \cup \{\varphi\}$ by the Rule of Generalization (Gen). Let us fill in this gap, leaving the details to the reader.

Demonstration. Assume that $\psi_0, \psi_1, \ldots, \psi_n$ is a proof of the formula ψ in the theory $T \cup \{\varphi\}$ and ψ_n follows from some previous formula ψ_k, where $0 \leq k < n$, by (Gen). Then ψ_k is provable in $T \cup \{\varphi\}$ and, by an induction argument, we can assume that the implication $\varphi \Rightarrow \psi_k$ is provable in T. Thus ψ has the form $(\forall x)\psi_k$ for some variable x. Now the following formulas are provable in T:

(1) $\varphi \Rightarrow \psi_k$

(2) $(\forall x)(\varphi \Rightarrow \psi_k)$ follows from (1) by (Gen)

(3) $(\forall x)(\varphi \Rightarrow \psi_k) \Rightarrow (\varphi \Rightarrow (\forall x)\psi_k)$ is an instance of the quantifier axiom scheme (QAx 1), as φ is closed so that x has no free occurrence in it

(4) $(\varphi \Rightarrow (\forall x)\psi_k)$ can be inferred from (2) and (3) by (MP)

Thus, finally, $T \vdash \varphi \Rightarrow \psi$.

Next we show that, in general, one cannot do without the assumption that φ is closed.

4.5.2 Example. Let L be the *language of pure equality*, i.e., the first-order language without any specific symbols ($F = C = R = \emptyset$) and $T = \emptyset$ be the theory without any specific axioms in L. Denote by φ the formula $x = y$ and by ψ the formula $x = z$. We claim that $T \cup \{\varphi\} \vdash \psi$, nevertheless, $T \nvdash \varphi \Rightarrow \psi$. $T \cup \{\varphi\} = \{\varphi\}$ is the theory with a single axiom $x = y$. It means that all the models of this theory have just a one--element base set. Applying (Gen) to this axiom, we see that $T \cup \{\varphi\} \vdash (\forall y)(x = y)$. Now, we have the quantifier axiom $(\forall y)(x = y) \Rightarrow x = z$, and using (MP) we infer that $T \cup \{\varphi\} \vdash x = z$. At the same time, the implication $\varphi \Rightarrow \psi$, i.e., $x = y \Rightarrow x = z$ is not provable just from logical axioms, i.e., in the theory T. Namely, if this were the case, then it would be satisfied in every L-structure $\mathcal{A}$. However, every $\mathcal{A}$ with at least a two-element base set with elements $a \neq b$ violates this implication, since substituting a for both x and y and b for z we have $a = a$, nevertheless $a \neq b$.

A theory T in a first-order language L is called *inconsistent* or *contradictory* if there is some closed L-formula φ such that $T \vdash \varphi$ as well as $T \vdash \neg\varphi$. Otherwise, T is

called *consistent* or *contradiction-free*. It can be easily verified that T is inconsistent if and only if every L-formula is provable in T.

The following results follow from the Deduction Theorem 4.5.1 in the same way as in Propositional Calculus.

4.5.3 Corollary on Proof by Contradiction. *Let T be a theory in a first-order language L and φ be a closed L-formula. Then $T \vdash \varphi$ if and only if the theory $T \cup \{\neg\varphi\}$ is contradictory (inconsistent).*

4.5.4 Corollary on Proof by Distinct Cases. *Let T be a theory in a first-order language L and φ, ψ be any L-formulas. If φ is closed then $T \vdash \psi$ if and only if $T \cup \{\varphi\} \vdash \psi$ and $T \cup \{\neg\varphi\} \vdash \psi$.*

4.5.5 Exercise. Find examples showing that the assumption that φ is closed cannot be omitted from the above Corollaries.

4.6 Complete Theories

Another important property of first-order theories closely related to consistency is that of their completeness. A theory T in a first-order language L is called *complete* if it is consistent and for any *closed* L-formula φ we have $T \vdash \varphi$ or $T \vdash \neg\varphi$. In other words, T is complete if and only if, for every L-sentence φ, either $T \vdash \varphi$ or $T \vdash \neg\varphi$ (but not both). We also use to say that T *can decide* every sentence φ.

4.6.1 Example. The reader may wonder why in the definition of complete theories we have required that T can decide just the closed formulas, ignoring the remaining ones. For the sake of explanation, consider the formula $x = y$ and its negation $x \neq y$. If the provability of one of them were included in the requirement of completeness of a theory T then, in the first case, T would have just one-element models, or, in the second case, it would be contradictory. As a consequence, any consistent theory "complete" in such a sense would have trivial models, only.

Using the Corollary 4.5.3 on Proof by Contradiction, complete theories can be characterized as maximal consistent theories in the following sense:

4.6.2 Corollary on Complete Theories. *Let T be a consistent theory in a first--order language L. Then T is complete if and only if, for every L-sentence φ, either $T \vdash \varphi$ or the theory $T \cup \{\varphi\}$ is contradictory.*

In other words, extending the axiom list of a complete theory T by any sentence φ makes no sense: either φ is already provable in T (in which case the sets of formulas provable in T and $T \cup \{\varphi\}$ coincide) or $T \cup \{\varphi\}$ turns inconsistent hence worthless.

Most of the relevant first-order theories occurring in mathematics are not complete. On the other hand, many of them have important complete extensions. In this place we just mention some examples of complete theories without proving their completeness.

4.6.3 Divisible Abelian Groups. The Theory of Groups is not complete: for instance, the fact that there exist both abelian as well as nonabelian groups shows that neither the commutativity law $(\forall\, x, y)(xy = yx)$ nor its negation can be proved in it. Here we describe some relatively simple complete extensions of the Theory of Abelian Groups.

An abelian group $\mathcal{G} = (G;\, +, 0)$ is called *nontrivial* if $(\exists\, x)(x \neq 0)$ holds in $\mathcal{G}$; it is called *divisible* if it is nontrivial and satisfies the condition

$$(\forall\, x)(\exists\, y)(n \times y = x)$$

for every integer $n \geq 2$, where $n \times x = x + \ldots + x$ with n-fold occurrence of x. $\mathcal{G}$ is called *torsion-free* if it satisfies all the conditions

$$n \times x = 0 \;\Rightarrow\; x = 0$$

for $n \geq 2$. Given a fixed $n \geq 1$, we say that $\mathcal{G}$ is a *group of exponent* n if it satisfies $(\forall\, x)(n \times x = 0)$. It is known that the *Theory of Divisible Torsion-Free Abelian Groups*, as well as every *Theory of Divisible Abelian Groups of Exponent* p, for a fixed prime number p, is complete.

4.6.4 Real Closed and Algebraically Closed Fields. It can be shown that the Theory of Real Closed Fields (both in its unordered as well as in its ordered version) is complete. On the other hand, the Theory of Algebraically Closed Fields is not complete. Nonetheless, its complete extensions can be fully described.

The *characteristic* of a unitary ring $\mathcal{A} = (A;\, +, \cdot, 0, 1)$ is the least integer $\mathrm{char}(\mathcal{A}) = n \geq 1$ such that $n \times 1 = 0$, or $\mathrm{char}(\mathcal{A}) = \infty$ if $n \times 1 \neq 0$ for each $n \geq 1$ (some authors put $\mathrm{char}(\mathcal{A}) = 0$ in this case). It is known that the characteristic of any field is either a prime or ∞. A field $(F,\, +, \cdot, 0, 1)$ has the prime characteristic p if and only if it satisfies $p \times 1 = 0$, it has the characteristic ∞ if and only if it satisfies all the conditions $p \times 1 \neq 0$ for every prime number p. Every Theory of Algebraically Closed Fields of a fixed prime Characteristic p, as well as the Theory of Algebraically Closed Fields of Characteristic ∞ is complete.

4.6.5 Dense Linear Order. A linearly ordered set $(A;\, <)$ is called *dense* if it contains at least two elements and satisfies the condition

$$(\forall\, x, y)\big(x < y \;\Rightarrow\; (\exists\, z)(x < z < y)\big)$$

$(A;\, <)$ is *without endpoints* if it satisfies the condition

$$(\forall\, x)(\exists\, y, z)(y < x < z)$$

The *Theory of Dense Linear Order without Endpoints* can be proved to be complete. Additionally, three complete extensions of the Theory of Dense Linear Order can obtained by the variation of the condition of the existence of endpoints in the obvious way.

4.6.6 Atomic and Atomless Boolean Algebras. An element $a \in B$ of a Boolean algebra $\mathcal{B} = (B; \wedge, \vee, ', 0, 1)$ is called an *atom* if $a \neq 0$ and there is no element $b \in B$ such that $0 < b < a$. Formally, we can extend the language of Boolean algebras by a new unary predicate $\mathrm{At}(x)$ defined by

$$\mathrm{At}(x) \;\Leftrightarrow\; x \neq 0 \wedge (\forall y)\big(0 \leq y \leq x \;\Rightarrow\; (y = 0 \vee y = x)\big)$$

using the previously defined partial order $\leq$ (see 4.3.5). Then $\mathcal{B}$ is called *atomic* if for every nonzero element of B there is an atom contained in it, i.e., if $\mathcal{B}$ satisfies the condition

$$(\forall x)(x \neq 0 \;\Rightarrow\; (\exists y)(\mathrm{At}(y) \wedge y \leq x))$$

$\mathcal{B}$ is called *atomless* if it has at least two elements and contains no atom. This can be expressed by the non-triviality condition $0 \neq 1$ and a kind of density axiom

$$(\forall x)(x \neq 0 \;\Rightarrow\; (\exists y)(0 < y < x))$$

The *Theory of Atomic Boolean Algebras with Infinitely Many Atoms* as well as every *Theory of Atomic Boolean Algebras with Precisely n Atoms*, for any $n \geq 0$, are complete. Similarly, the *Theory of Atomless Boolean Algebras* is complete, too. Moreover, all the complete extensions of the Theory of Boolean Algebras can be effectively described in terms of a pair of integer invariants. However, this description is fairly involved and goes beyond the scope of our exposition.

4.6.7 Presburger Arithmetic. Later on, when dealing with Gödel's Incompleteness Theorems, we shall see that not only Peano Arithmetic is not complete but also its completions cannot be effectively described. On the other hand, it has an interesting complete subtheory called *Presburger Arithmetic*, describing the structure of natural numbers with the operations of successor and addition, only. Its language contains the constant symbols 0 and 1 and the operation symbol $+$. Its axioms are obtained from the axioms of Peano Arithmetic quoted in 4.3.8 by omitting those containing the symbol of multiplication, i.e., the couple forming the right most column of the seven individual axioms of PA as well as all the instances of the Scheme of Induction where the formula $\varphi(x, \vec{u})$ contains the operation symbol $\cdot$.

4.7 Results on Language Extensions

When proving a universally quantified statement of the form $(\forall x_1, \ldots, x_k)\varphi(x_1, \ldots, x_k)$ in a first-order theory T, we usually begin with the phrase: *"Let x_1, ..., x_n be arbitrary elements..."* That, however, means that we do not consider x_1, ..., x_n in our proof as variables any more, and deal with them as with some unspecified constants. The following result shows that such a kind of argumentation is legitimate in First-Order Logic.

4.7.1 Lemma on Constants. *Let T be a theory in a first-order language L, $\varphi(x_1, \ldots, x_k)$ be an L-formula and $c_1, \ldots, c_k$ be pairwise distinct constant symbols not occurring in L. Then*

$$T \vdash \varphi(c_1, \ldots, c_k) \quad \textit{if and only if} \quad T \vdash (\forall x_1, \ldots, x_k)\varphi(x_1, \ldots, x_k)$$

Demonstration. Assume that $T \vdash \varphi(c_1, \ldots, c_k)$. Realize that T is a theory in the language L not containing the constants $c_1, \ldots, c_k$, so that the theory T "cannot know anything about them". Therefore, everything what can be proved in T about these constants can be proved about conveniently chosen distinct variables $x_1, \ldots, x_k$ not occurring in the original proof — it suffices to replace each occurrence of the symbol symbol c_i by the variable x_i. Then $T \vdash \varphi(x_1, \ldots, x_k)$ and the needed conclusion follows by the Rule of Generalization. The reversed implication is trivial.

It is clear that the above theorem remains true also in case when the constants $c_1, \ldots, c_k$ belong to L but they do not occur in any of the specific axioms of T.

When developing and building a mathematical theory we seldom keep its language fixed for all the time. Just the opposite, we often define new notions, corresponding to some operations, relations or distinguished elements, and introduce new symbols for them. These new symbols, as a rule, denote important or frequently occurring concepts, abbreviate otherwise cumbersome formulations and that way contribute to transparency and intelligibility of the theory. Even in our course we already did so several times in the parts devoted to examples of various first-order theories, without paying special attention to this point. For instance, we extended the language of Group Theory consisting of a single binary operation symbol $\cdot$ by the constant symbol e for the unit element and the unary operation symbol $^{-1}$ for taking inverses. We also extended both the language of the Theory of Boolean Algebras as well as the language of Peano Arithmetic by the order relation symbol $\leq$, etc. Now, we will treat this situation in general.

Let $L = (F, C, R, \nu)$ and $L' = (F', C', R', \nu')$ be two first-order languages. We say that the language L' is an *extension* of the language L if $F \subseteq F'$, $C \subseteq C'$, $R \subseteq R'$ and for each operational or relational symbol $s \in F \cup R$ we have $\nu'(s) = \nu(s)$, i.e., the arities of the symbol s in L and L' coincide. Is it the case, we write $L \subseteq L'$ or $L' \supseteq L$. Then any first-order theory T in the language L can be considered as a theory in the language L'. The other way round, from any L'-structure $\mathcal{A} = (A; I)$ one can obtain an L-structure $\mathcal{A} \restriction L = (A; I \restriction L)$, called the *restriction* of $\mathcal{A}$ to L, by leaving its base set A and the interpretations s^I of all the symbols of L unchanged and omitting the interpretations of the remaining symbols of L'. We are particularly interested in the case when the new symbols extending L are introduced by means of definitions by formulas of the language L.

The *unique existence quantification* $(\exists!\, x)\varphi$ is introduced as the abbreviation for

$$(\exists\, x)\big(\varphi \wedge (\forall\, y)(\varphi(y/x) \Rightarrow y = x)\big)$$

where y is any variable not occurring in φ.

Let T be a theory in a first-order language L. Dealing with constant, functional and relational symbols we distinguish three possibilities:

(a) Let $\varphi(x)$ be an L-formula such that $T \vdash (\exists!\, x)\varphi(x)$. We extend the language L by a new constant symbol d not occurring in L and the theory T by the axiom

$$x = d \;\Leftrightarrow\; \varphi(x)$$

(b) Let $\psi(x_1, \ldots, x_n, y)$ be an L-formula such that $T \vdash (\forall\, x_1, \ldots, x_n)(\exists!\, y)\psi(\vec{x}, y)$. We extend the language L by a new n-ary functional symbol g not occurring in

L and the theory T by the axiom

$$y = g(x_1, \ldots, x_n) \Leftrightarrow \psi(x_1, \ldots, x_n, y)$$

(c) Let $\rho(x_1, \ldots, x_n)$ be any L-formula. We extend the language L by a new n-ary relational symbol q not occurring in L and the theory T by the axiom

$$q(x_1, \ldots, x_n) \Leftrightarrow \rho(x_1, \ldots, x_n)$$

In any of the above cases we say that the constant symbol d or the functional symbol g or the relational symbol q, respectively, were introduced by the corresponding definition in T. The reader should realize that, regarding constant symbols as nullary functional symbols, (a) can be considered as a special case of (b).

We say that a theory T' in a first-order language L' is an *extension* of the theory T in the first-order language L *by definitions* if L' is an extension of L by finitely many specific symbols and the specific axioms of T' are obtained extending T by consecutive introduction of the new symbols of L' by definitions. Thus introducing a new symbol at some step we can use not just the means of the original language L in its definition but also the previously introduced symbols. Now it is clear that every model $\mathcal{A} = (A; I)$ of the theory T in the language L has a unique extension to a model $\mathcal{A}' = (A; I')$ of T' in the language L'. It is obtained by repeated interpretation of each newly introduced symbol in $\mathcal{A}'$ using its defining formula in the language L extended by the previously introduced symbols.

Although the new theory T' enables to express several concepts in a more concise and readable way, concerning statements in the original language L, it cannot prove more then the original theory T.

A theory T' in a first-order language L' extending a theory T in a first-order language $L \subseteq L'$ is called a *conservative extension* of T if for any closed L-formula φ we have

$$T' \vdash \varphi \quad \text{if and only if} \quad T \vdash \varphi$$

Obviously, every conservative extension of a consistent theory is itself consistent.

4.7.2 Theorem on Extension by Definitions. *Assume that the theory T' in a first-order language L' is obtained as an extension by definitions of a theory T in a first-order language $L \subseteq L'$. Then T' is a conservative extension of T.*

In order to demonstrate Theorem 4.7.2 it would be enough to deal with the case when L' and T' are obtained from L and T by introducing a single defined symbol. The idea of the proof is simple: it consists in replacing every instance the defined formula $x = d$, $y = g(x_1, \ldots, x_n)$ or $q(x_1, \ldots, x_n)$, respectively, by an appropriate instance of the corresponding L-formula defining it. However, its realization would require to take care of some technical details which we skip as they would hardly contribute to reader's understanding the issue.

4.7.3 Exercise. Extensions by defined constants, operations or relation are fairly frequent in Set Theory.

(a) Write explicitly the defining formulas for the empty set constant $\emptyset$, the operations of the unordered pair of elements $\{x, y\}$, of the union $\bigcup X$, as well as for all the operations of taking the subset $\{x \in X \colon \varphi(x, \vec{u})\}$ from the Scheme of Comprehension.

(b) Using appropriate instances of the Scheme of Comprehension introduce the binary operations of intersection $X \cap Y = \{u \colon u \in X \wedge u \in Y\}$ and set-theoretical difference $X \setminus Y = \{u \colon u \in X \wedge u \notin Y\}$.

(c) Write the defining formula for the subset relation $X \subseteq Y$ and, using it, introduce the power set operation $\mathcal{P}(X) = \{Y \colon Y \subseteq X\}$.

(d) Using the operations of unordered pair $\{x, y\}$ and union $\bigcup X$, introduce the binary operation of union $X \cup Y = \{u \colon u \in X \vee u \in Y\}$.

(e) Using the operation of unordered pair, introduce the operation of ordered pair as $(x, y) = \{\{x\}, \{x, y\}\}$ and prove that

$$(x, y) = (u, v) \Leftrightarrow x = u \wedge y = v$$

(f) Using the operations of ordered pair (x, y), binary union $X \cup Y$ and power set $\mathcal{P}(X)$, as well as an appropriate instance of the Scheme of Comprehension, introduce and justify the operation of cartesian product

$$X \times Y = \{(x, y) \colon x \in X \wedge y \in Y\}$$

4.8 Gödel's Completeness Theorem

Assume that $L = (F, C, R, \nu)$ is a first-order language containing at least one constant symbol (i.e., $C \neq \emptyset$). Denote by K the set of all constant terms of L. Then K becomes the base set of an L-structure $\mathcal{K} = (K; \dots)$, obtained by interpreting the specific symbols of L in the following natural way:

(a) for any n-ary functional symbol $f \in F$ and constant terms $t_1, \dots, t_n \in K$, $f^{\mathcal{K}}(t_1, \dots, t_n)$ is the constant term $f(t_1, \dots, t_n) \in K$;

(b) for any constant symbol $c \in C$, $c^{\mathcal{K}}$ is the constant term $c \in K$;

(c) for any n-ary relational symbol $r \in R$ and constant terms $t_1, \dots, t_n \in K$, we put $(t_1, \dots, t_n) \in r^{\mathcal{K}}$ if and only if $T \vdash r(t_1, \dots, t_n)$.

Additionally, we introduce the following binary relation $\sim_T$ on K:

$$t_1 \sim_T t_2 \Leftrightarrow T \vdash t_1 = t_2$$

for $t_1, t_2 \in K$. If there's no danger of confusion, we write just $\sim$ instead of $\sim_T$.

4.8.1 Exercise. Using the Axioms of Equality show that

$$\begin{aligned} t &\sim t \\ t_1 \sim t_2 &\Rightarrow t_2 \sim t_1 \\ (t_1 \sim t_2 \wedge t_2 \sim t_3) &\Rightarrow t_1 \sim t_3 \\ (t_1 \sim s_1 \wedge \dots \wedge t_n \sim s_n) &\Rightarrow f(t_1, \dots, t_n) \sim f(s_1, \dots, s_n) \\ (t_1 \sim s_1 \wedge \dots \wedge t_n \sim s_n) &\Rightarrow (r(t_1, \dots, t_n) \Leftrightarrow r(s_1, \dots, s_n)) \end{aligned}$$

for any $t, t_1, t_2, t_3 \in K$, all n-ary symbols $f \in F$, $r \in R$ and any $t_1, s_1, \dots, t_n, s_n \in K$.

The first three conditions express the fact that $\sim$ is reflexive, symmetric and transitive, i.e., it is an *equivalence relation* on the set K. For each $t \in K$ we denote by

$$\widetilde{t} = \{s \in K \colon s \sim t\}$$

the set of all constant terms equivalent with t, i.e., all such $s \in K$ for which the equality $s = t$ can be proved in T. We always have $t \in \widetilde{t}$, $\widetilde{t} = \widetilde{s}$ if and only if $t \sim s$, and $\widetilde{t} \cap \widetilde{s} = \emptyset$ if $t \not\sim s$. Thus we can form the *quotient set*

$$K/\!\sim \;= \{\widetilde{t} \colon t \in K\}$$

i.e., the *partition* of K into blocks of pairwise equivalent elements. Alternatively, $K/\!\sim$ can be viewed as the result of identifying or merging into a single element each block of pairwise equivalent elements of K. In other words, the equivalence relation $\sim$ is considered as a "new equality" on K. The last two *compatibility conditions* express the fact that both the operations and the relations in $\mathcal{K}$ respect the equivalence relation, i.e., the "new equality" $\sim$.

The quotient $M = K/\!\sim$ becomes the base set of an L-structure $\mathcal{M} = (M; \dots)$, again, obtained by interpreting the specific symbols of L in the following natural way:

(a) for any n-ary operation symbol $f \in F$ and equivalence blocks $\widetilde{t}_1, \dots, \widetilde{t}_n \in M$, $f^{\mathcal{M}}(\widetilde{t}_1, \dots, \widetilde{t}_n)$ is the equivalence block $f(t_1, \dots, t_n)\widetilde{\ } \in M$ of the constant term $f(t_1, \dots, t_n) \in K$;

(b) for any constant symbol $c \in C$, $c^{\mathcal{M}}$ is the equivalence block $\widetilde{c} \in M$ of the constant term $c \in K$;

(c) for any n-ary relational symbol $r \in R$ and equivalence blocks $\widetilde{t}_1, \dots, \widetilde{t}_n \in M$, we put $(\widetilde{t}_1, \dots, \widetilde{t}_n) \in r^{\mathcal{M}}$ if and only if $T \vdash r(t_1, \dots, t_n)$.

The compatibility conditions for $\sim$ guarantee that the above definitions of the interpretations $f^{\mathcal{M}}$, $r^{\mathcal{M}}$ of the operation and predicate symbols, respectively, are correct, i.e., they do not depend on the particular representatives of the equivalence blocks $\widetilde{t}_i$. Obviously, the interpretation of any constant term $t \in K$ in the structure $\mathcal{M}$ is the element $\widetilde{t} \in M$, i.e., $t^{\mathcal{M}} = \widetilde{t}$.

In order to stress the role of the theory T in the construction of the structure $\mathcal{M}$, we denote it by $\mathcal{M}(T) = \mathcal{M} = (M; \dots)$ and call it the *canonical structure* of the theory T.

4.8.2 Example. The constant terms in the language of Peano Arithmetic PA are composed of the constant symbols 0 and 1 by means of the operations of addition and multiplication. For instance, 1, $0 + 1$, $1 + 0$, $1 \cdot 1$, $0 + (1 \cdot 1)$ are five different constant terms, however, they all denote the same natural number 1, and, at the same time, the equality between any pair of them is provable in PA. In other words, $1 \sim_{\mathrm{PA}} 0 + 1 \sim_{\mathrm{PA}} 1 + 0 \sim_{\mathrm{PA}} 1 \cdot 1 \sim_{\mathrm{PA}} 0 + (1 \cdot 1)$. In fact, there are infinitely many constant terms t in the language of PA such that $t \sim_{\mathrm{PA}} 1$. Similarly, the natural number 2 denotes the equivalence block of the constant term $1 + 1$ or of any constant term provably equal to it, etc. The reader should realize that the canonical structure $\mathcal{M}(\mathrm{PA})$ of the theory PA coincides with its standard model $(\mathbb{N}; +, \cdot, 0, 1)$.

It would be nice if we could guarantee that $\mathcal{M}(T) \vDash T$, i.e., that the canonical structure $\mathcal{M}(T)$ is a model of T for any consistent first-order theory T (in a language L containing at least one constant symbol). Unfortunately, this is not always the case. Nevertheless, we can prove that $\mathcal{M}(T) \vDash T$ for theories satisfying a couple of conditions, to be formulated below.

The first of these conditions is the completeness of the theory T, similarly as in Propositional Calculus. The second condition has no propositional analogue. Given an L-formula $\varphi(x)$ (with a single free variable x), a constant L-term t is called a *witness* of the sentence $(\exists\, x)\varphi(x)$ in the theory T if

$$T \vdash (\exists\, x)\varphi(x) \Rightarrow \varphi(t)$$

(Notice that the substitution of a constant term for any variable is always admissible.) Since we always have $T \vdash \varphi(t) \Rightarrow (\exists\, x)\varphi(x)$, t is a witness of $(\exists\, x)\varphi(x)$ in T if and only if

$$T \vdash (\exists\, x)\varphi(x) \Leftrightarrow \varphi(t)$$

Is it the case, then we have $T \vdash (\exists\, x)\varphi(x)$ if and only if $T \vdash \varphi(t)$.

A theory T in a first-order language L (containing at least one constant symbol) is called a *Henkin theory* if every L-sentence of the form $(\exists\, x)\varphi(x)$ has a witness in T.

4.8.3 Proposition. *Let T be a complete Henkin theory in a first-order language L (containing at least one constant symbol). Then $\mathcal{M}(T) \vDash T$, in other words, the canonical structure $\mathcal{M}(T)$ of the theory T is a model of T.*

Demonstration. We will show that, for any closed L-formula φ, we have

$$T \vdash \varphi \qquad \text{if and only if} \qquad \mathcal{M}(T) \vDash \varphi \tag{$*$}$$

This already implies the needed conclusion $\mathcal{M}(T) \vDash T$. We will proceed by induction on the complexity of φ.

Every closed atomic L-formula φ has the form $t = s$ or $r(t_1, \ldots, t_n)$ where t, s and $t_1, \ldots, t_n$ are constant terms and r is an n-ary relational symbol. Thus for atomic sentences $(*)$ is true according to the definition of the structure $\mathcal{M}(T)$.

Now, it is enough to perform the induction steps for the logical connectives $\neg$ and $\wedge$, and the existential quantifier $\exists$.

Assuming $(*)$ for φ, we'll verify it for $\neg\varphi$ by showing that the conditions $T \vdash \neg\varphi$ and $\mathcal{M}(T) \vDash \neg\varphi$ are equivalent. $T \vdash \neg\varphi$ implies $T \nvdash \varphi$ since T is consistent; the reversed implication follows from the completeness of T. Thus the conditions $T \vdash \neg\varphi$ and $T \nvdash \varphi$ are equivalent. However, $T \nvdash \varphi$ is equivalent to $\mathcal{M}(T) \nvDash \varphi$ by the inductive assumption, and that is equivalent to $\mathcal{M}(T) \vDash \neg\varphi$.

Assuming $(*)$ for both φ and ψ, we'll verify it for $\varphi \wedge \psi$. Obviously, the following conditions are equivalent: $T \vdash \varphi \wedge \psi$; $T \vdash \varphi$ and $T \vdash \psi$; $\mathcal{M}(T) \vDash \varphi$ and $\mathcal{M}(T) \vDash \psi$; $\mathcal{M}(T) \vDash \varphi \wedge \psi$ (the inductive assumption is needed to ensure the equivalence of the second and the third condition).

Finally, assuming $(*)$ for all the sentences $\varphi(t)$ where t is a constant term, we will verify it for the sentence $(\exists\, x)\varphi(x)$. Since T is a Henkin theory, the sentence $(\exists\, x)\varphi(x)$ has some witness t in T, hence the condition $T \vdash (\exists\, x)\varphi(x)$ is equivalent

to the existence of some constant term t such that $T \vdash \varphi(t)$. By the inductive assumption, this is equivalent to the existence of some constant term t such that $\mathcal{M}(T) \vDash \varphi(t)$, i.e., $\mathcal{M}(T) \vDash \varphi\big(\widetilde{t}\big)$. Since $\mathcal{M}(T) = (M; \dots)$ and $M = K/\sim$ consists entirely of elements of the form $\widetilde{t}$ where t is a constant term, the last condition is equivalent to $\mathcal{M}(T) \vDash (\exists x)\varphi$.

4.8.4 Exercise. Assume that S is a contradictory theory in a first-order language L (containing at least one constant symbol). Describe its canonical structure $\mathcal{M}(S)$ and realize that it is not a model of S. Find complete (hence consistent) theory T in L such that $\mathcal{M}(S) = \mathcal{M}(T) \vDash T$.

Using the *Axiom of Choice* it is possible to prove the following theorem. The interested reader will find its proof in the final Section 4.11. Dealing with a fixed first-order language L, a *new symbol* (no matter whether a constant, functional or relational one) always means a specific symbol not occurring in L.

4.8.5 Theorem on Complete Henkin Extensions. *Let T be a consistent theory in a first-order language $L = (F, C, R, \nu)$. Then there is an extension of L by a set D of new constant symbols to a first-order language $L_D = (F, C \cup D, R, \nu)$ and an extension of T to a complete Henkin theory $\widehat{T} \supseteq T$ in the language L_D.*

We will use the last Theorem in the demonstration of the following result, which is an alternative version of the Completeness Theorem.

4.8.6 Gödel's Completeness Theorem. *Every consistent first-order theory T has some model $\mathcal{A} \vDash T$.*

The reader should realize that also the other way round, if a first-order theory has some model then it must be consistent, in other words, a contradictory first-order theory cannot have any model. (This is the alternative version of the Soundness Theorem.)

Demonstration. Let T be a consistent theory in a first-order language L, the first--order language L_D be an extension of L by certain set D of new constant symbols, and $\widehat{T} \supseteq T$ be a theory in L_D forming a complete Henkin extension of T. According to the last Proposition, the canonical structure $\mathcal{M}\big(\widehat{T}\big)$ of the theory $\widehat{T}$ is a model of $\widehat{T}$, i.e., $\mathcal{M}\big(\widehat{T}\big) \vDash \widehat{T}$. Since $T \subseteq \widehat{T}$, we have $\mathcal{M}\big(\widehat{T}\big) \vDash T$, hence $\mathcal{M}\big(\widehat{T}\big)$ is a model of T, as well.

Those who feel puzzled by the fact that T is a theory in the language L, while $\mathcal{M}\big(\widehat{T}\big)$ is an L_D-structure, can form the restriction $\mathcal{A} = \mathcal{M}\big(\widehat{T}\big) \restriction L$ of the L_D-structure $\mathcal{M}\big(\widehat{T}\big)$ to the language L. Then $\mathcal{A}$ is already an L-structure and, obviously, $\mathcal{A} \vDash T$.

Finally we can prove the original form of the Completeness Theorem. We state it in the form comprising the Soundness Theorem, as well.

4.8.7 Completeness Theorem. *Let T be a theory in a first-order language L. Then, for every L-formula ψ, $T \vDash \psi$ if and only if $T \vdash \psi$.*

Demonstration. If $T \vdash \psi$ then $T \vDash \psi$ by the Soundness Theorem 4.4.8. To show the converse, assume that $T \vDash \psi$, nevertheless $T \nvdash \psi$. Without loss of generality we can assume that ψ is closed. (Otherwise, we can replace ψ by the sentence $(\forall\, x_1, \ldots, x_n)\psi$, which we denote by $\overline{\psi}$, where $x_1, \ldots, x_n$ are all the variables occurring freely in ψ. Then we have $T \vdash \psi$ if and only if $T \vdash \overline{\psi}$, and $T \vDash \psi$ if and only if $T \vDash \overline{\psi}$.) As ψ is closed, from $T \nvdash \psi$ it follows that the theory $T \cup \{\neg\psi\}$ is consistent by Corollary 4.5.3 on Proof by Contradiction. Then, according to Gödel's Completeness Theorem, $T \cup \{\neg\psi\}$ has some model $\mathcal{A}$. Then $\mathcal{A}$ is a model of the theory T such that $\mathcal{A} \vDash \neg\psi$. However, since $T \vDash \psi$, we have $\mathcal{B} \vDash \psi$ for every model $\mathcal{B}$ of T; in particular, $\mathcal{A} \vDash \psi$. This contradiction proves that $T \vdash \psi$.

4.9 The Compactness Theorem

Once we have established Gödel's Completeness Theorem, the first-order version of the Compactness Theorem can be demonstrated as its corollary in essentially the same way as its Propositional Calculus version. We leave it to the reader as an exercise.

4.9.1 Compactness Theorem. *Let T be a theory in a first-order language L. Then T has some model if and only if every finite subtheory T_0 of T has some model.*

However, unless its Predicate Calculus version, the first-order version of the Compactness Theorem has several important consequences. We confine ourselves to just some few examples. At least in some of them the reader should experience the feeling that the Compactness Theorem enables to prove the existence of certain models of some theories almost — if not even literally — out of nothing.

To start with, the reader should realize the following immediate consequence of the Compactness Theorem.

4.9.2 Corollary. *Let T, S be two theories in a first-order language L. Then the theory $T \cup S$ has some model if and only if, for every finite subtheory $U \subseteq S$, the theory $T \cup U$ has some model.*

A first-order theory is said to have *arbitrarily big finite models* if for every natural number $n \geq 1$ there is a finite model $\mathcal{A} = (A; \ldots)$ of the theory T such that $|A| \geq n$.

4.9.3 Theorem. *Let T be a theory in a first-order language L. If T has arbitrarily big finite models, then T has some infinite model, as well.*

Demonstration. For every $n \geq 2$ we denote by σ_n the following sentence in the language of pure equality:

$$(\exists\, x_1, \ldots, x_n)\Big(\bigwedge_{1 \leq i < j \leq n} x_i \neq x_j\Big)$$

Then, for any L-structure $\mathcal{A} = (A; \ldots)$, we have $\mathcal{A} \vDash \sigma_n$ if and only if $|A| \geq n$.

For each $n \geq 2$ we denote by S_n the first-order theory with axioms $\sigma_2, \ldots, \sigma_n$ and by $S = \{\sigma_n \colon 2 \leq n \in \mathbb{N}\}$ the theory formed by all the axioms σ_n. Obviously, a first-order structure $\mathcal{A}$ is infinite if and only if $\mathcal{A} \vDash S$.

Assume that T has arbitrarily big finite models. It follows that each of the theories $T \cup S_n$, where $n \geq 2$, has some model. Then, however, for every finite subtheory $U \subseteq S$, there is some $n \geq 2$ such that $U \subseteq S_n$. Since any model of the theory $T \cup S_n$ is a model of $T \cup U$, the theory $T \cup U$ has some model. By the Compactness Theorem the theory $T \cup S$ has some model $\mathcal{A}$, as well. This $\mathcal{A}$ is an infinite model of T.

4.9.4 Exercise. Let T be a first-order theory in the language of the Theory of unitary and φ be a closed formula in this language.

(a) Show that if T has as its models unitary rings of arbitrarily big finite characteristic then it has as a model also a unitary ring of characteristic ∞.

(b) Show that the characteristic of a field is either a prime or ∞ and prove that if T has as its models fields of arbitrarily big prime characteristic then it has as a model also a field of characteristic ∞.

(c) Show the following *Robinson's Principle*: If the sentence φ is satisfied in every field of characteristic ∞ then there is a prime number p such that φ is satisfied in every field of prime characteristic $q \geq p$.

Another striking consequence of the Compactness Theorem is the existence of nonstandard models of Peano Arithmetic, first noticed by Thoralf Skolem.

4.9.5 Theorem. *Peano Arithmetic has some nonstandard models.*

Demonstration. Let us extend the language of PA by a new constant symbol q. Let χ_n denote the formula $q \neq n$ in this extended language (recall that every natural number n coincides with the constant term $(\ldots(0+1)+\ldots+1)+1$ in the language of PA, obtained by adding 1 repeatedly n times to 0).

We introduce the theories $S = \{\chi_n \colon n \in \mathbb{N}\}$ and $S_n = \{\chi_0, \chi_1, \ldots, \chi_n\}$ for each $n \in \mathbb{N}$. Interpreting the symbol q as the natural number $q^{\mathcal{M}_n} = n+1$, we obtain the model $\mathcal{M}_n = (\mathbb{N}; +, \cdot, 0, 1, n+1)$ of the theory $\mathrm{PA} \cup S_n$. By the Compactness Theorem it follows, that also the theory $\mathrm{PA} \cup S$ has some model $\mathcal{M} = (M; +, \cdot, 0, 1, q)$. In this model, the interpretation $q^{\mathcal{M}} \in M$ of the symbol q differs from all the constant terms $n \in \mathbb{N}$, thus $(M; +, \cdot, 0, 1)$ is a nonstandard model of PA.

Intuitively, $(M; +, \cdot, 0, 1)$ can be viewed as a number system extending the standard natural number system $(\mathbb{N}; +, \cdot, 0, 1)$ by some ideal "infinite" natural numbers, with one of them represented by (the interpretation of) the constant symbol q. Then of course, $q-1$, $q+1$, $2q$, q^2, etc., represent different infinite elements of M. Moreover, if $p \in M$ is any infinite element then so is $p-1$, since if $p-1$ were finite then $p = (p-1)+1$ would be finite, too. Thus the nonempty set of all infinite elements of M has no least element, seemingly contradicting the Well Ordering Principle implied by the Scheme of Induction of PA. Nonetheless, this paradox has a simple resolution: the sets of finite and infinite elements in M, respectively, are not first-order expressible. This means that there are no formulas $\varphi(x, u_1, \ldots, u_n)$, $\psi(x, u_1, \ldots, u_n)$ in the language of

PA and no (finite or infinite) elements $p_1, q_1, \ldots, p_n, q_n \in M$ such that

$$\{a \in M\colon a \text{ is finite}\} = \{a \in M\colon \mathcal{M} \vDash \varphi(a, p_1, \ldots, p_n)\}$$
$$\{a \in M\colon a \text{ is infinite}\} = \{a \in M\colon \mathcal{M} \vDash \psi(a, q_1, \ldots, q_n)\}$$

Similar accounts show that there are also nonstandard number systems $({}^*\mathbb{R}; +, \cdot, 0, 1)$, extending the standard number system $(\mathbb{R}; +, \cdot, 0, 1)$ of all real numbers and satisfying all the axioms of the Theory of Real Closed Fields, having the same first-order properties as $(\mathbb{R}; +, \cdot, 0, 1)$. Such number systems contain, besides standard reals, also infinite (infinitely big) and infinitesimal (infinitely small) number quantities. Using them, it is possible (among other things) to develop the infinitesimal (i.e., the differential and the integral) calculus in an intuitively appealing way, close to its historically original form, in the spirit of Newton, Leibniz, Euler and others, and that way to rehabilitate and justify the approach abandoned during the 19^{th} century in favor of the techniques of limits and the $\varepsilon\delta$-analysis.

4.10 Cardinality of Models and Skolem's Paradox

A detailed inspection of the proof of Gödel's Completeness Theorem (both in Section 4.8 as well as Section 4.11) shows that the construed model satisfies an additional cardinality specification.

The *cardinality of a first-order language* $L = (F, C, R, \nu)$ is defined as

$$\|L\| = |\operatorname{Form}(L)| = \max(|F|, |C|, |R|, \aleph_0)$$

A *first-order language* L is called *countable* if $\|L\| = \aleph_0$. Obviously, any first-order language with just finitely many specific symbols is countable.

It is clear that the set K of all constant terms of any first-order language L has the cardinality $|K| \leq |\operatorname{Term}(L)|$. The base set M of the canonical structure $\mathcal{M}(T) = (M; \ldots)$ of any theory T in the language L is a quotient $M = K/\!\sim_T$, therefore $|M| \leq |K|$. Thus the canonical structure $\mathcal{M}(T) = (M; \ldots)$ of any theory in a first--order language L has the cardinality

$$|M| \leq |K| \leq |\operatorname{Term}(L)| \leq |\operatorname{Form}(L)| = \|L\|$$

Similarly, the set D of new constant symbols, added to the language L for the sake of construction of the complete Henkin extension $\widehat{T} = T_H^+$ of T, has the cardinality $\|L\|$, again. Thus the new language L_D has the same cardinality as the original language L. Putting things together, we see that the canonical structure $\mathcal{M}\big(\widehat{T}\big) = (M; \ldots)$ still has the cardinality $|M| \leq \|L\|$. As a consequence, we obtain the following strengthening of Gödel's Completeness Theorem.

4.10.1 Theorem. *Let T be a consistent theory in a first-order language L of cardinality $\|L\| = \alpha$. Then T has a model $\mathcal{M} = (M; \ldots)$ of cardinality $|M| \leq \alpha$.*

4.10.2 Corollary. *Every consistent first-order theory in a countable language has a countable model, i.e., a model $\mathcal{M} = (M; \ldots)$ of cardinality at most $\aleph_0$.*

Realizing that the usual language of Set Theory has a single specific symbol, namely the binary relational symbol $\in$ for the membership relation, we readily obtain:

4.10.3 Corollary. *Any of the set theories* ZF, ZFC *(if it is consistent) has a countable model* $\mathcal{M} = \big(M; \in^{\mathcal{M}}\big)$.

Moreover, by *Mostowski Collapse Theorem*, we can arrange that $a \subseteq M$ for $a \in M$, and $a \in^{\mathcal{M}} b$ if and only if $a \in b$ for all $a, b \in M$. Then, according to the Soundness Theorem 4.4.8, everything that can be proved in Set Theory must be satisfied in $\mathcal{M}$. In particular, there are sets $\mathbb{N}^{\mathcal{M}}$, $\mathbb{R}^{\mathcal{M}} \in M$ playing in $\mathcal{M}$ the role of the set of all natural numbers and of the set of all real numbers, respectively. However, since $\mathbb{N}^{\mathcal{M}} \subseteq M$, $\mathbb{R}^{\mathcal{M}} \subseteq M$, both the sets $\mathbb{N}^{\mathcal{M}}$, $\mathbb{R}^{\mathcal{M}}$ are countable, hence (as it is clear that none of them can be finite) there is a bijective mapping $f\colon \mathbb{N}^{\mathcal{M}} \to \mathbb{R}^{\mathcal{M}}$. On the other hand, Cantor's Theorem "*the set* $\mathbb{R}$ *of all real numbers is uncountable*," which is provable in Set Theory, must be true in $\mathcal{M}$, as well. This sounds like a contradiction.

This paradox was discovered by the Norwegian mathematician Thoralf Skolem in 1922. However, Skolem derived it from the *Löwenheim-Skolem Downward Theorem* (which we will deal with later on) and not from Gödel's Completeness Theorem (though it was known to him well before Gödel proved and published it in 1930, but he neither proved it nor formulated it explicitly). Skolem's Paradox is not a contradiction proving the inconsistency of Set Theory. It can be resolved in the following way: The bijection $f\colon \mathbb{N}^{\mathcal{M}} \to \mathbb{R}^{\mathcal{M}}$ does not belong to the model $\mathcal{M}$; in fact there is no function $f \in M$ establishing a bijective correspondence $f\colon \mathbb{N}^{\mathcal{M}} \to \mathbb{R}^{\mathcal{M}}$. Thus Cantor's Theorem still holds in $\mathcal{M}$. Informally, the set $\mathbb{R}^{\mathcal{M}}$ is uncountable just from the internal point of view (i.e., as a set belonging to the model $\mathcal{M}$), while from the external point of view it is still countable. Nevertheless, Skolem's Paradox indicates that the notions like countability or uncountability, similarly as several other set-theoretical concepts concerning infinite cardinal numbers, are of a relative nature and lack an absolute character.

4.11 Proof of the Theorem on Complete Henkin Extensions

Let $(A; \leq)$ be a partially ordered set. An element $m \in A$ is called *maximal* if there is no element $a \in A$ such that $m < a$.

Any subset $\mathcal{T} \subseteq \mathcal{P}(X)$ of the powerset of any set X will be referred to a sa a *system of subsets* of X and automatically regarded as a partially ordered set $(\mathcal{T}; \subseteq)$ with the relation of set-theoretical inclusion. We say that a system $\mathcal{T} \subseteq \mathcal{P}(X)$ of subsets of a set X has *finite character* if for any $T \subseteq X$ we have $T \in \mathcal{T}$ if and only if $U \in \mathcal{T}$ for any finite set $U \subseteq T$.

We record without proof the following consequence of the Axiom of Choice; let's remark on the margin that it is even equivalent to it.

4.11.1 Teichmüller-Tukey Lemma. *Let* X *be any set and* $\mathcal{T} \subseteq \mathcal{P}(X)$ *be a system of finite character of subsets of* X. *Then for every* $T \in \mathcal{T}$ *there exists a maximal element* $M \in \mathcal{T}$ *such that* $T \subseteq M$.

Let us take for X the set $\Phi = \mathrm{Form}(L)$ of all formulas of some first-order language L and denote by $\mathcal{T} \subseteq \mathcal{P}(\Phi)$ the system of all consistent theories in L. As already noticed in the proof of the Compactness Theorem, a first-order theory is consistent if and only if every its finite subtheory $U \subseteq T$ is consistent. In other words, the system $\mathcal{T} \subseteq \mathcal{P}(\Phi)$ of all consistent theories in L has finite character. That way the Teichmüller-Tukey Lemma implies that every consistent theory T in a first-order language L can be extended to a maximal consistent theory $M \supseteq T$ in L.

4.11.2 Exercise. Let T be a theory in a first-order language L. We denote by

$$T^{\vdash} = \{\varphi \in \mathrm{Form}(L) \colon T \vdash \varphi\}$$

the set of all L-formulas provable in T. Show that T is complete if and only if $T^{\vdash}$ is a maximal consistent theory.

The above Exercise has furnished us with the last piece of knowledge needed in order to establish the following result.

4.11.3 Lindenbaum's Theorem. *Every consistent theory T in a first-order language L can be extended to a complete theory $T^{+} \supseteq T$ in L.*

4.11.4 Remark. Although (an equivalent alternative formulation of) the Axiom of Choice plays a crucial role in the proof of Lindenbaum's Theorem, it is known that this theorem is weaker than AC, in the sense that AC cannot be proved in Zermelo-Fraenkel Set Theory assuming Lindenbaum's Theorem, only.

Next we show a result on Henkin extensions of consistent theories.

4.11.5 Theorem on Conservative Henkin Extensions. *Let T be a consistent theory in a first-order language $L = (F, C, R, \nu)$. Then there is an extension of L by a set D of new constant symbols to a first-order language $L_D = (F, C \cup D, R, \nu)$ and a conservative extension of T to a Henkin theory $T_H \supseteq T$ in the language L_D.*

Demonstration. Our aim is to endow every L-sentence $(\exists\, x)\varphi(x)$ with a new witnessing constant d_φ and to extend the theory T by the corresponding witnessing axiom $(\exists\, x)\varphi(x) \Rightarrow \varphi(d_\varphi)$. However, doing so for all L-formulas $\varphi(x)$, the language L is likely extended and new sentences $(\exists\, x)\varphi(x)$ calling for their own witnessing constants arise. That's why we have to iterate the extension procedure recursively.

Let us denote by $L_0 = L$ the original first-order language and by Φ_0 the set of all L_0-formulas with a single free variable x. For every formula $\varphi \in \Phi_0$ we introduce a new constant symbol d_φ in such way that for different formulas $\varphi, \psi \in \Phi_0$ the symbols d_φ, d_ψ are distinct, as well. We denote by $D_0 = \{d_\varphi \colon \varphi \in \Phi_0\}$ the set of all these constants, by L_1 the extension of the language L_0 by the set D_0 of the new constants and by

$$W_0 = \{(\exists\, x)\varphi(x) \Rightarrow \varphi(d_\varphi) \colon \varphi \in \Phi_0\}$$

the set of all the witnessing axioms for the sentences $(\exists\, x)\varphi(x)$ where $\varphi \in \Phi_0$.

Assuming that the set of formulas Φ_n, the set $D_n = \{d_\varphi \colon \varphi \in \Phi_n\}$ of constant symbols, the language L_{n+1}, and the set $W_n = \{(\exists x)\varphi(x) \Rightarrow \varphi(d_\varphi) \colon \varphi \in \Phi_n\}$ of witnessing axioms are already defined, we denote by Φ_{n+1} the set of all formulas of the language L_{n+1} with a single free variable x not belonging to the union $\bigcup_{k=0}^{n} \Phi_k$. For every $\varphi \in \Phi_{n+1}$ we introduce a new (i.e., not occurring in the language L_{n+1}) constant symbol d_φ, with distinct symbols d_φ, d_ψ corresponding to different formulas φ, ψ. Next we denote by $D_{n+1} = \{d_\varphi \colon \varphi \in \Phi_{n+1}\}$ the set of all the recently added constants, by L_{n+2} the extension of the language L_{n+1} by the set D_{n+1} of these constants and by

$$W_{n+1} = \{(\exists x)\varphi(x) \Rightarrow \varphi(d_\varphi) \colon \varphi \in \Phi_{n+1}\}$$

the set of all witnessing axioms for the sentences $(\exists x)\varphi(x)$ where $\varphi \in \Phi_{n+1}$.

Finally we put $D = \bigcup_{n\in\mathbb{N}} D_n$, $W = \bigcup_{n\in\mathbb{N}} W_n$ and denote by L_D the extension of the language L by the set of the new constant symbols D. We claim that $T_H = T \cup W$ is a Henkin theory in the language L_D and a conservative extension of T.

It can be easily seen that every L_D-sentence of the form $(\exists x)\varphi(x)$ has a witness in the theory T_H: Since φ contains just finitely many constant symbols from D (if any), there is the smallest $n \in \mathbb{N}$ such that φ contains no symbol from D_m for all $m \geq n$. Then $\varphi \in \Phi_n$ and the sentence $(\exists x)\varphi(x) \Rightarrow \varphi(d_\varphi)$ belongs to W_n hence to T_H. Thus the constant symbol $d_\varphi \in D$ is a witness of the sentence $(\exists x)\varphi(x)$ in T_H.

In order to show that T_H is a conservative extension of T it is enough to verify that any L-sentence ψ provable in T_H is provable already in T. If $T_H \vdash \psi$ then there are finitely many witnessing axioms θ_i of the form $(\exists x)\varphi_i(x) \Rightarrow \varphi_i(d_i)$ with $1 \leq i \leq k$, where we write d_i instead of d_{φ_i}, such that $T \cup \{\theta_1, \ldots, \theta_k\} \vdash \psi$. If $k = 0$ then already $T \vdash \psi$ and we are done; thus we can assume that $k \geq 1$. Then it is enough to show that the number k of witnessing axioms can be anytime reduced by 1.

There's again the smallest n such that none of the formulas $\varphi_1, \ldots, \varphi_k$ contain any constant symbol from D_m for all $m \geq n$. Then $\varphi_j \in \Phi_n$ for some $j \in \{1, \ldots, k\}$ and none of the witnessing sentences θ_i for $i \neq j$ contains the symbol d_j. For brevity's sake we denote φ_j by φ, d_j by d and $\Theta = \{\theta_i \colon 1 \leq i \leq k,\ i \neq j\}$. As a consequence of the Deduction Theorem 4.5.1, we have

$$T \cup \Theta \vdash \big((\exists x)\varphi(x) \Rightarrow \varphi(d)\big) \Rightarrow \psi$$

Now the reader is asked to realize that, for any propositional forms A, B, C, both the propositional forms

$$((A \Rightarrow B) \Rightarrow C) \Rightarrow (\neg A \Rightarrow C) \qquad \text{and} \qquad ((A \Rightarrow B) \Rightarrow C) \Rightarrow (B \Rightarrow C)$$

are tautologies, therefore, by the Post Completeness Theorem 3.8.3, they are provable just from the propositional logical axioms. In particular, both the formulas

$$\big(\big((\exists x)\varphi(x) \Rightarrow \varphi(d)\big) \Rightarrow \psi\big) \Rightarrow \big(\neg(\exists x)\varphi(x) \Rightarrow \psi\big)$$
$$\big(\big((\exists x)\varphi(x) \Rightarrow \varphi(d)\big) \Rightarrow \psi\big) \Rightarrow (\varphi(d) \Rightarrow \psi)$$

are provable just from the logical axioms of Predicate Calculus. Then, by Modus Ponens, we have both

$$T \cup \Theta \vdash \neg(\exists x)\varphi(x) \Rightarrow \psi \qquad \text{as well as} \qquad T \cup \Theta \vdash \varphi(d) \Rightarrow \psi$$

Since the symbol d doesn't occur in any of the specific axioms of the theory $T \cup \Theta$, applying Lemma 4.7.1 on Constants to the latter item of the last couple of relations we obtain

$$T \cup \Theta \vdash (\forall x)(\varphi(x) \Rightarrow \psi)$$

Since ψ is closed, the variable x is not free in ψ, thus, according to (d) and the third equivalence in (b) of Exercise 4.4.11 on the prenex normal form,

$$T \cup \Theta \vdash (\exists x)\varphi(x) \Rightarrow \psi$$

This together with the former item of the above couple of relations gives $T \cup \Theta \vdash \psi$ in view of Corollary 4.5.4 on Proof by Distinct Cases. Since the set Θ consists of $k-1$ witnessing axioms, only, we are done.

4.11.6 Exercise. Show directly, i.e., without referring to the last Theorem, that the Henkin extension T_H of the consistent theory T constructed in its demonstration is consistent. (It can be done in a similar but slightly simpler way than in the demonstration of conservativeness of the extension T_H.)

Combining the two recently established theorems, we can finally prove the announced result.

4.8.5 Theorem on Complete Henkin Extensions. *Let T be a consistent theory in a first-order language $L = (F, C, R, \nu)$. Then there is an extension of L to a first--order language $L_D = (F, C \cup D, R, \nu)$ by a set D of new constant symbols and an extension of T to a complete Henkin theory $\widehat{T} \supseteq T$ in the language L_D.*

Demonstration. Let T_H be the conservative Henkin extension of the theory T in the language $L_D = (F, C \cup D, R, \nu)$ construed as above. Then, by Lindenbaum's Theorem, there is a complete theory $\widehat{T} = T_H^+ \supseteq T_H$ in the same language L_D. Obviously, $\widehat{T}$ is a Henkin theory, as well.

5 Gödel's Incompleteness Theorems

Gödel's Incompleteness Theorems belong to the most remarkable achievements of the 20^{th} century mathematics, shedding light on the limitations of formal methods and still raising philosophical questions about the nature of human thought, its possibilities and relations to our brains, to computers, etc.

> *Kurt Gödel's achievement in modern logic is singular and monumental — indeed, it is more than a monument, it is a landmark which will remain visible far in space and time.* [...] *The subject of logic has certainly completely changed its nature and possibilities with Gödel's achievement.*
>
> (John von Neumann)

5.1 Liar Paradox and the Paradoxes of Russell, Cantor and Berry

Let us recall the famous ancient *Liar Paradox*, also known as the *Epimenides Paradox*, usually ascribed to Eubulides of Miletus. In it the Cretan Epimenides pronounces the following sentence:

"All the Cretans are liars,"

tacitly assuming that "liar" means a person that always lies. If both the sentence and the tacit assumption are true, then Epimenides must be a liar and his statement must be a lie, as well. Then at least one Cretan is not a liar (in the sense that he does not lie all the time). Thus, finally, the sentence is not true, so Epimenides has told us a lie. This is not a contradiction, however, the fact that pronouncing a single false sentence can guarantee the existence of a person not lying all the time is still fairly paradoxical. The strong version of the *Liar Paradox* is due to the medieval French scholar Jean Buridan:

"What I am telling right now is a lie."

Even a simpler formulation is given by the following self-referential sentence:

"This sentence is not true."

At least at a glance it looks like a proposition, thus is seems legitimate to ask the question: "Is it true or false?" If it is false, then it must be true. Similarly, if it is true, then it cannot be true, hence it must be false. We can conclude that it is true if and only if it is not true. This is what we have in mind calling it the *strong version of Liar Paradox*.

In everyday life we need not to worry too much about the Liar Paradox. We can do away with it simply by marking that sentence as making no sense and not to care of it any more. However, the situation changes radically if such a self-referential sentence

could be formulated within some formal deductive system like, e.g., an axiomatic first--order theory. Such a theory would be necessarily inconsistent. This namely happened to the original version of Cantor's "naive" Set Theory.

Cantor's Set Theory used the unlimited version of the *Comprehension Principle* in forming sets:

> *For any "reasonable" property $P(x)$ one can form the set $\{x\colon P(x)\}$ of all objects x having this property.*

However intuitively appealing this principle might appear, it is fairly hazy, unless we make clear which properties we consider as "reasonable". What's even worse, this principle enables to formulate a set-theoretical version of Liar Paradox, namely *Russell's Paradox*, named after the British logician and philosopher Bertrand Russell:

> Cantor's Comprehension Principle allows us to form the set
> $$R = \{x\colon x \text{ is a set and } x \notin x\}$$
> of all sets x not belonging to itself.

Then the question: "Does the set R belong to itself?" immediately produces a contradiction. Indeed, we have $R \in R \Leftrightarrow R \notin R$.

Thus the original version of Cantor's Set Theory is inconsistent; the unlimited Comprehension Principle makes it possible to reproduce the Liar Paradox inside of this theory.

Liar Paradox can be avoided by restricting Cantor's Comprehension Principle to the following limited form:

> *For every set M and any "reasonable" property $P(x)$ one can form the set $\{x \in M\colon P(x)\}$ of all objects x from the set M having this property.*

Then the previous formation of the set R becomes illegal, and Russell's Paradox disappears. Instead, it is transformed to the following fact:

> *There is no set of all sets.*

Indeed, if there were the set V of all sets, then we could legally form the set
$$R = \{x \in V\colon x \notin x\}$$
of all sets x not belonging to itself, and obtain the contradiction $R \in R \Leftrightarrow R \notin R$ once again.

Another set-theoretical paradox was discovered by Cantor himself. One of his famous theorems states that, for every set X, the cardinality of its power set
$$\mathcal{P}(X) = \{A\colon A \subseteq X\}$$
is less then that of X; in symbols $|X| < |\mathcal{P}(X)|$. Taking for X the set V of all sets we obtain $|V| < |\mathcal{P}(V)|$. On the other hand, as each element A of the power set $\mathcal{P}(V)$ is a set again, we have $\mathcal{P}(V) \subseteq V$, henceforth $|\mathcal{P}(V)| \leq |V|$, which is a contradiction. Also this paradox can be avoided by assuming that the totality V of all sets is not a set. In Cantor's own words, V is an "inconsistent set". Therefore, neither the powerset construction nor Cantor's theorem can legitimately be applied to it.

Berry's Paradox demonstrates the need to clarify the vague concept of a "reasonable property" and that way to make clear which properties can be used even in the limited Comprehension Principle.

Consider the set A of all natural numbers which can be defined by some phrase of English language consisting of less than twenty words. Since the English language has a finite vocabulary, there are just finitely many English phrases consisting of less than twenty words. Hence the set A is finite, and, as the set $\mathbb{N}$ of all natural numbers is infinite, there exist natural numbers not belonging to the set A. In other words the complement $\mathbb{N} \setminus A$ is nonempty, thus, according to the Well Ordering Principle, it contains the smallest element. Then this natural number is defined by the English phrase

> *"The smallest natural number which cannot be defined by any English phrase consisting of less than twenty words"*

which has eighteen words, only. Hence the smallest element of the set $\mathbb{N} \setminus A$ belongs to the set A, as well. However, this is a contradiction, since $A \cap (\mathbb{N} \setminus A) = \emptyset$.

5.2 In Quest for a Way Out of the Crisis

The discovery of paradoxes in Cantor's Set Theory at the turn of the 19^{th} and 20^{th} century threw the mathematics of that time into a deep crisis. Moreover, it happened shortly after Set Theory had become widely accepted and recognized as the universal foundations of the whole of mathematics, providing it with a general common language and a firm ground on which all mathematical branches could be formulated and presented in a uniform way. Therefore the task to find a way out of the crisis became highly acute.

Some mathematicians reacted by refusing completely the conception of actual infinity forming one of the cornerstones of Set Theory (H. Poincaré, L. E. J. Brouwer).

Namely Brouwer established the doctrine of *intuitionism*, insisting that the infinity can be treated just as a potential and never completed process of growth or decay beyond any limit. He also proposed a revision of logic, refusing some classical logical laws (e.g., the Law of Excluded Middle $\varphi \vee \neg\varphi$, or the quantifier law $\neg(\forall x)\neg\varphi(x) \Rightarrow (\exists x)\varphi(x)$) as inapplicable within the realm of potentially infinite domains. The competing doctrine of *logicism* suggested to develop mathematics as a branch of logic (G. Frege, B. Russell, A. N. Whitehead) and to avoid the self-reference phenomenon, which they found responsible for the contradictions, by means of a fairly complicated hierarchy of the *Theory of Types*. However, none of these conceptions could compete with the approaches offered by the Set Theory making use of the full power of classical logic and, at the same time, avoiding the cumbersome hierarchy of the Theory of Types, along with preserving the conception of actually infinite sets.

The axiomatic system of Set Theory designed by Ernst Zermelo, and later on upgraded by Abraham Fraenkel, became the generally accepted foundations of most of the modern mathematics. The Paradoxes of Russell, Cantor and Berry (and some similar ones) were avoided by a cautious formulation of the Scheme of Comprehension, allowing to single out new sets just as *subsets of sets given in advance* by means of *properties described by set-theoretical formulas*. Three exceptions of sets, still de-

scribed by set-theoretical formulas, but not singled out from any in advance given set, are allowed by the Axioms of Pair, Union and Power Set. (See 2.7 and 4.3.9.)

No one was able to reproduce the known paradoxes, nor to produce any contradiction within the Zermelo-Fraenkel axiomatic system with the Axiom of Choice ZFC. Unfortunately, this does not exclude the possibility that, all the same, there are some contradictions, hidden deeply under the surface. This raised the task to *prove* the consistency of ZFC or of some other axiomatic system of Set Theory, capable to undertake the role of the foundations of mathematics. The project of proving the consistency of the foundations of mathematics was formulated by David Hilbert, the leading figure of the that time mathematics, who also designed the central notions and methods necessary for that purpose. The project is known under the name *Hilbert's Program*.

> *Hilbert's Program was an ambitious and wide-ranging project in the philosophy and foundations of mathematics. In order to "dispose of the foundational questions in mathematics once and for all", Hilbert proposed a two-pronged approach in 1921: first, classical mathematics should be formalized in axiomatic systems; second, using only restricted, "finitary" means, one should give proofs of the consistency of these axiomatic systems. Although Gödel's Incompleteness Theorems show that the program, as originally conceived, cannot be carried out, it had many partial successes, and generated important advances in logical theory and meta-theory, both at the time and since.*
>
> (Richard Zach, *Hilbert's Program Then and Now*, arXiv:math/0508572)

5.3 Gödel's Incompleteness Theorems. Preliminary Accounts

Consider the following self-referential sentence:

"This sentence is unprovable."

tacitly assuming that every sentence which is provable, is necessarily true. Once again we find legitimate to ask the question: "Is that sentence true or false?" If it is false, then it is provable, hence it must be true. This contradiction shows that it cannot be false, hence it is true. That way we have proved that this sentence is true, in other words, *we have proved* the sentence. Thus it is provable, hence, since it declares its own unprovability, it is false. At the same time, provable sentences must be true. It seems that we once again obtained a contradiction, conspicuously reminding of Liar Paradox.

However, this conclusion can be avoided by making precise the concept of provability. If it means *provability within some formal axiomatic system* (e.g., within some first-order theory), then our proof of the above sentence is just an informal intuitive argumentation showing that it is true, and not a proof within that system. Moreover, statements about provability within a given formal system in general *do not belong* to that system, hence the question of their provability within that system makes no sense. Thus it seems that the threatening paradox can be swept away from the very beginning.

All the same, let us admit that some formal systems could perhaps satisfy the following two properties:

(1) There is a sufficiently extensive distinguished class of statements formulated in the language of that system such that all statements from this class which are provable in the system are true in some intuitively appealing meaning of this word.

(2) There is a statement belonging to the above mentioned distinguished class declaring its own unprovability within the system.

Then that system is necessarily incomplete in the following sense:

(3) There are intuitively true statements formulated in the language of the system (and even belonging to that distinguished class) which are unprovable within that system.

Namely the statement belonging to that distinguished class and declaring its own unprovability within the system is an example of an intuitively true statement which is not provable within the system.

Now, the reader probably can hardly suppress the feeling that the existence of such formal axiomatic systems (first-order theories) is merely hypothetical, and in fact it should be possible to show that nothing like that can exist. Thus it might be rather surprising to realize what the young Austrian mathematician, logician and philosopher Kurt Gödel (1906, Brno – 1978, Princeton) has proved in 1930. Namely, according to his *First Incompleteness Theorem*, Peano Arithmetic, as well as any first-order theory capable to serve as the foundations of a reasonable fragment of mathematics, like ZF or ZFC or Principia Mathematica, provide examples of such axiomatic systems. According to his *Second Incompleteness Theorem*, such systems are capable to formulate a statement declaring their own consistency, nonetheless, if they are consistent, they are unable to prove it, though, in that case, the statement itself is true. As one of the consequences of Gödel's discoveries it became manifest that the goals of Hilbert's Program cannot be achieved.

5.4 The First Gödel Incompleteness Theorem

It is worth mentioning that Gödel worked within the intentions of Hilbert's Program and his Incompleteness Theorems appeared surprisingly on the way, without having been planned or anticipated in advance. We will skip almost all technical issues of Gödel's proof and begin with displaying some final results of his coding of formulas and proofs by natural numbers and representation of the provability relation by certain arithmetical predicate. It should be noted that our presentation differs considerable from Gödel's original one.

Informally, the First Gödel Incompleteness Theorem states that any consistent formal system which is sufficiently ample to include Peano Arithmetic is necessarily incomplete, either in the sense that it contains some true propositions about natural numbers which it cannot prove (semantic version), or in the sense that it contains certain arithmetical propositions which it can neither prove nor refute (syntactic version).

Let's begin with introducing some concepts necessary for describing more precisely the variety of first-order theories to which Gödel's results apply. A first-order theory T in a language with finitely many specific symbols is called *recursively axiomatizable*

if it has just finitely many axioms or its axioms can be effectively recognized by some algorithm (e.g., by a computer program). A first-order theory T is called *arithmetical* if there is some interpretation of Peano Arithmetic in this theory. This is to say that there are some formulas $\mathrm{Nat}(x)$, $\mathrm{Add}(x,y,z)$, $\mathrm{Mult}(x,y,z)$, $\mathrm{Zero}(x)$, $\mathrm{One}(x)$ in the language of T defining the concept of natural number, the operations of addition and multiplication of natural numbers and the distinguished objects 0 and 1, respectively, in such a way that for the structure of natural numbers thus obtained all the axioms of PA can be proved in T. If T is an arithmetical theory then a formula φ in the language of T is called *arithmetical* if it is built out of the "new" atomic formulas of the form $x = y$, $\mathrm{Add}(x,y,z)$, $\mathrm{Mult}(x,y,z)$, $\mathrm{Zero}(x)$, $\mathrm{One}(x)$ by means of logical connectives and *bounded quantifications* $(\forall x)(\mathrm{Nat}(x) \Rightarrow \varphi)$, $(\exists x)(\mathrm{Nat}(x) \wedge \varphi)$. An arithmetical theory T is called *arithmetically correct* if all the arithmetical sentences provable in T are satisfied in $(\mathbb{N}; +, \cdot, 0, 1)$.

An obvious example of a recursively axiomatizable arithmetically correct theory is the Peano Arithmetic itself. Other paradigmatic examples of such theories are recursive extensions of PA by axioms which are true in $(\mathbb{N}; +, \cdot, 0, 1)$, as well as various set theories like, e.g., ZF or ZFC.

Given an arithmetical sentence θ we will say that θ is *true* or *valid* or *satisfied* if it is satisfied in the standard model of Peano Arithmetic $(\mathbb{N}; +, \cdot, 0, 1)$. For an arithmetical sentence of the form $\psi(k_1, \ldots, k_n)$, where $\psi(x_1, \ldots, x_n)$ is an arithmetical formula and $k_1, \ldots, k_n$ are concrete natural numbers (constant arithmetical terms), we also use to say that $\psi(k_1, \ldots, k_n)$ *holds* or, simply, $\psi(k_1, \ldots, k_n)$ in that case.

Gödel developed a method of *coding* or *enumeration* by means of which all the arithmetical formulas in the language of an arithmetical theory T with a single free variable x can be lined up in a sequence $\varphi_0(x), \varphi_1(x), \ldots, \varphi_n(x), \ldots$ in such a way that, for each n, the formula $\varphi_n(x)$ can be effectively constructed (e.g., by a program), and vice versa, for each arithmetical formula $\psi(x)$, its number n such that $\psi(x)$ coincides with $\varphi_n(x)$ can be effectively determined. If T is additionally recursively axiomatizable then also all proofs in T can be lined up in a sequence $\Delta_0, \Delta_1, \ldots, \Delta_k, \ldots$ in such a way that the correspondence $k \leftrightarrow \Delta_k$ can be effectively described (e.g., executed by some programs) in either direction. Moreover, in that case Gödel constructed two effectively decidable ternary arithmetical predicates $P(x,y,z)$ and $R(x,y,z)$ of provability and refutability, respectively, such that for any natural numbers k, m, n the following conditions are satisfied:

$P(m,n,k)$ if and only if Δ_k is a proof of the sentence $\varphi_n(m)$ in T

$R(m,n,k)$ if and only if Δ_k is a proof of the sentence $\neg\varphi_n(m)$ in T

At the same time the algorithmic decidability of the predicates $P(x,y,z)$ and $R(x,y,z)$ ensures that, for any $m, n, k \in \mathbb{N}$, the satisfaction of any of the statements $P(m,n,k)$, $\neg P(m,n,k)$, $R(m,n,k)$, $\neg R(m,n,k)$, respectively, in $(\mathbb{N}; +, \cdot, 0, 1)$ is equivalent to its provability in PA, henceforth in T. Namely the algorithm deciding whether $P(m,n,k)$ holds or not provides the proof either of the statement $P(m,n,k)$ or of its negation, and similarly for $R(m,n,k)$. Summing up, we have:

5.4.1 Theorem. *Assume that T is a consistent recursively axiomatizable arithmetical theory. Then, for any natural numbers m, n, k, the three conditions in each of*

the following four rows are equivalent:

$$\begin{array}{rll} P(m,n,k), & T \vdash P(m,n,k), & \Delta_k \text{ is a proof of the sentence } \varphi_n(m) \text{ in } T \\ R(m,n,k), & T \vdash R(m,n,k), & \Delta_k \text{ is a proof of the sentence } \neg\varphi_n(m) \text{ in } T \\ \neg P(m,n,k), & T \nvdash P(m,n,k), & T \vdash \neg P(m,n,k) \\ \neg R(m,n,k), & T \nvdash R(m,n,k), & T \vdash \neg R(m,n,k) \end{array}$$

In particular, T decides both the statements $P(m,n,k)$ and $R(m,n,k)$ for any m, n, k.

Now, we have all the necessary ingredients needed for the formulation of Gödel's results. Consider the formula $\neg(\exists\, z)P(x,x,z)$. It has a single free variable, namely x, hence it occurs in the sequence $\{\varphi_n(x)\}_{n\in\mathbb{N}}$ under some number — let's denote it g. Thus $\varphi_g(x)$ is the above formula, and substituting the natural number g into it for x we obtain the sentence $\varphi_g(g)$, i.e., $\neg(\exists\, z)P(g,g,z)$, saying that, for no $z = k$, Δ_k is the proof of the sentence $\varphi_g(g)$. In other words, the meaning of that sentence is:

$$\varphi_g(g)\colon \quad \textit{"The sentence } \varphi_g(g) \textit{ is not provable in } T\textit{."}$$

Hence $\varphi_g(g)$ is an example of a self-referential sentence in the language of T declaring its own unprovability. On the other hand, the reader should keep in mind that $\varphi_g(g)$ is an arithmetical statement, like, e.g., $\neg(\exists\, x,y,z)\big((x+1)^2+(y+2)^3=(z+3)^4\big)$, saying that the diophantic equation $(x+1)^2+(y+2)^3=(z+3)^4$ has no solution in the domain of all natural numbers.

Thus our preliminary accounts (see 5.3) entitle us to state the semantic version of the First Gödel Incompleteness Theorem.

5.4.2 First Gödel Incompleteness Theorem. [Semantic version] *If T is a recursively axiomatizable arithmetically correct first-order theory then the Gödel's sentence $\varphi_g(g)$ is true in $(\mathbb{N};+,\cdot,0,1)$, nonetheless, it is unprovable in T. Thus T is incapable to prove all the true arithmetical statements about natural numbers.*

In particular, neither PA nor any of the set theories like, like ZF or ZFC, can prove all the true arithmetical statements about natural numbers.

Since we have no direct access to the infinite domain $\mathbb{N}$ of all natural numbers, the semantic concept of arithmetical truth playing a key role in the semantic version of the First Gödel Incompleteness Theorem "smells of metaphysics" and may evoke some bewilderment in the reader. It relies on our belief that $(\mathbb{N};+,\cdot,0,1)$ is a model of PA, which, however, can hardly be considered as an obvious or firmly and doubtlessly established fact. Nonetheless, this belief is even stronger than the belief in the consistency of PA, which still lacks a direct, immediate evidence. Anyway, it will be interesting to see what we can infer from this weaker syntactic assumption.

5.4.3 First Gödel Incompleteness Theorem. [Syntactic version] *Let T be a recursively axiomatizable arithmetical theory.*

(a) *If T is consistent then the Gödel's sentence $\varphi_g(g)$ is unprovable in T.*

(b) *If T is ω-consistent then neither the sentence $\neg\varphi_g(g)$ is provable in T.*

Thus the assumption of ω-consistency of T implies that T is incomplete.

Let us remark that ω-consistency is a technical condition, stronger than mere consistency, which we will formulate in the course of the demonstration of (b).

Demonstration. (a) Assume that the sentence $\varphi_g(g)$ is provable in T. From this point on we can proceed in twodifferent ways. We will present both of them.

First, the provability of $\varphi_g(g)$ means that this sentence has some proof, say Δ_k, in T. Then $P(g,g,k)$ holds, and due to the algorithmic nature of the predicate $P(x,y,z)$, the statement $P(g,g,k)$ is provable in T. It follows that $(\exists\, z)P(g,g,z)$, which is equivalent to $\neg\varphi_g(g)$, is provable in T, as well. Thus we have both $T \vdash \varphi_g(g)$ and $T \vdash \neg\varphi_g(g)$, contradicting the consistency of T.

The second argument starts with realizing the form of $\varphi_g(g)$: in fact we have assumed that $T \vdash \neg(\exists\, z)P(g,g,z)$, hence $T \vdash (\forall\, z)\neg P(g,g,z)$, since the second sentence is equivalent to the first one. It follows that $T \vdash \neg P(g,g,k)$ for each $k \in \mathbb{N}$. Therefore, $\neg P(g,g,k)$ holds for each k, by Theorem 5.4.1. It means that none of the proofs Δ_k is a proof of the sentence $\varphi_g(g)$ in T, in other words, $\varphi_g(g)$ is unprovable in T. This contradicts our original assumption which is henceforth wrong. Therefore $\varphi_g(g)$ is unprovable in T.

(b) Assume that the sentence $\neg\varphi_g(g)$, which is equivalent to $(\exists\, z)P(g,g,z)$, is provable in T. If there were some $k \in \mathbb{N}$ such that $P(g,g,k)$, we could infer that Δ_k is a proof of $\varphi_g(g)$ in T. Then both $\varphi_g(g)$ as well as $\neg\varphi_g(g)$ were provable in T, and we could refute our initial assumption that $T \vdash \neg\varphi_g(g)$ as contradicting the mere consistency of T (and get through without the assumption of its ω-consistency).

So does the provability of the arithmetical sentence $(\exists\, z)P(g,g,z)$ imply that there is indeed some $k \in \mathbb{N}$ such that $P(g,g,k)$? The positive answer to this question seems obvious at a glance. If there were a constructive proof of the statement $(\exists\, z)P(g,g,z)$, it would give us some concrete k such that $P(g,g,k)$. Unfortunately, we cannot exclude that the proof of the statement $(\exists\, z)P(g,g,z)$ proceeds in an indirect nonconstructive way, just deriving a contradiction from the assumption $\neg(\exists\, z)P(g,g,z)$, and giving not even a hint how the k such that $P(g,g,k)$ could be found. To conclude, our optimism was precocious, and our original idea of demonstration doesn't work. To get through we need something more.

An arithmetical theory T is called *ω-consistent* if, for no arithmetical formula $\psi(z)$, all the sentences $(\exists\, z)\psi(z)$, $\neg\psi(0)$, $\neg\psi(1)$, ..., $\neg\psi(k)$, ... are provable in T. Obviously, any ω-consistent arithmetical theory must be consistent.

Now assuming that T is ω-consistent and $T \vdash (\exists\, z)P(g,g,z)$, we can conclude that $T \nvdash \neg P(g,g,k)$ for some k. Then $P(g,g,k)$ holds for this k by Theorem 5.4.1. From this point on the original argument can be applied.

The following Example illustrates the difference between a purely existential and a constructive proof of an existential statement.

5.4.4 Example. We will prove the theorem:

"There exist irrational numbers $a, b > 0$ such that the number a^b is rational."

Proof. It is known (and easy to show) that $\sqrt{2}$ is an irrational number. Then the number $\sqrt{2}^{\sqrt{2}}$ is either rational or irrational. If it is rational, we are done by taking

$a = b = \sqrt{2}$. If $\sqrt{2}^{\sqrt{2}}$ is irrational, we put $a = \sqrt{2}^{\sqrt{2}}$ and $b = \sqrt{2}$. Then both a, b are irrational and

$$a^b = \left(\sqrt{2}^{\sqrt{2}}\right)^{\sqrt{2}} = \left(\sqrt{2}\right)^{\left(\sqrt{2}\,\sqrt{2}\right)} = \left(\sqrt{2}\right)^2 = 2$$

Since $a^b = 2$ is obviously rational, we are done again.

The reader should realize that our proof is purely existential, making use of the Law of Excluded Middle. We do not know whether the number $\sqrt{2}^{\sqrt{2}}$ is rational or irrational, therefore we do not know which one of the couple of possibilities really works. From the intuitionistic or constructivist viewpoint such proofs are unacceptable. A constructive proof would require to decide whether $\sqrt{2}^{\sqrt{2}}$ is rational or irrational and provide an explicit unambiguous choice of the pair a, b.

In fact it is known (but not so easy to show) that the number $\sqrt{2}^{\sqrt{2}}$ is irrational, hence the second possibility takes place in the proof above.

5.4.5 Exercise. Assume that Peano Arithmetic is a consistent Henkin theory. Show that, in such a case, it is necessarily ω-consistent. (*Hint*: Use the fact that every constant term in the language of PA is provably in PA equal to some term of the form $(\ldots(0+1)+\cdots+1)+1$ (with n occurrences of 1), i.e., to the standard natural number n.)

Later on, in 1936, Barkley Rosser formulated a modification of Gödel's statement $\neg(\exists\, z)P(g, g, z)$ making possible to avoid the assumption of ω-consistency and to prove the incompleteness of recursively axiomatizable arithmetic theories assuming their mere consistency.

Consider the arithmetical formula $(\forall\, z)(P(x, x, z) \Rightarrow (\exists\, u \leq z)R(x, x, u))$. Let us denote by r its number in the list $\{\varphi_n(x)\}_{n\in\mathbb{N}}$. Substituting r for x into the formula $\varphi_r(x)$ we obtain the self-referential sentence $\varphi_r(r)$, i.e.,

$$(\forall\, z)(P(r, r, z) \Rightarrow (\exists\, u \leq z)R(r, r, u))$$

Its meaning can be deciphered as follows:

$\varphi_r(r)$: *"If the sentence $\varphi_r(r)$ is provable in T by some proof of a given number then, among the proofs with at most that number, there is a proof of its negation $\neg\varphi_r(r)$ in T."*

Needless to say, the following version of the First Incompleteness Theorem is of syntactic nature.

5.4.6 Gödel-Rosser Incompleteness Theorem. *Let T be any recursively axiomatizable arithmetical theory. If T is consistent then neither the Rosser sentence $\varphi_r(r)$ nor its negation $\neg\varphi_r(r)$ are provable in T. Hence, if T is consistent then it is incomplete.*

Demonstration. Assume that $\varphi_r(r)$ is provable and Δ_k is its proof in T. Then both the statements $P(r,r,k)$ and $(\exists\, u \le k)R(r,r,u)$ are provable in T, as well. The latter is equivalent to the alternative

$$R(r,r,0) \vee R(r,r,1) \vee \ldots \vee R(r,r,k)$$

Then, however, it suffices to check the proofs Δ_0, Δ_1, $\ldots$, Δ_k and it is guaranteed that one from among them is a proof of the sentence $\neg\varphi_r(r)$ in T, contradicting its consistency.

Now, assume that $\neg\varphi_r(r)$ is provable in T by a proof Δ_l. Then we have $R(r,r,l)$, and the algorithm verifying this fact provides a proof of $R(r,r,l)$ in T. Realizing that $\neg\varphi_r(r)$ is equivalent to the sentence

$$(\exists\, z)\big(P(r,r,z) \wedge (\forall\, u)(R(r,r,u) \Rightarrow z < u)\big)$$

we can infer that the statement $(\exists\, z < l)P(r,r,z)$ is provable in T. Then necessarily $l > 0$, and the last statement is equivalent to the alternative

$$P(r,r,0) \vee P(r,r,1) \vee \ldots \vee P(r,r,l-1)$$

Hence among the proofs Δ_0, Δ_1, $\ldots$, Δ_{l-1}, there is a proof of the sentence $\varphi_r(r)$ in T, contradicting the consistency of T, again.

5.5 The Second Gödel Incompleteness Theorem

Informally, the Second Gödel Incompleteness Theorem states that any formal system which is sufficiently ample to include Peano Arithmetic cannot prove its own consistency. However, it should be realized that the statement that some formal system is consistent is a statement *about the system* which is not even formulated in the language of the system, thus the question of its provability within the system makes no sense. Hence it is a highly important fact that some formal systems, in particular, all recursively axiomatizable arithmetical first-order theories, indeed allow for formulation of such statements.

Given an arithmetical theory T, we say that an arithmetical sentence θ is a *consistency statement* for T if the consistency of T is equivalent to the validity of θ in $(\mathbb{N};\, +, \cdot, 0, 1)$. If T is additionally recursively axiomatizable then there are several possibilities how to formulate the consistency statement for T.

(1) Let us recall that a theory T in a first-order language L is inconsistent if and only if there is a sentence ψ in the language L such that both ψ and $\neg\psi$ are provable in T. Accordingly, a consistency statement for a recursively axiomatizable arithmetical theory T can be formulated in the following fairly suggestive way:

$$\mathrm{Cons}_1(T)\colon \neg(\exists\, x, y, z, u)(P(x,y,z) \wedge R(x,y,u))$$

excluding the existence of any sentence of the form $\varphi_n(m)$ such that both $\varphi_n(m)$, $\neg\varphi_n(m)$ were provable in T.

(2) Equivalently, T is inconsistent if and only if *every* L-sentence is provable in T. Thus T is consistent if and only if there is at least one L-sentence ψ not provable

in T. We have a considerable freedom of choice for this sentence. In particular, we can follow the "way of economy" suggested by John von Neumann and take Gödel's statement $\varphi_g(g)$ for that purpose. Indeed, if $\varphi_g(g)$ is provable in T then, as we already have seen, T is inconsistent. The other way round, if $\varphi_g(g)$ is unprovable in T then, of course, T is consistent. Thus T is consistent if and only if $\varphi_g(g)$ is not provable in T. This gives us the consistency statement

$$\mathrm{Cons}_2(T)\colon \neg(\exists\, z)P(g,g,z),$$

which coincides with the formerly introduced Gödel's statement $\varphi_g(g)$.

(3) Last but not least, we can take some logical axiom or some axiom of PA; then the requirement of unprovability of its negation is clearly equivalent to the consistency of T. In particular, let $s \in \mathbb{N}$ be the number of the formula $x \neq x$. Then $\varphi_s(0)$ is the sentence $0 \neq 0$. That way we obtain yet another consistency statement:

$$\mathrm{Cons}_3(T)\colon \neg(\exists\, z)P(0,s,z)$$

expressing the unprovability of the sentence $0 \neq 0$ in T.

5.5.1 Exercise. (a) When dealing with the syntactic version of the First Gödel's Incompleteness Theorem, we have shown that from the provability of Gödel's sentence $\varphi_g(g)$ in T there follows the provability of its negation $\neg\varphi_g(g)$ in T. Taking for granted that the implication

$$(\exists\, z)P(g,g,z) \Rightarrow (\exists\, u)R(g,g,u)$$

formalizing that account is provable in T, show that the implication

$$\neg(\exists\, x,y,z,u)(P(x,y,z) \wedge R(x,y,u)) \Rightarrow \neg(\exists\, z)P(g,g,z)$$

is provable in T, as well. Therefore the provability of the consistency statement $\mathrm{Cons}_1(T)$ from (1) in T implies the same for Gödel's sentence $\varphi_g(g)$.

(b) Similarly as in (a), we can infer that from the provability of Gödel's sentence $\varphi_g(g)$ there follows the provability of the sentence $0 \neq 0$ in T. Take for granted that the implication

$$(\exists\, z)P(g,g,z) \Rightarrow (\exists\, u)P(0,s,u)$$

formalizing this account is provable in T and show that the provability of the consistency statement $\mathrm{Cons}_3(T)$ from (3) in T implies the same for Gödel's sentence $\varphi_g(g)$, again.

Thus for a recursively axiomatizable arithmetical theory T with the provability predicate $P(x,y,z)$ and, possibly, with the refutability predicate $R(x,y,z)$, the consistency statement $\mathrm{Cons}(T)$ can be formulated within its language by any of the three sentences $\mathrm{Cons}_1(T)$, $\mathrm{Cons}_2(T)$, $\mathrm{Cons}_3(T)$ mentioned above (as well as by many more ones). At the same time, the assumption of provability of any of these statements in T yields the provability of Gödel's sentence $\varphi_g(g)$ in T. Summing up we have:

5.5.2 Second Gödel Incompleteness Theorem. *Let T be a recursively axiomatizable arithmetical theory. Then T allows for the formulation of its own consistency statement* $\mathrm{Cons}(T)$. *However, if T is consistent then any of its consistency statements* $\mathrm{Cons}_1(T)$, $\mathrm{Cons}_2(T)$, $\mathrm{Cons}_3(T)$ *from the above list is unprovable in T.*

If T is a consistent recursively axiomatizable arithmetical theory then, by Gödel's Second Incompleteness Theorem, the statement $\mathrm{Cons}(T)$ is not provable in T, thus, due to the Corollary 4.5.3 on Proof by Contradiction, its extension $T \cup \{\neg\, \mathrm{Cons}(T)\}$ is consistent, as well. However, as T is consistent, the new axiom $\neg\, \mathrm{Cons}(T)$ is not satisfied in $(\mathbb{N};\, +, \cdot, 0, 1)$, hence $T \cup \{\neg\, \mathrm{Cons}(T)\}$ cannot be arithmetically correct even if T were so. Next we denote, for definiteness' sake, by $\mathrm{Cons}(T)$ the Gödel's statement $\varphi_g(g)$. Then the statement $(\exists\, z)P(g, g, z)$, being logically equivalent to $\neg\, \mathrm{Cons}(T)$, is provable in $T \cup \{\neg\, \mathrm{Cons}(T)\}$. However, since T is consistent, none of the proofs Δ_k is a proof of the sentence $\varphi_g(g)$, i.e., of $\mathrm{Cons}(T)$, in T, therefore all the statements $\neg P(g, g, k)$, for $k \in \mathbb{N}$, are true in $(\mathbb{N};\, +, \cdot, 0, 1)$, hence provable in T and the more in $T \cup \{\neg\, \mathrm{Cons}(T)\}$. That way $T \cup \{\neg\, \mathrm{Cons}(T)\}$ is an example of a consistent theory which is not ω-consistent. On the other hand, if T is arithmetically correct then so is $T \cup \{\mathrm{Cons}(T)\}$.

In the following two exercises T denotes a recursively axiomatizable arithmetical theory with the provability predicate $P(x, y, z)$ and refutability predicate $R(x, y, z)$.

5.5.3 Exercise. The initial account in (2) suggests the following formalization of the consistency statement for T

$$\mathrm{Cons}_4(T)\colon\, (\exists\, x, y)(\forall\, z)\neg P(x, y, z)$$

declaring the existence of some sentence $\varphi_n(m)$ unprovable in T. However, the fact that its syntactic complexity (due to the quantifier prefix $\exists\forall$) is one step higher than that of the previous three consistency statements causes that it is not so easy to derive any conclusions from the assumption of the provability of $\mathrm{Cons}_4(T)$ in T.

(a) Show that $\mathrm{Cons}_4(T)$ is a consistency statement for T.

(b) Examine the provability status of the consistency statement $\mathrm{Cons}_4(T)$ in T. Realize that the mere assumption that T is consistent still does not allow us to show neither that $\mathrm{Cons}_4(T)$ is provable nor that it is unprovable in T.

(c) Observe that the implication $\varphi_g(g) \Rightarrow \mathrm{Cons}_4(T)$ is logically valid. Next, show that if T is ω-consistent then the negation $\neg\, \mathrm{Cons}_4(T)$ is not provable in T.

5.5.4 Exercise. (a) Show that the Rosser sentence $\varphi_r(r)$ is not a consistency statement for T. What about its negation $\neg\varphi_r(r)$?

(b) Show that both the sentences $\neg(\exists\, z)P(r, r, z)$, $\neg(\exists\, u)R(r, r, u)$ are consistency statements for T. What is their provability status?

In view of Gödel's results it is perhaps surprising but anyway worthwhile to mention that Solomon Feferman in 1960, at the cost of higher complexity, constructed a consistency statement $\mathrm{Cons}^*(\mathrm{PA})$ for Peano Arithmetic which, nevertheless, is *provable* in PA. However, for such a consistency statement neither the equivalence

$\mathrm{Cons}^*(\mathrm{PA}) \Leftrightarrow \mathrm{Cons}_i(\mathrm{PA})$ nor even the implication $\mathrm{Cons}^*(\mathrm{PA}) \Rightarrow \mathrm{Cons}_i(\mathrm{PA})$, for any $i = 1, 2, 3$, is provable in PA (unless PA is inconsistent).

The other way round, the consistency of Peano Arithmetic itself (and not only some of its consistency statements) can be proved within some stronger theories than PA (whose consistency is, for that reason, more doubtful than that of PA). The list obviously includes any reasonable version of Set Theory, as they allow for the construction of the standard model $(\mathbb{N}; +, \cdot, 0, 1)$ of PA. However, some fairly moderate extensions of PA manage that job already. Gerhard Gentzen in 1936 proved the consistency of PA by *transfinite induction* over the ordinal numbers less then the countable ordinal $\varepsilon_0 = \omega^{\omega^{\omega^{\cdot^{\cdot^{\cdot}}}}}$. (An outline of this method will appear in Section 5.8 on Goodstein sequences.) Gödel himself in 1958 gave a proof of consistency of PA using *recursive functionals*.

5.6 Attempts at Completion

There naturally arises the question whether Peano Arithmetic cannot be completed by adding to it some new axioms of which we know that they are satisfied in the standard model $(\mathbb{N}; +, \cdot, 0, 1)$. One possible candidate could be recursively constructed as follows: Let T_0 be the theory PA itself. Given the theory T_q, for $q \in \mathbb{N}$, we construct the sequence $\Delta_0^q, \Delta_1^q, \ldots, \Delta_k^q, \ldots$ of all proofs in T_q and the provability predicate $P_q(x, y, z)$ for T_q such that, for any $k, m, n \in \mathbb{N}$,

$$P_q(m, n, k) \quad \text{if and only if} \quad \Delta_k^q \text{ is a proof of the sentence } \varphi_n(m) \text{ in } T_q$$

Then we put

$$T_{q+1} = T_q \cup \{\mathrm{Cons}(T_q)\}$$

where $\mathrm{Cons}(T_q)$ is any of the consistency statements $\mathrm{Cons}_i(T_q)$ for fixed $i = 1, 2, 3$. In other words, T_{q+1} is the extension of T_q by the consistency axiom $\mathrm{Cons}(T_q)$ for T_q. Obviously, every T_q is a recursively axiomatizable arithmetically correct theory. Thus putting

$$\widehat{T} = \bigcup_{q \in \mathbb{N}} T_q$$

we obtain an arithmetically correct theory in which all the consistency statements $\mathrm{Cons}(T_q)$ can be proved. However, $\widehat{T}$ is still recursively axiomatizable, hence all the previous incompleteness results apply to it. In particular, $\widehat{T}$ is incomplete, it can formulate its own consistency statement $\mathrm{Cons}\big(\widehat{T}\big)$ which, nevertheless, it is incapable to prove. Moreover, as shown by Alan Turing in 1939, Peano Arithmetic cannot be completed even by transfinite iteration of the procedure of extending it by adding consecutive consistency statements to it.

One of the aspects of the incompleteness of Peano Arithmetic and related theories can be more specifically identified as the phenomenon of *ω-incompleteness*, i.e., a kind of "nonuniformity" of provability in them. For instance, if $\psi(x)$ is a formula in the language of PA then the provability in PA of all the sentences $\psi(m)$ for any $m \in \mathbb{N}$ still does not imply the provability of its universal closure $(\forall x)\psi(x)$ in PA. It can namely happen that the proofs of the particular instances $\psi(0)$, $\psi(1)$, …, $\psi(m)$, …

differ to such an extent that it is impossible to compose a uniform proof of the universally quantified statement $(\forall\, x)\psi(x)$ out of them. An arithmetical theory T is called *ω-complete* if this cannot happen, i.e., if, for every arithmetical formula $\psi(x)$, the provability in T of all the particular instances $\psi(m)$ for all $m \in \mathbb{N}$ already implies the provability of its universal closure $(\forall\, x)\psi(x)$ in T.

A construction of a complete extension of Peano Arithmetic based on the removal of the ω-incompleteness phenomenon was proposed by S. Feferman in 1962. However, in order to be able to extend PA to both a complete and ω-complete theory he had to sacrifice the condition of recursive axiomatization. By transfinite recursion over the ordinal numbers less than certain limit ordinal $\zeta \leq \omega^{\omega^{\omega}}$ he constructed a sequence of arithmetical theories $\{T_\alpha\}_{\alpha<\zeta}$ and a sequence of provability predicates $\big\{P_\alpha(x,y,z)\big\}_{\alpha<\zeta}$ for these theories such that

$$\begin{aligned} T_0 &= \mathrm{PA} \\ T_{\alpha+1} &= T_\alpha \cup \big\{(\forall\, x)(\exists\, z)P_\alpha(x,n,z) \Rightarrow (\forall\, x)\varphi_n(x)\colon n \in \mathbb{N}\big\} \quad \text{for each } \alpha < \zeta \\ T_\lambda &= \bigcup_{\alpha<\lambda} T_\alpha \quad \text{for any limit ordinal } \lambda < \zeta \end{aligned}$$

Adding the new axioms

$$(\forall\, x)(\exists\, z)P_\alpha(x,n,z) \Rightarrow (\forall\, x)\varphi_n(x)$$

for $n \in \mathbb{N}$ to the axioms of T_α guarantees the provability of every universally quantified statement $(\forall\, x)\varphi_n(x)$ in $T_{\alpha+1}$, once all its particular instances $\varphi_n(m)$ for $m \in \mathbb{N}$ are provable in T_α.

Finally, it can be shown that the arithmetical theory

$$\widetilde{T} = \bigcup_{\alpha<\zeta} T_\alpha$$

is not only ω-complete but also complete and ω-consistent. However, in order to derive at this conclusion we have to assume that PA is consistent. Assuming that PA is arithmetically correct, we can infer that so is $\widetilde{T}$.

5.7 The Theorems of Tarski and Church-Turing

To complete the picture we formulate two further incompleteness results by Alfred Tarski, and Alonzo Church and Alan Turing, respectively. *Tarski's Theorem on Undefinability of Truth* states informally that the property of arithmetical sentences "to be true" cannot be defined by any formula in the language of those sentences. More precisely, it says that the *satisfaction relation* for arithmetical formulas in the standard model $(\mathbb{N};\, +, \cdot, 0, 1)$ cannot be expressed by any arithmetical formula. In the theorems below we once again refer to the sequence $\{\varphi_n(x)\}_{n\in\mathbb{N}}$ of arithmetical formulas.

5.7.1 Tarski's Theorem on Undefinability of Truth. *Let T be any consistent arithmetical theory. Then there is no arithmetical formula $\sigma(x, y)$ in the language of T such that for any $m, n \in \mathbb{N}$ we have*

$$(\mathbb{N}; +, \cdot, 0, 1) \vDash \varphi_n(m) \Leftrightarrow \sigma(m, n)$$

Demonstration. Admit that such a formula $\sigma(x, y)$ exists. Then $\neg\sigma(x, x)$ is an arithmetical formula with a single free variable x, thus it can be found in the sequence $\{\varphi_n(x)\}_{n\in\mathbb{N}}$ under some index $t \in \mathbb{N}$. Then the following statements are equivalent in $(\mathbb{N}; +, \cdot, 0, 1)$: $\sigma(t, t)$, $\varphi_t(t)$, $\neg\sigma(t, t)$. Hence

$$(\mathbb{N}; +, \cdot, 0, 1) \vDash \sigma(t, t) \Leftrightarrow \neg\sigma(t, t)$$

which is contradiction.

Tarski's Theorem imposes severe limitations on the possibility of self-representation of arithmetical theories. In order to be able to define a satisfaction formula $\sigma(x, y)$ for such a theory T it is necessary to extend T to some theory T' in a "metalanguage" whose expressive power goes beyond that of T. For example, a satisfaction formula for Peano Arithmetic can be defined in the Second-Order Arithmetic or in the Zermelo-Fraenkel Set Theory.

When dealing with decidability questions, both A. Church and A. Turing were heavily influenced by the work of K. Gödel on completeness of the First-Order Logic and even more by his work on incompleteness of Peano Arithmetic and other arithmetical theories. While Church developed the so called *λ-calculus* and used it as a paradigmatic model of general computations, Turing designed ideal models of computing devices which became known as *Turing machines*. Soon it became clear that both approaches are equivalent. The proof of their Undecidability Theorem is beyond the scope of our course.

5.7.2 Church-Turing Undecidability Theorem. *Let T be any consistent recursively axiomatizable arithmetical theory. Then there is no algorithm which could decide whether any given arithmetical sentence in the language of T is provable in T. In particular, there is no algorithm which could decide the question of provability of the sentence $\varphi_n(m)$ in T for every input $(m, n) \in \mathbb{N} \times \mathbb{N}$.*

Church also proved that there is no algorithm which could decide whether a sentence in a first-order language L with at least one binary operational or relational symbol or with at least two unary operation symbols is a "first-order tautology", i.e., whether it is satisfied in all L-structures (or, which is the same, whether it is provable just from the logical axioms). Thus there is a striking difference between First-Order Logic and Propositional Calculus in which the tautologies can be effectively recognized by the truth table algorithm.

While Tarski's Theorem imposes some limits to what can be *expressed* by formal languages, Church-Turing Theorem sets up some limits to what can be *computed* by any mechanical or electronic device or *effectively decided* by means of an algorithmic computational procedure. However, they both, together with Gödel's Incompleteness Theorems, of course, raise various questions about the relation of computers, human brains and human mind or spirit.

5.8 Goodstein Sequences

It can be objected that the true statements unprovable in PA constructed by Gödel and Rosser, like $\varphi_g(g)$, $\varphi_r(r)$, Cons(PA) (no matter which possibility we choose), are highly artificial and deprived of proper mathematical meaning and content. However, there are indeed several known arithmetical theorems of combinatorial or number theoretic character which, nonetheless, are unprovable in PA. As a rule, they illustrate the ω-incompleteness phenomenon at the same time. Some examples of universally quantified statements of the form $(\forall x)\psi(x)$ unprovable in PA, nonetheless true in $(\mathbb{N}; +, \cdot, 0, 1)$ in the sense that all the particular instances $\psi(m)$ for each $m \in \mathbb{N}$ are even provable in PA, are provided by the *Paris-Harrington* strengthening of *Ramsey's Theorem* or by the *Goodstein sequences*. Both these results are in fact equivalent to the consistency of Peano Arithmetic. We will briefly explain the nature of the latter example.

Given a natural number $b \geq 2$, the *hereditary base b expansion* of any natural number m is obtained from the its usual base b expansion by expanding all its exponents at the base b, again, doing the same with the exponents of exponents, and repeating this procedure until all the numbers bigger than b are eliminated from this expression. For instance, the hereditary base 2 expansion of the number $m = 357$ reads as follows:

$$\begin{aligned} 357 &= 2^8 + 2^6 + 2^5 + 2^2 + 1 = 2^{2^3} + 2^{2^2+2} + 2^{2^2+1} + 1 \\ &= 2^{2^{2+1}} + 2^{2^2+2} + 2^{2^{2+1}} + 2^{2^2+1} + 1 \end{aligned}$$

Its hereditary base 3 expansion is

$$357 = 3^5 + 3^4 + 3^3 + 2 \cdot 3 = 3^{3+2} + 3^{3+1} + 3^3 + 2 \cdot 3$$

Similarly, the hereditary base 2 expansion of the number $m = 1\,000$ is

$$\begin{aligned} 1\,000 &= 2^9 + 2^8 + 2^7 + 2^6 + 2^5 + 2^3 \\ &= 2^{2^3+1} + 2^{2^3} + 2^{2^2+2+1} + 2^{2^2+2} + 2^{2^2+1} + 2^{2+1} \\ &= 2^{2^{2+1}+1} + 2^{2^{2+1}} + 2^{2^2+2+1} + 2^{2^2+2} + 2^{2^2+1} + 2^{2+1} \end{aligned}$$

On the other hand, its hereditary base 5 expansion coincides with its plane base 5 expansion:

$$1\,000 = 5^4 + 3 \cdot 5^3$$

For any natural number m we construct the *Goodstein sequence* of natural numbers

$$G(m, 0), G(m, 1), G(m, 2), \ldots, G(m, n), G(m, n+1), \ldots$$

corresponding to m, which starts with $G(m, 0) = m$, and having arrived at the number $G(m, n)$, if $G(m, n) > 0$ then the next item $G(m, n+1)$ is obtained by replacing every occurrence of the number $n+2$ in the hereditary base $n+2$ expansion of $G(m, n)$ by the number $n+3$ and subtracting 1 from the result; if $G(m, n) = 0$ then $G(m, n+1) = 0$,

as well. For example, for $m = 29$, we get

$$\begin{aligned}
G(m,0) &= 2^4 + 2^3 + 2^2 + 1 = 2^{2^2} + 2^{2+1} + 2^2 + 1 \\
G(m,1) &= 3^{3^3} + 3^{3+1} + 3^3 + 1 - 1 = 3^{3^3} + 3^{3+1} + 3^3 = 7\,625\,597\,485\,095 \\
G(m,2) &= 4^{4^4} + 4^{4+1} + 4^4 - 1 = 4^{4^4} + 4^{4+1} + 3 \cdot 4^3 + 3 \cdot 4^2 + 3 \cdot 4 + 3 \approx 134 \cdot 10^{152} \\
G(m,3) &= 5^{5^5} + 5^{5+1} + 3 \cdot 5^3 + 3 \cdot 5^2 + 3 \cdot 5 + 3 - 1 \\
&= 5^{5^5} + 5^{5+1} + 3 \cdot 5^3 + 3 \cdot 5^2 + 3 \cdot 5 + 2 \sim 10^{2\,200} \\
G(m,4) &= 6^{6^6} + 6^{6+1} + 3 \cdot 6^3 + 3 \cdot 6^2 + 3 \cdot 6 + 2 - 1 \\
&= 6^{6^6} + 6^{6+1} + 3 \cdot 6^3 + 3 \cdot 6^2 + 3 \cdot 6 + 1 \sim 10^{36\,305} \\
G(m,5) &= 7^{7^7} + 7^{7+1} + 3 \cdot 7^3 + 3 \cdot 7^2 + 3 \cdot 7 + 1 - 1 \\
&= 7^{7^7} + 7^{7+1} + 3 \cdot 7^3 + 3 \cdot 7^2 + 3 \cdot 7 \sim 10^{696\,000} \quad \ldots
\end{aligned}$$

The above computations indicate that the Goodstein sequences $\{G(m,n)\}_{n=0}^{\infty}$ grow rapidly for any m, and not just for the particular value $m = 29$. Thus the following result proved by R. L. Goodstein in 1944 is highly surprising and unexpected.

5.8.1 Goodstein's Theorem. *For every natural number m there exists a natural number n such that $G(m,n) = 0$.*

In fact, for $m \leq 3$, the sequence $\{G(m,n)\}_{n=0}^{\infty}$ assumes the value 0 fairly quickly. The reader can easily verify that $G(0,n) = 0$, for each n, $G(1,0) = 1$, $G(1,n) = 0$ for $n \geq 1$, $G(2,0) = G(2,1) = 2$, $G(2,2) = 1$ and $G(2,n) = 0$ for $n \geq 3$. For $m = 3$ we have

$$\begin{aligned}
&G(3,0) = 3 = 2 + 1, \quad G(3,1) = 3 + 1 - 1 = 3, \quad G(3,2) = 4 - 1 = 3, \\
&G(3,3) = 3 - 1 = 2, \quad G(3,4) = 2 - 1 = 1, \qquad G(3,5) = 0 = G(3,n) \quad \text{for } n > 5
\end{aligned}$$

For $m = 4$ the first n such that $G(4,n) = 0$ equals the immense value $3\big(2^{402\,653\,211} - 1\big)$.

Formally, the proof of Goodstein's Theorem uses transfinite induction over the countable well-ordered set of all ordinal numbers less than the ordinal

$$\varepsilon_0 = \omega^{\omega^{\omega^{\cdot^{\cdot^{\cdot}}}}}$$

i.e., the first-ordinal α satisfying $\omega^\alpha = \alpha$. However, the main idea of this proof can be explained easily. It consists in dominating every sequence $\{G(m,n)\}_{n=0}^{\infty}$, with m fixed, by a sequence $\{\Gamma(m,n)\}_{n=0}^{\infty}$ of ordinal numbers $\Gamma(m,n) < \varepsilon_0$, such that $G(m,n) \leq \Gamma(m,n)$ and $\Gamma(m,n) > \Gamma(m,n+1)$ whenever $\Gamma(m,n) > 0$, for each n. Since the set of all ordinals $< \varepsilon_0$ is well-ordered by the relation $<$, it cannot contain any infinite strictly decreasing sequence. Hence each of the sequences $\{\Gamma(m,n)\}_{n=0}^{\infty}$ must eventually stabilize at the value $\Gamma(m,n) = 0$ for some n. Then $G(m,n) = 0$, as well.

The ordinal number $\Gamma(m,n)$ is obtained by replacing each occurrence of the term $n+2$ in the hereditary base $n+2$ expansion of the number $G(m,n)$ by the ordinal ω.

In the particular case $m = 29$, we have

$$\begin{aligned}
G(m,0) &= 2^{2^2} + 2^{2+1} + 2^2 + 1 \\
&< \omega^{\omega^\omega} + \omega^{\omega+1} + \omega^\omega + 1 = \Gamma(m,0) \\
G(m,1) &= 3^{3^3} + 3^{3+1} + 3^3 \\
&< \omega^{\omega^\omega} + \omega^{\omega+1} + \omega^\omega = \Gamma(m,1) \\
G(m,2) &= 4^{4^4} + 4^{4+1} + 3\cdot 4^3 + 3\cdot 4^2 + 3\cdot 4 + 3 \\
&< \omega^{\omega^\omega} + \omega^{\omega+1} + 3\cdot\omega^3 + 3\cdot\omega^2 + 3\cdot\omega + 3 = \Gamma(m,2) \\
G(m,3) &= 5^{5^5} + 5^{5+1} + 3\cdot 5^3 + 3\cdot 5^2 + 3\cdot 5 + 2 \\
&< \omega^{\omega^\omega} + \omega^{\omega+1} + 3\cdot\omega^3 + 3\cdot\omega^2 + 3\cdot\omega + 2 = \Gamma(m,3) \\
G(m,4) &= 6^{6^6} + 6^{6+1} + 3\cdot 6^3 + 3\cdot 6^2 + 3\cdot 6 + 1 \\
&< \omega^{\omega^\omega} + \omega^{\omega+1} + 3\cdot\omega^3 + 3\cdot\omega^2 + 3\cdot\omega + 1 = \Gamma(m,4) \\
G(m,5) &= 7^{7^7} + 7^{7+1} + 3\cdot 7^3 + 3\cdot 7^2 + 3\cdot 7 \\
&< \omega^{\omega^\omega} + \omega^{\omega+1} + 3\cdot\omega^3 + 3\cdot\omega^2 + 3\cdot\omega = \Gamma(m,5) \ \ldots
\end{aligned}$$

Then the sequence of ordinals

$$\begin{aligned}
\Gamma(m,0) &= \omega^{\omega^\omega} + \omega^{\omega+1} + \omega^\omega + 1 \\
&> \Gamma(m,1) = \omega^{\omega^\omega} + \omega^{\omega+1} + \omega^\omega \\
&> \Gamma(m,2) = \omega^{\omega^\omega} + \omega^{\omega+1} + 3\cdot\omega^3 + 3\cdot\omega^2 + 3\cdot\omega + 3 \\
&> \Gamma(m,3) = \omega^{\omega^\omega} + \omega^{\omega+1} + 3\cdot\omega^3 + 3\cdot\omega^2 + 3\cdot\omega + 2 \\
&> \Gamma(m,4) = \omega^{\omega^\omega} + \omega^{\omega+1} + 3\cdot\omega^3 + 3\cdot\omega^2 + 3\cdot\omega + 1 \\
&> \Gamma(m,5) = \omega^{\omega^\omega} + \omega^{\omega+1} + 3\cdot\omega^3 + 3\cdot\omega^2 + 3\cdot\omega \\
&> \ldots
\end{aligned}$$

cannot decrease for ever, hence it must eventually stabilize at the value $\Gamma(m,n) = 0$ for some unimaginably huge value of n. For that n also $G(m,n) = 0$.

As shown by L. Kirby and J. Paris in 1982, Goodstein's Theorem cannot be proved just by means of the Peano Arithmetic alone.

5.8.2 Kirby-Paris Theorem. *In* PA *it is provable that Goodstein's Theorem implies the consistency statement* Cons(PA)*. As a consequence, if* PA *is consistent then Goodstein's Theorem is not provable in* PA*.*

On the other hand, the existential statement $(\exists\, y)(G(m,y) = 0)$ is provable in PA for any fixed $m \in \mathbb{N}$. We know this though already for rather small values of m we not only do not know the precise value of such a $y = n$ but we even do not dispose of any explicit proof of that statement in PA. We only know that the primitive step-by-step computation must eventually produce the result. However, this computation will not terminate within the existence not only of the mankind but of the entire universe. At the same time, as an illustration of the ω-incompleteness phenomenon mentioned in

connection with Feferman's construction, it should be realized that it is impossible to extract out of those particular proofs-computations any general common idea which would allow to convert them into a single proof of the universal-existential sentence $(\forall x)(\exists y)(G(x,y)=0)$ in PA.

5.9 Philosophical Consequences. Themes for an Essay

There is a vast literature dealing with mathematical, philosophical, metaphysical and others extra-mathematical consequences of Gödel's Incompleteness Theorems and some related results. Let us confine to a brief list of some traditionally inferred conclusions:

(1) Human knowledge is necessarily incomplete and we never can be sure that it is free of contradictions.

(2) Human knowledge cannot be reduced to any formal system. By realizing the incompleteness phenomena inherent for such systems we are capable to transcend their limitations.

(3) Computers can compute and prove just within the scope of some formal system. Humans, however, are able to seize and reveal some truths unprovable within any formal system. It follows that human brain — in spite of the fact that with respect to some parameters (as, e.g., the speed of computation) it is far behind the computers — still possesses some capabilities making it superior to any computer.

It is extremely interesting to present some Gödel's ideas upon these issues here. Gödel namely went a step farther beyond (3). According to him, we all probably agree that computers can compute and prove just within the scope of some formal system given in advance. Similarly, the activity of human brain can in principle be simulated by certain computer (though we do not dispose of such computers at present). However, human beings are capable of viewing or grasping even some truths unprovable within any formal system. It follows that human mind or human intellect or human spirit, however we call it, is endowed not only with some capabilities which make it superior to any computer but also with some faculties which cannot be explained as a mere manifestation of the activity and functioning of human brain.[5]

Try to ponder over the above quoted conclusions and opinions. To which degree you agree or disagree with any of them and why? To which degree can the above conclusions be justified by the incompleteness results we have been dealing with? Discuss those points and try to make them more precise, finally arriving at some formulations you can agree with. To which degree follow your conclusions from the results of Gödel, Rosser, Tarski, Church and Turing? Try to put these questions into the context of the recent turbulent development in Artificial Intelligence.

[5] Freely quoted according to Hao Wang [1996].

6 Substructures and Homomorphisms

This chapter marks the beginning of our discussion on *model theory*. From the extensive material covered by this mathematical discipline, we will limit ourselves to a brief introduction to its fundamental concepts and methods, which generalize several results that the reader has likely encountered in an algebra course. While model theory can, in a certain sense, be viewed as the "metamathematics of algebra", our approach will — unlike the previous three chapters — be more mathematical than metamathematical. Among other things, this will be reflected in the fact that the proofs of the stated results will mostly be presented under the usual heading "proof", instead of the previously used term "demonstration".

Throughout this chapter, we denote by $L = (F, R, \nu)$ some fixed but otherwise arbitrary first-order language, with constant symbols understood as nullary functional symbols. Unless confusion might arise or explicitly stated otherwise, the term "structure", or "first-order structure", will refer to a structure of the language L, and the term "theory", or "first-order theory", will refer to a theory in the language L. Similarly, words such as "term" or "formula" will be understood as a term or formula of the language L.

6.1 Substructures

Let $\mathcal{A} = (A; \dots)$, $\mathcal{B} = (B; \dots)$ be two first-order structures. We say that $\mathcal{B}$ is a *substructure* of $\mathcal{A}$, denoted $\mathcal{B} \subseteq \mathcal{A}$, if $B \subseteq A$ and for any n-ary functional symbol $f \in F$ and relational symbol $r \in R$, and for all $b_1, \dots, b_n \in B$, we have:

$$f^{\mathcal{B}}(b_1, \dots, b_n) = f^{\mathcal{A}}(b_1, \dots, b_n)$$
$$(b_1, \dots, b_n) \in r^{\mathcal{B}} \;\Leftrightarrow\; (b_1, \dots, b_n) \in r^{\mathcal{A}}$$

In other words, $f^{\mathcal{B}} = f^{\mathcal{A}} \restriction B^n$ and $r^{\mathcal{B}} = r^{\mathcal{A}} \cap B^n$. Specifically, for a constant symbol $f \in F$, this means that $f^{\mathcal{B}} = f^{\mathcal{A}}$. We also say that $\mathcal{A}$ is an *extension* of $\mathcal{B}$, denoted $\mathcal{A} \supseteq \mathcal{B}$.

The "substructure relationship" is clearly transitive, meaning that for any structures $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, we have:

$$\mathcal{A} \subseteq \mathcal{B} \,\wedge\, \mathcal{B} \subseteq \mathcal{C} \;\Rightarrow\; \mathcal{A} \subseteq \mathcal{C}.$$

The underlying set B of the substructure $\mathcal{B} \subseteq \mathcal{A}$ must be closed under all operations in the structure $\mathcal{A}$, i.e., for any n-ary functional symbol $f \in F$ and all $b_1, \dots, b_n \in B$, we have $f^{\mathcal{A}}(b_1, \dots, b_n) \in B$. Specifically, $f^{\mathcal{A}} \in B$ for every constant symbol $f \in F$.

Conversely, any nonempty subset $M \subseteq A$ of the domain of $\mathcal{A}$ which is closed under all operations of $\mathcal{A}$ uniquely determines a substructure $\mathcal{M} = (M; \dots)$ of $\mathcal{A}$

with carrier M. The operations and relations in $\mathcal{M}$ are then defined in the only possible way:

$$f^{\mathcal{M}} = f^{\mathcal{A}} \restriction M^n$$
$$r^{\mathcal{M}} = r^{\mathcal{A}} \cap M^n$$

for every n-ary functional or relational symbol $f \in F$, $r \in R$. We express these equalities by saying that the operations and relations of $\mathcal{M}$ are *inherited* from $\mathcal{A}$. This observation provides a certain flexibility in terminology: depending on the context, the term substructure of $\mathcal{A}$ may refer either to the structure $\mathcal{M}$ satisfying $\mathcal{M} \subseteq \mathcal{A}$, or to a nonempty subset $M \subseteq A$ that is closed under the operations of $\mathcal{A}$ (where the corresponding structure $\mathcal{M} = (M; \dots)$ then inherits its operations and relations from $\mathcal{A}$).

6.1.1 Example. Vector spaces over a given field F can be viewed as structures $\big(V; F \cup \{+\}\big)$ in a first-order language with the binary addition operation $+$, unary scalar multiplication operations $f \in F$, and no relational symbols (see Example 4.3.3). In this case, the substructures of the vector space $\big(V; F \cup \{+\}\big)$ are precisely its linear subspaces.

6.1.2 Exercise. Let $\mathcal{A}$, $\mathcal{B}$ be first-order structures with $\mathcal{A} \subseteq \mathcal{B}$, and let $t(x_1, \dots, x_n)$ be a term, and $\varphi(x_1, \dots, x_n)$ an atomic formula. Prove that, for all $b_1, \dots, b_n \in B$, we have:

$$t^{\mathcal{A}}(b_1, \dots, b_n) \in B$$
$$\mathcal{B} \vDash \varphi(b_1, \dots, b_n) \;\Leftrightarrow\; \mathcal{A} \vDash \varphi(b_1, \dots, b_n)$$

Remark. A reader with some knowledge of graph theory should realize that the concept of a substructure corresponds to the notion of an *induced subgraph* of a given graph $G = (V; E)$, rather than to the concept of its arbitrary subgraph.

The properties of substructures exhibit a certain sensitivity to the language. The following exercise illustrates what we mean by this.

6.1.3 Exercise. We have already mentioned three possible definitions of groups as structures: $(G; \cdot)$, $(G; \cdot, e)$, and $\big(G; \cdot, e, {}^{-1}\big)$ in languages of varying richness.

(a) Find several examples of groups $\mathcal{G} = (G; \cdot)$, and $\mathcal{G} = (G; \cdot, e)$, along with their substructures that are not themselves groups. Based on this, realize that in both of these languages, a substructure of a group does not necessarily have to be its subgroup.

(b) Prove that every substructure of a group $\mathcal{G} = \big(G; \cdot, e, {}^{-1}\big)$ is itself a group. In this case, it can rightfully be called a subgroup of the given group.

In connection with substructures of groups that are not subgroups, a general question arises: Which properties of a given structure, expressed by first-order formulas, are preserved in its substructures or in its extensions? Certain classes of such formulas can be described relatively easily.

A formula φ is called *open* if it contains no quantifiers. A formula φ is called *universal* if it has the form $(\forall x_1, \ldots, x_n)\psi$, where ψ is an open formula. Similarly, a formula φ is called *existential* if it has the form $(\exists x_1, \ldots, x_n)\psi$, where ψ is an open formula. Clearly, the negation of a universal (existential) formula is logically equivalent to an existential (universal) formula.

6.1.4 Proposition. *Let $\mathcal{A}$, $\mathcal{B}$ be first-order structures with $\mathcal{B} \subseteq \mathcal{A}$.*
(a) *If $\varphi(x_1, \ldots, x_n)$ is an open formula, then for all $b_1, \ldots, b_n \in B$, we have:*

$$\mathcal{B} \vDash \varphi(b_1, \ldots, b_n) \;\Leftrightarrow\; \mathcal{A} \vDash \varphi(b_1, \ldots, b_n)$$

(b) *If $\varphi(x_1, \ldots, x_n)$ is a universal formula, then for all $b_1, \ldots, b_n \in B$, we have:*

$$\mathcal{A} \vDash \varphi(b_1, \ldots, b_n) \;\Rightarrow\; \mathcal{B} \vDash \varphi(b_1, \ldots, b_n)$$

(c) *If $\varphi(x_1, \ldots, x_n)$ is an existential formula, then for all $b_1, \ldots, b_n \in B$, we have:*

$$\mathcal{B} \vDash \varphi(b_1, \ldots, b_n) \;\Rightarrow\; \mathcal{A} \vDash \varphi(b_1, \ldots, b_n)$$

Proof. (a) The statement can be easily proven by induction on complexity with respect to the logical connectives $\neg$ and (for example) $\wedge$. For an atomic formula φ, it suffices to refer to Exercise 6.1.2. The inductive steps for $\neg$ and $\wedge$ are left to the reader.

(b) A universal formula φ has the form $(\forall y_1, \ldots, y_m)\psi(x_1, \ldots, x_n, y_1, \ldots, y_m)$ for some open formula ψ. The condition $\mathcal{A} \vDash \varphi(b_1, \ldots, b_n)$ means that

$$(\forall a_1, \ldots, a_m \in A)\big(\mathcal{A} \vDash \psi(b_1, \ldots, b_n, a_1, \ldots, a_m)\big)$$

Since $B \subseteq A$, it follows that

$$(\forall a_1, \ldots, a_m \in B)\big(\mathcal{A} \vDash \psi(b_1, \ldots, b_n, a_1, \ldots, a_m)\big)$$

The required conclusion follows from (a).

(c) can be proven analogously to (b) or derived from (b) using the fact that every existential formula is logically equivalent to the negation of a universal formula.

6.1.5 Corollary. Let $\mathcal{A}$, $\mathcal{B}$ be first-order structures such that $\mathcal{B} \subseteq \mathcal{A}$.
(a) If φ is a closed universal formula, then from $\mathcal{A} \vDash \varphi$ it follows that $\mathcal{B} \vDash \varphi$.
(b) If φ is a closed existential formula, then from $\mathcal{B} \vDash \varphi$ it follows that $\mathcal{A} \vDash \varphi$.

In other words, universal properties are preserved under substructures, and existential properties are preserved under extensions. Conversely, this implies that properties that are not preserved under substructures (extensions) cannot be expressed by universal (existential) formulas. In the following exercise, we demonstrate some applications of this observation.

6.1.6 Exercise. (a) Using a formula in a language where the only specific symbol is the binary operation $\cdot$, express the condition for the existence of a neutral element for this operation. Prove that this condition cannot be expressed by either a universal or an existential formula in this language.

(b) Using a formula in a language where the only specific symbols are the binary operation $\cdot$ and the constant symbol e, express the condition that e is the neutral element of this operation using a universal formula. Prove that this condition cannot be expressed by an existential formula.

(c) In the same language as in (b) (and assuming that e serves as the neutral element of the operation $\cdot$), express the condition for the existence of an inverse element for any element under this operation. Prove that this condition cannot be expressed by either a universal or an existential formula in this language.

(d) Independently find additional examples of properties that cannot be expressed by universal or existential formulas in a given language.

6.1.7 Exercise. Consider the field of all real numbers as the structure $(\mathbb{R}; +, \cdot, 0, 1)$.

(a) Find examples of substructures of this structure that are rings but not fields. Also, find examples of its substructures that are not even rings.

(b) What is the smallest subfield of the field $(\mathbb{R}; +, \cdot, 0, 1)$? What is its smallest subfield containing all integers? What is the smallest subfield of $(\mathbb{R}; +, \cdot, 0, 1)$ that contains the number $\sqrt{2}$, the number $\sqrt[3]{5}$, or the number π?

(c) What is the smallest possible substructure of the field $(\mathbb{R}; +, \cdot, 0, 1)$? Is it a field (or a ring)?

In this context, natural questions arise as to whether all properties that are preserved under substructures (extensions) can be expressed using universal (existential) formulas. *A priori*, it cannot be ruled out that some properties expressed by formulas of other syntactic forms might also be preserved in this way. We reveal in advance that the answer to both of these questions is affirmative, but we will have to wait a little longer for the proofs of these results.

6.2 Homomorphisms

A *homomorphism* from a structure $\mathcal{A} = (A; \dots)$ to a structure $\mathcal{B} = (B; \dots)$ is a mapping $h\colon A \to B$ such that for any n-ary functional symbol $f \in F$ or relational symbol $r \in R$ and all $a_1, \dots, a_n \in A$, the following conditions hold:

$$h\big(f^{\mathcal{A}}(a_1, \dots, a_n)\big) = f^{\mathcal{B}}(h(a_1), \dots, h(a_n))$$
$$(a_1, \dots, a_n) \in r^{\mathcal{A}} \;\Rightarrow\; (h(a_1), \dots, h(a_n)) \in r^{\mathcal{B}}$$

In particular, for a constant symbol $f \in F$, this means that $h\big(f^{\mathcal{A}}\big) = f^{\mathcal{B}}$. The fact that the mapping $h\colon A \to B$ is a homomorphism from $\mathcal{A}$ to $\mathcal{B}$ is denoted symbolically as $h\colon \mathcal{A} \to \mathcal{B}$ or $h\colon (A; \dots) \to (B; \dots)$. Simply put, homomorphisms are mappings between structures that preserve their fundamental operations (including nullary ones) and relations.

6.2.1 Exercise. Let $h\colon \mathcal{A} \to \mathcal{B}$ be a homomorphism, $t(x_1, \dots, x_n)$ be a term, and $\varphi(x_1, \dots, x_n)$ be an atomic formula. Then for all $a_1, \dots, a_n \in A$, the following holds:

$$h\big(t^{\mathcal{A}}(a_1, \dots, a_n)\big) = t^{\mathcal{B}}(h(a_1), \dots, h(a_n))$$
$$\mathcal{A} \vDash \varphi(a_1, \dots, a_n) \;\Rightarrow\; \mathcal{B} \vDash \varphi(h(a_1), \dots, h(a_n))$$

Prove this.

The reader has likely encountered homomorphisms of groups or rings before. Likewise, linear mappings between vector spaces over a given field F are, in essence, nothing other than homomorphisms between structures of the form $\big(V; F \cup \{+\}\big)$, where $+$ is the binary operation of addition and the unary operations correspond to scalar multiplication by elements $f \in F$.

It is worth noting that, unlike substructures, homomorphisms between groups exhibit a significant degree of independence from the language. The explanation of this fact will occur later in this section.

6.2.2 Exercise. Let $\big(G; \cdot, e, {}^{-1}\big)$ and $\big(H; \cdot, e, {}^{-1}\big)$ be groups, and let $h\colon G \to H$ be an arbitrary mapping. Prove that the following conditions are equivalent:

(i) h is a homomorphism $\big(G; \cdot, e, {}^{-1}\big) \to \big(H; \cdot, e, {}^{-1}\big)$;

(ii) h is a homomorphism $(G; \cdot, e) \to (H; \cdot, e)$;

(iii) h is a homomorphism $(G; \cdot) \to (H; \cdot)$.

The composition of homomorphisms is again a homomorphism. Furthermore, homomorphisms preserve substructures of the given structures "in both directions". We leave the simple proofs of these facts as exercises for the reader.

6.2.3 Proposition. *Let $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ be first-order structures, and let $h\colon \mathcal{A} \to \mathcal{B}$ and $g\colon \mathcal{B} \to \mathcal{C}$ be homomorphisms. Then the composed mapping $g \circ h\colon \mathcal{A} \to \mathcal{C}$ is also a homomorphism.*

6.2.4 Proposition. *Let $\mathcal{A}$ and $\mathcal{B}$ be first-order structures, and let $h\colon \mathcal{A} \to \mathcal{B}$ be a homomorphism.*

(a) *If $M \subseteq A$ is a substructure of $\mathcal{A}$, then the set*

$$h[M] = \{h(a)\colon a \in M\} \subseteq B$$

is a substructure of $\mathcal{B}$.

(b) *If $N \subseteq B$ is a substructure of $\mathcal{B}$, and the set*

$$h^{-1}[N] = \{a \in A\colon h(a) \in N\} \subseteq A$$

is nonempty, then $h^{-1}[N]$ is a substructure of $\mathcal{A}$.

In short, the homomorphic image of a substructure is a substructure, and the homomorphic preimage of a substructure, as long as it is nonempty, is also a substructure. Note that if the language L contains at least one constant symbol, the condition $h^{-1}[N] \neq \emptyset$ is automatically satisfied for every substructure N of $\mathcal{B}$.

We say that a structure $\mathcal{B}$ is a *homomorphic image* of a structure $\mathcal{A}$ if there exists a surjective homomorphism $h : \mathcal{A} \to \mathcal{B}$. We are interested in determining which properties are preserved under homomorphic images. A formula φ is called *positive* if it is built from atomic formulas using only the logical connectives $\wedge$, $\vee$, and the quantifiers $\forall$, $\exists$.

6.2.5 Proposition. *Let $\mathcal{A}$ and $\mathcal{B}$ be first-order structures, and let $h\colon \mathcal{A} \to \mathcal{B}$ be a surjective homomorphism. Then for any positive formula $\varphi(x_1, \ldots, x_n)$ and all $a_1, \ldots, a_n \in A$, we have*

$$\mathcal{A} \vDash \varphi(a_1, \ldots, a_n) \;\Rightarrow\; \mathcal{B} \vDash \varphi(h(a_1), \ldots, h(a_n))$$

Proof. We proceed by induction on the complexity of the formula. For an atomic formula φ, it suffices to refer to Exercise 6.2.1. The inductive steps for both logical connectives and quantifiers will be performed simultaneously. Let $*$ denote any of the logical connectives $\wedge$, $\vee$, and let Q denote any of the quantifiers $\forall$, $\exists$ (both in the language L and in natural language).

Assume that the implication to be proven holds for each of the formulas $\varphi(x_1, \ldots, x_n)$ and $\psi(x_1, \ldots, x_n)$, and choose arbitrary elements $a_1, \ldots, a_n \in A$. Then the condition

$$\mathcal{A} \vDash (\varphi * \psi)(a_1, \ldots, a_n)$$

is equivalent to

$$\mathcal{A} \vDash \varphi(a_1, \ldots, a_n) \;*\; \mathcal{A} \vDash \psi(a_1, \ldots, a_n)$$

By the induction hypothesis, we obtain

$$\mathcal{B} \vDash \varphi(h(a_1), \ldots, h(a_n)) \;*\; \mathcal{B} \vDash \psi(h(a_1), \ldots, h(a_n))$$

which is equivalent to

$$\mathcal{B} \vDash (\varphi * \psi)(h(a_1), \ldots, h(a_n))$$

Now assume that the implication holds for a formula $\varphi(x, x_1, \ldots, x_n)$ and choose arbitrary elements $a_1, \ldots, a_n \in A$. Then the condition

$$\mathcal{A} \vDash (\mathrm{Q}\, x)\varphi(x, a_1, \ldots, a_n)$$

is equivalent to

$$(\mathrm{Q}\, a \in A)\big(\mathcal{A} \vDash \varphi(a, a_1, \ldots, a_n)\big)$$

By the induction hypothesis, this implies

$$(\mathrm{Q}\, a \in A)\big(\mathcal{B} \vDash \varphi(h(a), h(a_1), \ldots, h(a_n))\big)$$

From this, we obtain

$$(\mathrm{Q}\, b \in B)\big(\mathcal{B} \vDash \varphi(b, h(a_1), \ldots, h(a_n))\big)$$

For the existential quantifier $\exists$, the result is clear: we can simply take $b = h(a)$. An analogous conclusion for the universal quantifier $\forall$ follows from the surjectivity of h, as every $b \in B$ has the form $b = h(a)$ for some $a \in A$. In both cases, we conclude

$$\mathcal{B} \vDash (\mathrm{Q}\, x)\varphi(x, h(a_1), \ldots, h(a_n))$$

6.2.6 Corollary. *Let the structure $\mathcal{B}$ be a homomorphic image of the structure $\mathcal{A}$. Then, for any closed positive formula φ, from $\mathcal{A} \models \varphi$ it follows that $\mathcal{B} \models \varphi$.*

In other words, positive properties are preserved under homomorphic images.

A closer examination of the proof of Proposition 6.2.5 reveals that the surjectivity condition of the mapping φ was used only in the inductive step for the universal quantifier. This simultaneously proves the following result.

6.2.7 Proposition. *Let $\mathcal{A}$, $\mathcal{B}$ be first-order structures and let $h\colon \mathcal{A} \to \mathcal{B}$ be a homomorphism. Then, for any existential positive formula $\varphi(x_1, \ldots, x_n)$ and all $a_1, \ldots, a_n \in A$, it holds that*

$$\mathcal{A} \models \varphi(a_1, \ldots, a_n) \;\Rightarrow\; \mathcal{B} \models \varphi(h(a_1), \ldots, h(a_n))$$

6.2.8 Exercise. Consider groups as structures $(G; \cdot)$ in a language with a single binary operation symbol $\cdot$. Express by formulas of this language the property $\varepsilon(u)$: "u is the identity element of the operation $\cdot$", and the relation $\iota(x, y)$: "y is the inverse element of x with respect to the operation $\cdot$". Verify that both $\varepsilon(u)$ and $\iota(x, y)$ are (universal) positive formulas. Is this fact alone sufficient to explain the implication (iii) $\Rightarrow$ (i) from Exercise 6.2.2? Under what circumstances is it sufficient, and under what circumstances is it not? Think about how we can "rescue the situation" in this case.

From Proposition 6.2.5, it follows, among other things, that properties that are not preserved under homomorphic images cannot be expressed using positive formulas.

6.2.9 Exercise. A ring $(A; +, \cdot, 0)$ is called an *integral domain* if it satisfies the condition

$$(\forall\, x, y)(x \cdot y = 0 \;\Rightarrow\; (x = 0 \;\vee\; y = 0))$$

Prove that the property of rings "being an integral domain" cannot be expressed by a positive formula. Find other examples of properties of first-order structures that cannot be expressed by positive formulas.

Later on, we will see that, conversely, first-order properties that are preserved under homomorphic images can be expressed using positive formulas.

A formula φ is called *negative* if it is the negation of a positive formula. The proofs of the following results on the backward transfer of negative properties from homomorphic images to the original structure are left as an exercise for the reader.

6.2.10 Proposition. *Let $\mathcal{A}$, $\mathcal{B}$ be first-order structures, and let $h\colon \mathcal{A} \to \mathcal{B}$ be a surjective homomorphism. Then, for any negative formula $\varphi(x_1, \ldots, x_n)$ and all elements $a_1, \ldots, a_n \in A$, it holds that*

$$\mathcal{B} \models \varphi(h(a_1), \ldots, h(a_n)) \;\Rightarrow\; \mathcal{A} \models \varphi(a_1, \ldots, a_n)$$

If φ is additionally closed, then from $\mathcal{B} \models \varphi$ it follows that $\mathcal{A} \models \varphi$.

6.2.11 Proposition. *Let $\mathcal{A}$, $\mathcal{B}$ be first-order structures, and let $h\colon \mathcal{A} \to \mathcal{B}$ be a homomorphism. Then, for any universal negative formula $\varphi(x_1, \ldots, x_n)$ and all elements $a_1, \ldots, a_n \in A$ it holds that*

$$\mathcal{B} \vDash \varphi(h(a_1), \ldots, h(a_n)) \;\Rightarrow\; \mathcal{A} \vDash \varphi(a_1, \ldots, a_n)$$

If φ is additionally closed, then from $\mathcal{B} \vDash \varphi$ it follows that $\mathcal{A} \vDash \varphi$.

6.3 Isomorphisms and Embeddings

A homomorphism $h\colon \mathcal{A} \to \mathcal{B}$ is called an *isomorphism* if h is a bijective mapping and its inverse mapping $h^{-1}\colon B \to A$ is also a homomorphism $h^{-1}\colon \mathcal{B} \to \mathcal{A}$. The fact that h is an isomorphism from the structure $\mathcal{A}$ to the structure $\mathcal{B}$ is denoted by $h\colon \mathcal{A} \xrightarrow{\sim} \mathcal{B}$. We say that structures the $\mathcal{A}$ and $\mathcal{B}$ are *isomorphic* if there exists at least one isomorphism $h\colon \mathcal{A} \xrightarrow{\sim} \mathcal{B}$; we write $\mathcal{A} \cong \mathcal{B}$.

The following observation is evident.

6.3.1 Proposition. *Let $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ be first-order structures.*

(a) *The identity mapping $\mathrm{Id}_A\colon A \to A$ is an isomorphism $\mathrm{Id}_A\colon \mathcal{A} \xrightarrow{\sim} \mathcal{A}$.*

(b) *If $h\colon \mathcal{A} \xrightarrow{\sim} \mathcal{B}$ is an isomorphism, then its inverse mapping $h^{-1}\colon \mathcal{B} \to \mathcal{A}$ is also an isomorphism.*

(c) *If both $h\colon \mathcal{A} \xrightarrow{\sim} \mathcal{B}$ and $g\colon \mathcal{B} \xrightarrow{\sim} \mathcal{C}$ are isomorphisms, then the composite mapping $g \circ h\colon \mathcal{A} \xrightarrow{\sim} \mathcal{C}$ is also an isomorphism.*

Consequently, for any structures $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$, the following holds:

$$\mathcal{A} \cong \mathcal{A}$$
$$\mathcal{A} \cong \mathcal{B} \;\Rightarrow\; \mathcal{B} \cong \mathcal{A}$$
$$\mathcal{A} \cong \mathcal{B} \wedge \mathcal{B} \cong \mathcal{C} \;\Rightarrow\; \mathcal{A} \cong \mathcal{C}$$

That is, the isomorphism relation is *reflexive*, *symmetric*, and *transitive*; in other words, it is an *equivalence relation* on the class of all structures of the language L.

The reader should note that while every isomorphism is a bijective homomorphism, not every bijective homomorphism is an isomorphism.

6.3.2 Example. Let the language L contain a single specific symbol — a unary predicate symbol P. Further, let $\mathcal{A}_1 = (A; P_1)$ and $\mathcal{A}_2 = (A; P_2)$ be two structures of the language L with the same underlying set A, where the subsets $P_1, P_2 \subseteq A$ serve as interpretations of the symbol P in the structures $\mathcal{A}_1$ and $\mathcal{A}_2$, respectively. Then, the identity mapping $\mathrm{Id}_A\colon A \to A$ is certainly bijective. It is a homomorphism $\mathrm{Id}_A\colon \mathcal{A}_1 \to \mathcal{A}_2$ precisely when $P_1 \subseteq P_2$, and it is an isomorphism $\mathrm{Id}_A\colon \mathcal{A}_1 \xrightarrow{\sim} \mathcal{A}_2$ precisely when $P_1 = P_2$. Thus, if P_1 is a proper subset of P_2, then $\mathrm{Id}_A\colon \mathcal{A}_1 \to \mathcal{A}_2$ serves as an example of a bijective homomorphism that is not an isomorphism.

Under certain additional conditions, however, every bijective homomorphism is indeed an isomorphism.

6.3.3 Proposition. *Assume that the language L contains no relational symbols. Then every bijective homomorphism $h\colon \mathcal{A} \to \mathcal{B}$ between structures of the language L is an isomorphism.*

Proof. Let $h\colon \mathcal{A} \to \mathcal{B}$? be a bijective homomorphism. It suffices to verify that for every n-ary operation symbol $f \in F$ and all $b_1, \ldots, b_n \in B$, the following holds:

$$h^{-1}\big(f^{\mathcal{B}}(b_1, \ldots, b_n)\big) = f^{\mathcal{A}}\big(h^{-1}(b_1), \ldots, h^{-1}(b_n)\big)$$

Let $a_i = h^{-1}(b_i)$ for $1 \leq i \leq n$; then $b_i = h(a_i)$. Since h is a homomorphism and $h^{-1} \circ h = \mathrm{Id}_A$, we have:

$$\begin{aligned} h^{-1}\big(f^{\mathcal{B}}(b_1, \ldots, b_n)\big) &= h^{-1}\big(f^{\mathcal{B}}(h(a_1), \ldots, h(a_n))\big) \\ &= \big(h^{-1} \circ h\big)\big(f^{\mathcal{A}}(a_1, \ldots, a_n)\big) \\ &= f^{\mathcal{A}}\big(h^{-1}(b_1), \ldots, h^{-1}(b_n)\big) \end{aligned}$$

6.3.4 Corollary. *Let $\mathcal{A}$ and $\mathcal{B}$ be first-order structures. Then a mapping $h\colon A \to B$ is an isomorphism $h\colon \mathcal{A} \xrightarrow{\sim} \mathcal{B}$ if and only if it is bijective and for every n-ary functional symbol $f \in F$ or relational symbol $r \in R$, and all $a_1, \ldots, a_n \in A$, the following hold:*

$$\begin{aligned} &h\big(f^{\mathcal{A}}(a_1, \ldots, a_n)\big) = f^{\mathcal{B}}(h(a_1), \ldots, h(a_n)) \\ &(a_1, \ldots, a_n) \in r^{\mathcal{A}} \;\Leftrightarrow\; (h(a_1), \ldots, h(a_n)) \in r^{\mathcal{B}} \end{aligned}$$

A mapping $h\colon A \to B$ between the underlying sets of structures $\mathcal{A}$ and $\mathcal{B}$ is called an *embedding* of structure $\mathcal{A}$ into structure $\mathcal{B}$ if h is an isomorphism of $\mathcal{A}$ onto the substructure $h[A]$ of $\mathcal{B}$. In such a case, we write $h\colon \mathcal{A} \hookrightarrow \mathcal{B}$.

A fundamental and simplest example of an embedding is the embedding of a substructure $\mathcal{B} \subseteq \mathcal{A}$ into the structure $\mathcal{A}$ via the identity mapping $\mathrm{Id}_B\colon B \hookrightarrow A$. It is also evident that the composition of embeddings $h\colon \mathcal{A} \hookrightarrow \mathcal{B}$ and $g\colon \mathcal{B} \hookrightarrow \mathcal{C}$ is itself an embedding $g \circ h\colon \mathcal{A} \hookrightarrow \mathcal{C}$.

From the preceding Corollary 6.3.4, the following more detailed characterization of embeddings immediately follows.

6.3.5 Proposition. *Let $\mathcal{A}$ and $\mathcal{B}$ be first-order structures. Then a mapping $h\colon A \to B$ is an embedding $h\colon \mathcal{A} \hookrightarrow \mathcal{B}$ if and only if it is injective and for every n-ary functional symbol $f \in F$ or relational symbol $r \in R$, and all $a_1, \ldots, a_n \in A$, the following hold:*

$$\begin{aligned} &h\big(f^{\mathcal{A}}(a_1, \ldots, a_n)\big) = f^{\mathcal{B}}(h(a_1), \ldots, h(a_n)) \\ &(a_1, \ldots, a_n) \in r^{\mathcal{A}} \;\Leftrightarrow\; (h(a_1), \ldots, h(a_n)) \in r^{\mathcal{B}} \end{aligned}$$

It follows that if the language L contains no relational symbols, then every injective homomorphism is an embedding.

Moreover, note that the very property of injectivity of the mapping h

$$a = b \;\Leftrightarrow\; h(a) = h(b)$$

for all $a, b \in A$, is nothing but the condition of preserving the equality relation in both directions. While the implication $\Rightarrow$ is trivially satisfied "always", injectivity is given precisely by the reverse implication $\Leftarrow$.

6.3.6 Exercise. Let $\mathcal{A}$, $\mathcal{B}$ be first-order structures. Prove that a mapping $h\colon A \to B$ is an embedding $h\colon \mathcal{A} \hookrightarrow \mathcal{B}$ if and only if for any atomic formula $\varphi(x_1, \ldots, x_n)$ and all $a_1, \ldots, a_n \in A$, the following holds:

$$\mathcal{A} \vDash \varphi(a_1, \ldots, a_n) \;\Leftrightarrow\; \mathcal{B} \vDash \varphi(h(a_1), \ldots, h(a_n))$$

6.3.7 Proposition. *Let $(A; <)$ and $(B; <)$ be partially ordered sets. If $(A; <)$ is linearly ordered, then every homomorphism $h\colon (A; <) \to (B; <)$ is also an embedding. If h is additionally surjective, then it is an isomorphism $h\colon (A; <) \xrightarrow{\sim} (B; <)$.*

Proof. Let $h\colon (A; <) \to (B; <)$ be a homomorphism. First, we show that h is injective. Choose two distinct elements $a, b \in A$. Since the ordering $<$ on set A is linear, either $a < b$ or $b < a$ holds. In the first case, we have $h(a) < h(b)$; in the second, $h(b) < h(a)$. In either case, $h(a) \neq h(b)$. Next, we need to verify that for any $a, b \in A$, the condition $h(a) < h(b)$ implies $a < b$. Since h is injective, $h(a) < h(b)$ implies $a \neq b$. Then either $a < b$ or $b < a$. However, the latter would lead to $h(b) < h(a)$, contradicting the assumption $h(a) < h(b)$. From this, the second part of the proposition immediately follows.

6.3.8 Exercise. Formulate and prove an analogous statement for homomorphisms and embeddings of partially ordered sets with the partial order $\leq$.

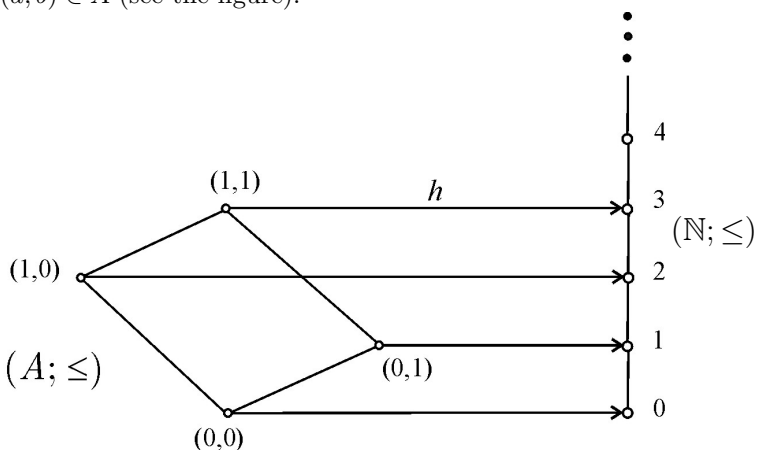
6.3.9 Example. Let $(A; \leq)$ be the set $A = \{0, 1\} \times \{0, 1\}$ equipped with the componentwise partial order, defined as:

$$(a_1, b_1) \leq (a_2, b_2) \;\Leftrightarrow\; a_1 \leq a_2 \wedge b_1 \leq b_2$$

for all $(a_1, b_1), (a_2, b_2) \in A$. Let $(\mathbb{N}; \leq)$ be the set of all natural numbers with the usual (linear) order. Define a mapping $h\colon A \to \mathbb{N}$ by:

$$h(a, b) = 2a + b$$

for $(a, b) \in A$ (see the figure).



At first glance, it is evident that $h\colon (A;\leq) \to (\mathbb{N};\leq)$ is an injective homomorphism but not an embedding. In $(\mathbb{N};\leq)$, we have $h(0,1) = 1 \leq 2 = h(1,0)$, yet in $(A;\leq)$, $(0,1) \leq (1,0)$ does not hold — the elements $(0,1)$ and $(1,0)$ are namely incomparable.

An immediate generalization of Corollary 6.1.5 is as follows.

6.3.10 Proposition. *Let $\mathcal{A}$ and $\mathcal{B}$ be first-order structures, and let $h\colon \mathcal{A} \hookrightarrow \mathcal{B}$ be an embedding.*

(a) *If φ is a closed universal formula, then from $\mathcal{B} \vDash \varphi$ it follows that $\mathcal{A} \vDash \varphi$.*

(b) *If φ is a closed existential formula, then from $\mathcal{A} \vDash \varphi$ it follows that $\mathcal{B} \vDash \varphi$.*

6.3.11 Exercise. (a) Formulate (and if you feel the need, prove) an analogous generalization of Proposition 6.1.4.

(b) Which first-order properties are preserved under isomorphisms?

6.4 Elementary Equivalence and Elementary Substructures

Isomorphic L-structures $\mathcal{A}$ and $\mathcal{B}$ can be considered "much the same," and if necessary, identified using the corresponding isomorphism. However, for many purposes in model theory, isomorphism is "too strict" an equivalence. From a certain perspective, structures $\mathcal{A}$ and $\mathcal{B}$ can be regarded as "the same" if they share the same properties expressible in the language L.

We say that structures $\mathcal{A}$ and $\mathcal{B}$ are *elementarily equivalent*, denoted $\mathcal{A} \equiv \mathcal{B}$, if for every closed formula φ:

$$\mathcal{A} \vDash \varphi \;\Leftrightarrow\; \mathcal{B} \vDash \varphi$$

It is immediately clear that elementary equivalence is reflexive, symmetric, and transitive, i.e., for any structures $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$:

$$\mathcal{A} \equiv \mathcal{B}$$
$$\mathcal{A} \equiv \mathcal{B} \;\Rightarrow\; \mathcal{B} \equiv \mathcal{A}$$
$$\mathcal{A} \equiv \mathcal{B} \wedge \mathcal{B} \equiv \mathcal{C} \;\Rightarrow\; \mathcal{A} \equiv \mathcal{C}$$

making it an equivalence relation on the class of all structures of language L. It is also evident that isomorphic structures are elementarily equivalent, i.e.,

$$\mathcal{A} \cong \mathcal{B} \;\Rightarrow\; \mathcal{A} \equiv \mathcal{B}$$

for any $\mathcal{A}$ and $\mathcal{B}$. We will soon see examples of elementarily equivalent structures that are not isomorphic (e.g., due to differing cardinalities).

6.4.1 Exercise. Let at least one of the structures $\mathcal{A}$, $\mathcal{B}$ be finite. Then $\mathcal{A} \equiv \mathcal{B}$ if and only if $\mathcal{A} \cong \mathcal{B}$. Prove this. (*Hint*: First, prove that if, for example, the structure $\mathcal{A}$ is finite and $|A| = n$, then $\mathcal{A} \equiv \mathcal{B}$ implies $|B| = n$. Then try to prove the implication $\mathcal{A} \equiv \mathcal{B} \Rightarrow \mathcal{A} \cong \mathcal{B}$ under the additional assumption that the language L has only finitely many specific symbols. Finally, try to reduce the general case to this situation.)

The *theory of the structure* $\mathcal{A}$ is defined as the set

$$\mathrm{Th}(\mathcal{A}) = \{\varphi \in \mathrm{Form}(L) \colon \varphi \text{ is closed and } \mathcal{A} \vDash \varphi\}$$

of all closed formulas in the language L that hold in the structure $\mathcal{A}$. Clearly, for any structures $\mathcal{A}$, $\mathcal{B}$, we have

$$\mathcal{A} \equiv \mathcal{B} \;\Leftrightarrow\; \mathrm{Th}(\mathcal{A}) = \mathrm{Th}(\mathcal{B})$$

Since for any structure $\mathcal{A}$, the theory $\mathrm{Th}(\mathcal{A})$ is complete, the following three conditions are equivalent for any structures $\mathcal{A}$, $\mathcal{B}$:

$$\mathrm{Th}(\mathcal{A}) = \mathrm{Th}(\mathcal{B}), \quad \mathrm{Th}(\mathcal{A}) \subseteq \mathrm{Th}(\mathcal{B}), \quad \mathrm{Th}(\mathcal{A}) \supseteq \mathrm{Th}(\mathcal{B})$$

This means that the logical equivalence in the definition of elementary equivalence can be replaced by either of the implications $\Rightarrow$, $\Leftarrow$.

6.4.2 Proposition. *For any structures $\mathcal{A}$, $\mathcal{B}$, the following conditions are equivalent:*

(i) $\mathcal{A} \equiv \mathcal{B}$

(ii) *For every closed formula φ, we have* $\mathcal{A} \vDash \varphi \;\Rightarrow\; \mathcal{B} \vDash \varphi$

(iii) *For every closed formula φ, we have* $\mathcal{B} \vDash \varphi \;\Rightarrow\; \mathcal{A} \vDash \varphi$.

The same conclusion can be reached based on the fact that the system of all closed formulas in the language L is closed under negation. Think about how.

An even closer relation of similarity between structures is the notion of an elementary substructure. We say that a structure $\mathcal{A}$ is an *elementary substructure* of a structure $\mathcal{B}$, or equivalently, that the structure $\mathcal{B}$ is an *elementary extension* of the structure $\mathcal{A}$, if $A \subseteq B$ and for *every* formula $\varphi(x_1, \ldots, x_n)$ in the language L and all $a_1, \ldots, a_n \in A$, we have

$$\mathcal{A} \vDash \varphi(a_1, \ldots, a_n) \;\Leftrightarrow\; \mathcal{B} \vDash \varphi(a_1, \ldots, a_n)$$

This fact is denoted symbolically as $\mathcal{A} \prec \mathcal{B}$ or $\mathcal{B} \succ \mathcal{A}$. Consider that from the condition $\mathcal{A} \prec \mathcal{B}$, it follows that $\mathcal{A} \subseteq \mathcal{B}$, meaning every elementary substructure of $\mathcal{B}$ is indeed a substructure of it.

6.4.3 Exercise. Verify that for any structures $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, we have

$$\mathcal{A} \prec \mathcal{A}$$
$$\mathcal{A} \prec \mathcal{B} \Rightarrow \mathcal{A} \equiv \mathcal{B}$$
$$\mathcal{A} \prec \mathcal{B} \wedge \mathcal{B} \prec \mathcal{C} \;\Rightarrow\; \mathcal{A} \prec \mathcal{C}$$
$$\mathcal{A} \prec \mathcal{C} \wedge \mathcal{B} \prec \mathcal{C} \wedge \mathcal{A} \subseteq \mathcal{B} \;\Rightarrow\; \mathcal{A} \prec \mathcal{B}$$

Since the system of all L-formulas is also closed under negation, logical equivalence can be replaced by any of the implications $\Rightarrow$, $\Leftarrow$ in the definition of an elementary substructure.

6.4.4 Proposition. *For any structures $\mathcal{A}$, $\mathcal{B}$ such that $A \subseteq B$, the following conditions are equivalent:*

(i) $\mathcal{A} \prec \mathcal{B}$

(ii) *For any L-formula $\varphi(x_1, \ldots, x_n)$ and all $a_1, \ldots, a_n \in A$, we have:*

$$\mathcal{A} \vDash \varphi(a_1, \ldots, a_n) \;\Rightarrow\; \mathcal{B} \vDash \varphi(a_1, \ldots, a_n)$$

(iii) *For any L-formula $\varphi(x_1, \ldots, x_n)$ and all $a_1, \ldots, a_n \in A$, we have:*

$$\mathcal{B} \vDash \varphi(a_1, \ldots, a_n) \;\Rightarrow\; \mathcal{A} \vDash \varphi(a_1, \ldots, a_n)$$

6.4.5 Example. The field $\mathcal{Q} = (\mathbb{Q}; +\,\cdot, 0, 1)$ of all rational numbers is a substructure of the field $\mathcal{R} = (\mathbb{R}; +, \cdot, 0, 1)$ of all real numbers. However, these fields are not elementarily equivalent. Since $\sqrt{2}$ does not exist in $\mathbb{Q}$, we have:

$$\mathcal{Q} \vDash \neg(\exists\, x)(x \cdot x = 1 + 1)$$

whereas $\sqrt{2} \in \mathbb{R}$, thus

$$\mathcal{R} \vDash (\exists\, x)(x \cdot x = 1 + 1)$$

Therefore, $\mathcal{Q} \prec \mathcal{R}$ cannot hold. Analogously, try to explain why the field $\mathcal{R}$ is not elementarily equivalent to the field $(\mathbb{C}; +\,\cdot, 0, 1)$ of all complex numbers, hence it cannot be its elementary substructure. On the other hand, as proved by Alfred Tarski in 1930, the field $(\mathbb{A}; +, \cdot, 0, 1)$ of all algebraic numbers (i.e., solutions of polynomial equations with rational coefficients) is an elementary substructure of the field $(\mathbb{C}; +\,\cdot, 0, 1)$. Similarly, the field $(\mathbb{A} \cap \mathbb{R}; +, \cdot, 0, 1)$ of all real algebraic numbers is an elementary substructure of $\mathcal{R}$. Moreover, since the sets $\mathbb{R}$ and $\mathbb{C}$ have the cardinality of the continuum, while the sets $\mathbb{A} \cap \mathbb{R}$ and $\mathbb{A}$ are countable, the elementarily equivalent fields $(\mathbb{A}; +, \cdot, 0, 1) \equiv (\mathbb{C}; +, \cdot, 0, 1)$, as well as $(\mathbb{A} \cap \mathbb{R}; +, \cdot, 0, 1) \equiv (\mathbb{R}; +, \cdot, 0, 1)$ cannot be isomorphic.

The following result, known as the Tarski or also the Tarski-Vaught criterion, essentially localizes the question of whether $\mathcal{A} \prec \mathcal{B}$ exclusively within the "larger" of the two structures $\mathcal{B} \supseteq \mathcal{A}$.

6.4.6 Theorem. *Let $\mathcal{A}$, $\mathcal{B}$ be arbitrary structures, with $\mathcal{A} \subseteq \mathcal{B}$. Then the following conditions are equivalent:*

(i) $\mathcal{A} \prec \mathcal{B}$

(ii) *For any formula $\varphi(x, x_1, \ldots, x_n)$ and all $a_1, \ldots, a_n \in A$, we have:*

$$(\exists\, b \in B)\big(\mathcal{B} \vDash \varphi(b, a_1, \ldots, a_n)\big) \;\Rightarrow\; (\exists\, a \in A)\big(\mathcal{B} \vDash \varphi(a, a_1, \ldots, a_n)\big)$$

Condition (ii) means: if the problem "find x such that $\mathcal{B} \vDash \varphi(x, a_1, \ldots, a_n)$" with parameters $a_1, \ldots, a_n \in A$ has at least one solution $b \in B$, then this problem also has some solution $a \in A$.

Proof. (i) $\Rightarrow$ (ii): Let $\mathcal{A} \prec \mathcal{B}$. Assume that $\varphi(x, x_1, \ldots, x_n)$ is a formula and $a_1, \ldots, a_n \in A$ are such that $(\exists b \in B)\big(\mathcal{B} \vDash \varphi(b, a_1, \ldots, a_n)\big)$, i.e., $\mathcal{B} \vDash (\exists x)\varphi(x, a_1, \ldots, a_n)$.

Then also $\mathcal{A} \vDash (\exists x)\varphi(x, a_1, \ldots, a_n)$, i.e., $(\exists a \in A)\big(\mathcal{A} \vDash \varphi(a, a_1, \ldots, a_n)\big)$. For such a, we also have $\mathcal{B} \vDash \varphi(a, a_1, \ldots, a_n)$, hence $(\exists a \in A)\big(\mathcal{B} \vDash \varphi(a, a_1, \ldots, a_n)\big)$.

(ii) $\Rightarrow$ (i): Assume (ii). By induction on complexity, we prove that for every formula $\varphi(x_1, \ldots, x_n)$ and all $a_1, \ldots, a_n \in A$,

$$\mathcal{A} \vDash \varphi(a_1, \ldots, a_n) \;\Leftrightarrow\; \mathcal{B} \vDash \varphi(a_1, \ldots, a_n)$$

For atomic φ, this holds because $\mathcal{A} \subseteq \mathcal{B}$. The inductive steps for logical connectives $\neg$ and $\wedge$ are straightforward and left to the reader. Thus, we focus only on the inductive step for the existential quantifier $\exists$. Assume that the above equivalence holds for the formula $\varphi(x, x_1, \ldots, x_n)$ and all $a, a_1, \ldots, a_n \in A$. We need to verify its validity for the formula $(\exists x)\varphi(x, x_1, \ldots, x_n)$ and all $a_1, \ldots, a_n \in A$. The following conditions are equivalent:

$$\mathcal{A} \vDash (\exists\, x)\varphi(x, a_1, \ldots, a_n) \tag{1}$$

$$(\exists\, a \in A)\big(\mathcal{A} \vDash \varphi(a, a_1, \ldots, a_n)\big) \tag{2}$$

$$(\exists\, a \in A)\big(\mathcal{B} \vDash \varphi(a, a_1, \ldots, a_n)\big) \tag{3}$$

$$(\exists\, b \in B)\big(\mathcal{B} \vDash \varphi(b, a_1, \ldots, a_n)\big) \tag{4}$$

$$\mathcal{B} \vDash (\exists\, x)\varphi(x, a_1, \ldots, a_n) \tag{5}$$

To explain: the equivalences (1) $\Leftrightarrow$ (2) and (4) $\Leftrightarrow$ (5) follow from the definition of formula satisfaction in structures, the equivalence (2) $\Leftrightarrow$ (3) follows from the induction hypothesis, the implication (3) $\Rightarrow$ (4) is trivial, while the reverse implication (4) $\Rightarrow$ (3) is precisely condition (ii) of the theorem being proved.

A mapping $h\colon A \to B$ between the underlying sets of structures $\mathcal{A}$, $\mathcal{B}$ is called an *elementary embedding* of $\mathcal{A}$ into $\mathcal{B}$ if h is an isomorphism of $\mathcal{A}$ onto the *elementary* substructure $h[A]$ of $\mathcal{B}$. In this case, we write $h\colon \mathcal{A} \xrightarrow{\prec} \mathcal{B}$.

The simplest example of an elementary embedding is the embedding of an elementary substructure $\mathcal{B} \prec \mathcal{A}$ into $\mathcal{A}$ via the identity mapping $\mathrm{Id}_B\colon B \to A$. It is also clear that the composition of elementary embeddings $h\colon \mathcal{A} \xrightarrow{\prec} \mathcal{B}$, $g\colon \mathcal{B} \xrightarrow{\prec} \mathcal{C}$ is itself an elementary embedding $g \circ h\colon \mathcal{A} \xrightarrow{\prec} \mathcal{C}$.

6.4.7 Exercise. (a) Formulate and prove generalizations of Proposition 6.4.4 and Theorem 6.4.6 from elementary substructures to elementary embeddings.

(b) Let $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ be structures, $h\colon \mathcal{A} \to \mathcal{B}$ an embedding, and $g\colon \mathcal{B} \xrightarrow{\prec} \mathcal{C}$ an elementary embedding such that the composition $g \circ h\colon \mathcal{A} \xrightarrow{\prec} \mathcal{C}$ is also an elementary embedding. Prove that then $h\colon \mathcal{A} \xrightarrow{\prec} \mathcal{B}$ is also an elementary embedding. Which previously stated fact about elementary substructures does this statement generalize?

6.5 Diagrams of Structures

In this section, we briefly introduce the *diagram method*, based on expanding the basic language with new constant symbols. This method allows us to construct new

structures with prescribed properties from given ones, reflecting certain properties of the original structures as models of appropriate theories in an extended language.

Let L be a first-order language, $\mathcal{A} = (A; \dots)$ an L-structure, and $M \subseteq A$. The language obtained from L by extending it with a set of new constants $\{c_a \colon a \in M\}$ (where the constant symbols c_a corresponding to different elements of M are distinct) will be denoted by L_M. The structure of the language L_M, which arises from the structure $\mathbb{A}$ by keeping the interpretations of the original symbols of L unchanged and interpreting each new constant c_a as the element $a \in M$, is denoted as

$$\mathcal{A}_M = (\mathcal{A}, a)_{a \in M}$$

The most important special case for us is when $M = A$; then $\mathcal{A}_A = (\mathcal{A}, a)_{a \in A}$. Similarly, if $\mathcal{B}$ is another L-structure and $h \colon A \to B$ is some mapping, then $\big(\mathcal{B}, h(a)\big)_{a \in A}$ denotes the structure of the language L_A, in which the individual constants c_a are interpreted as elements $h(a)$ (and the original symbols of the language L are interpreted the same way as in $\mathcal{B}$).

For brevity's sake, negations of atomic formulas will be referred to as *negatomic* ones. The *atomic diagram*, or simply the *diagram* of the L-structure $\mathcal{A}$, is defined as the set of formulas

$$\mathrm{D}(\mathcal{A}) = \{\varphi \in \mathrm{Form}(L_A) \colon \varphi \text{ is a closed atomic or negatomic formula and } \mathcal{A} \vDash \varphi\}.$$

Note that every closed formula φ of the language L_A has the form $\psi(c_{a_1}, \dots, c_{a_n})$ for some formula $\psi(x_1, \dots, x_n)$ of the original language L. In this case, the condition $\mathcal{A}_A \vDash \varphi$, i.e.,

$$\mathcal{A}_A \vDash \psi(c_{a_1}, \dots, c_{a_n})$$

is equivalent to the condition

$$\mathcal{A} \vDash \psi(a_1, \dots, a_n)$$

so we may freely switch between the second-to-last and last notations.

The positive atomic diagram of the L-structure $\mathcal{A}$ is defined as the set of formulas

$$\mathrm{D}^+(\mathcal{A}) = \{\varphi \in \mathrm{Form}(L_A) : \varphi \text{ is a closed atomic formula and } \mathcal{A}_A \vDash \varphi\}.$$

Clearly, we have $\mathrm{D}^+(\mathcal{A}) \subseteq \mathrm{D}(\mathcal{A})$, where the "smaller" of the two diagrams, $\mathrm{D}^+(\mathcal{A})$, consists of all atomic formulas from the "larger" diagram $\mathrm{D}(\mathcal{A})$. It thus seems that the "larger" diagram $\mathrm{D}(\mathcal{A})$ contains more information than the "smaller" one, $\mathrm{D}^+(\mathcal{A})$. However, this is not the case — in fact, both diagrams contain the same amount of information. The "larger" diagram $\mathrm{D}(\mathcal{A})$ can be reconstructed from the "smaller" one $\mathrm{D}^+(\mathcal{A})$ in a straightforward way. From each pair of formulas $\varphi(a_1, \dots, a_n)$, $\neg\varphi(a_1, \dots, a_n)$, where $\varphi(x_1, \dots, x_n)$ is an atomic L-formula and $a_1, \dots, a_n \in A$, precisely one belongs to the diagram $\mathrm{D}(\mathcal{A})$. If $\varphi(a_1, \dots, a_n) \in D^+(\mathcal{A})$, then it is $\varphi(a_1, \dots, a_n)$; if $\varphi(a_1, \dots, a_n) \notin D^+(\mathcal{A})$, then it is $\neg\varphi(a_1, \dots, a_n)$. The reasons for introducing the seemingly redundant and unnecessarily large diagram $\mathrm{D}(\mathcal{A})$ will become apparent shortly.

6.5.1 Example. Let $\mathcal{A} = (A; \cdot)$ be a finite set with a single binary operation $\cdot$. Then the diagrams $\mathrm{D}(\mathcal{A})$, $\mathrm{D}^+(\mathcal{A})$ contain the same information as the Cayley multiplication table of the structure $\mathcal{A}$. An element $c \in A$ appears in the row of element $a \in A$ and the column of element $b \in A$ in this table precisely when $a \cdot b = c$ holds in $\mathcal{A}$.

The following statements express some of the already known relationships between structures in terms of the newly introduced diagrams.

6.5.2 Proposition. *Let $\mathcal{A}$, $\mathcal{B}$ be structures of the language L.*
(a) *Let $A \subseteq B$. Then $\mathcal{A} \subseteq \mathcal{B}$ if and only if $\mathcal{B}_A \models \mathrm{D}(\mathcal{A})$.*
(b) *Let $h\colon A \to B$. Then h is an embedding of $\mathcal{A}$ into $\mathcal{B}$ if and only if $\big(\mathcal{B}, h(a)\big)_{a\in A} \models \mathrm{D}(\mathcal{A})$.*

6.5.3 Proposition. *Let $\mathcal{A}$, $\mathcal{B}$ be structures of the language L, and let $h\colon A \to B$. Then h is a homomorphism from $\mathcal{A}$ to $\mathcal{B}$ if and only if $\big(\mathcal{B}, h(a)\big)_{a\in A} \models \mathrm{D}^+(\mathcal{A})$.*

From the couple of the above propositions, we can clearly see the different roles played by the diagrams $\mathrm{D}(\mathcal{A})$ and $\mathrm{D}^+(\mathcal{A})$.

If we do not restrict ourselves only to atomic formulas, we can introduce another type of diagrams. The *elementary* or *complete diagram* of an L-structure $\mathcal{A}$ is the set of formulas

$$\mathrm{Th}(\mathcal{A}_A) = \{\varphi \in \mathrm{Form}(L_A)\colon \varphi \text{ is closed and } \mathcal{A}_A \models \varphi\}$$

The following proposition is a direct analogue of Proposition 6.5.2.

6.5.4 Proposition. *Let $\mathcal{A}$, $\mathcal{B}$ be structures of the language L.*
(a) *Let $A \subseteq B$. Then $\mathcal{A} \prec \mathcal{B}$ if and only if $\mathcal{B}_A \models \mathrm{Th}(\mathcal{A}_A)$.*
(b) *Let $h\colon A \to B$. Then h is an elementary embedding of $\mathcal{A}$ into $\mathcal{B}$ if and only if $\big(\mathcal{B}, h(a)\big)_{a\in A} \models \mathrm{Th}(\mathcal{A}_A)$.*

For completeness, let us also recall that the *theory of the structure* $\mathcal{A}$ in the language L

$$\mathrm{Th}(\mathcal{A}) = \{\varphi \in \mathrm{Form}(L)\colon \varphi \text{ is closed and } \mathcal{A} \models \varphi\}$$

can be considered as a type of diagram that expresses the relationship of elementary equivalence.

6.5.5 Proposition. *Let $\mathcal{A}$, $\mathcal{B}$ be structures of the language L. Then $\mathcal{A} \equiv \mathcal{B}$ if and only if $\mathcal{B} \models \mathrm{Th}(\mathcal{A})$.*

6.5.6 Exercise. Using an extension of the language L by appropriate constant symbols, simplify the formulation of condition (ii) in the Tarski-Vaught criterion for elementary substructures and simplify the notation of some parts of its proof.

6.6 The Löwenheim-Skolem-Tarski Theorems

In this section, we will prove two results — the Löwenheim-Skolem-Tarski "downward" and "upward" theorems — which guarantee the existence of many elementary substructures or elementary extensions of a given infinite structure $\mathcal{A}$ with a prescribed cardinality, satisfying certain minimal constraints formulated in terms of the cardinality of the language L and of the structure $\mathcal{A}$ itself. We will make use of extensions of the original language by appropriate constant symbols, and in the proof of the second theorem, we will also encounter a simple application of the diagram method (namely, the elementary diagram $\mathrm{Th}(\mathcal{A}_A)$ of the structure $\mathcal{A}$).

Recall that the *cardinality of the language* $L = (F, R, \nu)$ is the cardinal number

$$\|L\| = |\operatorname{Form}(L)| = \max(|F|, |R|, \aleph_0)$$

6.6.1 Löwenheim-Skolem-Tarski Downward Theorem. *Let $\mathcal{A}$ be a structure of infinite cardinality $|A| = \alpha \geq |L|$ and β be a cardinal number such that $\|L\| \leq \beta \leq \alpha$. Then, for every subset $M \subseteq A$ of cardinality $|M| \leq \beta$, there exists an elementary substructure $\mathcal{B} \prec \mathcal{A}$ of $\mathcal{A}$ with cardinality $|B| = \beta$ such that $M \subseteq B$.*

Proof. Without loss of generality, we may assume that $|M| = \beta$ (otherwise, we can replace M with a set M' of power β such that $M \subseteq M' \subseteq A$). We construct a sequence $B_0 \subseteq B_1 \subseteq \ldots \subseteq B_n \subseteq B_{n+1} \subseteq \ldots$ of subsets of A recursively. Initially, set $B_0 = M$. Next, for each given n, construct B_{n+1} by ensuring that for every formula $\varphi(x)$ of the language L_{B_n}, subject to

$$\mathcal{A} \vDash (\exists x)\varphi(x)$$

we choose precisely one element $b \in A$, such that $\mathcal{A} \vDash \varphi(b)$, and add it to B_n. It is easy to verify that the set

$$B = \bigcup_{n \in \mathbb{N}} B_n$$

has cardinality β and forms a substructure of $\mathcal{A}$. According to the construction, the substructure $\mathcal{B} = (B; \ldots)$ of $\mathcal{A}$ is elementary by the Tarski-Vaught criterion (Theorem 6.4.6). (Fill in the missing steps and consider where we have used the axiom of choice.)

In the case where $\|L\| \leq \beta < \alpha$, the elementary substructure $\mathcal{B}$ of the structure $\mathcal{A}$, whose existence is guaranteed by the theorem just proved, is necessarily proper. However, in the case where $\|L\| \leq \beta = \alpha$, we cannot guarantee this solely based on this theorem.

6.6.2 Löwenheim-Skolem-Tarski Upward Theorem. *Let $\mathcal{A}$ be an L-structure of infinite cardinality $|A| = \alpha$, and β be a cardinal number such that*

$$|L_A\| \ = \max(\alpha, \|L\|) \leq \beta$$

Then there exists an elementary extension $\mathcal{B} \succ \mathcal{A}$ of the structure $\mathcal{A}$ with cardinality $|B| = \beta$.

Proof. Let us form the language L_A and extend it by adding a new set of constant symbols D of cardinality $|D| = \beta$. We denote this extension of the language L (and consequently of L_A) as $L^+ = (L_A)_D$. Next, we define U as the theory in the language L^+ that requires all constants $d \in D$ to denote distinct elements. In other words, the axioms of the theory U consist precisely of all inequalities $d_1 \neq d_2$ for any two distinct constant symbols $d_1, d_2 \in D$. Since $|D| = \beta$, every model of the theory U will have a cardinality of at least β. We claim that the theory $\mathrm{Th}(\mathcal{A}_A) \cup U$ in the language L^+ has some model $\mathcal{M}$. By the compactness theorem, it suffices to verify that for any finite subtheory $U_0 \subseteq U$, the theory $\mathrm{Th}(\mathcal{A}_A) \cup U_0$ has a model. Let D_0 be the set of all constant symbols $d \in D$ appearing in the axioms of U_0. Clearly, D_0 is a finite set. We take the structure $\mathcal{A}_A$ of the language L_A and extend it to a structure $\mathcal{A}^+$ of the language L^+ by keeping the interpretations of the symbols of L_A unchanged, interpreting the constants $d \in D_0$ as distinct elements of the set A (which is always possible since A is infinite), and interpreting the remaining constants $d \in D \setminus D_0$ as arbitrary elements of A. It is obvious that $\mathcal{A}^+ \models \mathrm{Th}(\mathcal{A}_A) \cup U_0$.

Now, let $\mathcal{M}$ be a structure of the language L^+ such that $\mathcal{M} \models \mathrm{Th}(\mathcal{A}_A) \cup U$. Since $\mathcal{M} \models U$, it follows that $|M| \geq \beta$. Because $\mathcal{M} \models \mathrm{Th}(\mathcal{A}_A)$, the reduct $\mathcal{M}^- = \mathcal{M} \restriction L$ to the language L (obtained by omitting the interpretations of the constants c_a and $d \in D$) is, by Theorem 6.5.4, an elementary extension of an isomorphic copy of the structure $\mathcal{A}$. By Löwenheim-Skolem-Tarski Downward Theorem 6.6.1, $\mathcal{M}^-$ contains an elementary substructure $\mathcal{B} \prec \mathcal{M}^-$ of cardinality $|B| = \beta$ such that $A \subseteq B$. Then, we have $\mathcal{A} \prec \mathcal{M}^-$, $\mathcal{B} \prec \mathcal{M}^-$, and $A \subseteq B$, which implies $\mathcal{A} \prec \mathcal{B}$

In the case where $\alpha < \beta$, the structure $\mathcal{B}$ is again a proper elementary extension of $\mathcal{A}$. However, in the case where $\alpha = \beta$, we cannot guarantee this solely based on the theorem just proved. The question of the existence of proper elementary substructures $\mathcal{B} \prec \mathcal{A}$ or proper elementary extensions $\mathcal{B} \succ \mathcal{A}$ of a given structure $\mathcal{A}$ with the same cardinality $|B| = |A| = \alpha$ is a more subtle matter. Under certain circumstances, the existence of a proper elementary extension $\mathcal{B}$ of $\mathcal{A}$ of the same cardinality $|B| = |A|$ can be proved using the *ultrapower* construction (see Exercise 8.4.4).

7 Preservation Theorems

In this chapter, we will examine in more detail the relationship between the syntactic form of axioms of first order theories and the preservation of these theories under fundamental constructions with their models. Specifically, we will characterize theories that are preserved under transition to substructures or extensions, theories that are preserved under unions of chains, and theories preserved under homomorphic images, as theories that can be axiomatized using axioms of certain well-described syntactic form.

We begin our considerations with "an inventory" of the most important tools we will use. These include Gödel's Completeness Theorem 4.8.6, the Compactness Theorem 4.9.1 along with its Corollary 4.9.2, as well as Lemma 4.7.1 on constants. We will become familiar with additional tools in the next section.

7.1 Lemma on Mutual Compatibility and the Axiomatization Lemma

Let T and S be consistent theories in a first-order language L. We say that T and S are *compatible* if the theory $T \cup S$ is consistent. Otherwise, we say that T and S are *incompatible*.

7.1.1 Lemma on Mutual Compatibility. *Let T and S be theories in a first-order language L. Then T and S are incompatible if and only if there exist axioms $\sigma_1(x_1, \ldots, x_k), \ldots, \sigma_n(x_1, \ldots, x_k)$ of the theory S such that*

$$T \vdash (\exists\, x_1, \ldots, x_k)\big(\neg\sigma_1(x_1, \ldots, x_k) \vee \ldots \vee \neg\sigma_n(x_1, \ldots, x_k)\big)$$

Demonstration. If such formulas $\sigma_1, \ldots, \sigma_n$ exist, then

$$S \vdash (\forall\, x_1, \ldots, x_k)\big(\sigma_1(x_1, \ldots, x_k) \wedge \ldots \wedge \sigma_n(x_1, \ldots, x_k)\big)$$

as a result of which $T \cup S$ is clearly inconsistent.

Conversely, if $T \cup S$ is inconsistent, then some theory of the form $T \cup S_0$ is already inconsistent, where $S_0 = \{\sigma_1, \ldots, \sigma_n\}$ is a finite subtheory of the theory S. Let $x_1, \ldots, x_k$ be all variables that are free in some of the formulas σ_i. Then the theory

$$T \cup \big\{(\forall\, x_1, \ldots, x_k)(\sigma_1(x_1, \ldots, x_k) \wedge \ldots \wedge \sigma_n(x_1, \ldots, x_k))\big\}$$

is also inconsistent, which means that the negation of this formula is provable in T.

We denote the class of all models of the theory T in the language L as $\mathrm{Mod}(T)$. We say that a set of formulas Γ in the language L is an *axiom set of the theory* T, if

$\mathrm{Mod}(T) = \mathrm{Mod}(\Gamma)$, i.e., for every L-structure $\mathcal{A}$, it holds that $\mathcal{A} \models T$ if and only if $\mathcal{A} \models \Gamma$.

7.1.2 Axiomatization Lemma. *Let T be a consistent theory in the language L, and let Δ be a set of sentences in the language L closed under finite disjunctions. The following conditions are equivalent:*

(i) *T has an axiom set $\Gamma \subseteq \Delta$.*

(ii) *For any structures $\mathcal{A}$, $\mathcal{B}$ in the language L, the following holds:*

$$\mathcal{A} \models T \ \wedge\ \mathcal{B} \models \mathrm{Th}(\mathcal{A}) \cap \Delta \ \Rightarrow\ \mathcal{B} \models T$$

The set Δ should be viewed as a set of formulas (logically equivalent to formulas) of a certain syntactic form, e.g., universal formulas, existential formulas, positive formulas, etc.

Proof. (i) $\Rightarrow$ (ii): Assume that T has an axiom set $\Gamma \subseteq \Delta$. Further, assume that $\mathcal{A} \models T$ and $\mathcal{B} \models \mathrm{Th}(\mathcal{A}) \cap \Delta$. Then clearly $\mathcal{B} \models \Gamma$, thus $\mathcal{B} \models T$.

(ii) $\Rightarrow$ (i): Assume that (ii) holds. Denote by Γ the set of all sentences φ in the language L such that $\varphi \in \Delta$ and $T \vdash \varphi$. We show that Γ is an axiom set for T. It suffices to prove the inclusion $\mathrm{Mod}(\Gamma) \subseteq \mathrm{Mod}(T)$ (the reverse inclusion is namely obvious — why?). Let $\mathcal{B} \in \mathrm{Mod}(\Gamma)$. In order to prove that $\mathcal{B} \in \mathrm{Mod}(T)$, i.e., $\mathcal{B} \models T$, using condition (ii), we need to find a structure $\mathcal{A} \models T$ such that $\mathcal{B} \models \mathrm{Th}(\mathcal{A}) \cap \Delta$. This last condition can be equivalently rewritten as $\mathcal{A} \models \Sigma$, where

$$\Sigma = \{\neg\delta \colon \delta \in \Delta,\ \mathcal{B} \models \neg\delta\}$$

(think about why). We need to ensure that the theory $T \cup \Sigma$ has some model $\mathcal{A}$. For this, it suffices to prove that it is consistent. Otherwise, according to Lemma 7.1.1 on mutual compatibility, there would exist some formulas $\delta_1, \ldots, \delta_n \in \Delta$ such that $\neg\delta_1, \ldots, \neg\delta_n \in \Sigma$ and $T \vdash \delta_1 \vee \ldots \vee \delta_n$. Denote by δ the last formula. Clearly, $\delta \in \Delta$, hence also $\delta \in \Gamma$, and since $\mathcal{B} \models \Gamma$, it follows that $\mathcal{B} \models \delta$. However, this contradicts the fact that $\mathcal{B} \models \neg\delta_1$, …, $\mathcal{B} \models \neg\delta_n$, thus $\mathcal{B} \models \neg\delta$.

7.2 Universal Theories and Substructures

We say that a *theory T is preserved under substructures* if any substructure of a model of the theory T is also a model of T.

7.2.1 Theorem. *Let T be a consistent theory. Then the following conditions are equivalent:*

(i) *T has a set of universal axioms.*

(ii) *T is preserved under substructures.*

Proof. We focus only on the non-trivial implication (ii) $\Rightarrow$ (i). Assume that (ii) holds. Denote by Π^0_1 the set of all sentences in the language L that are equivalent to universal formulas. Clearly, Π^0_1 is closed under finite disjunctions. Using the Axiomatization Lemma 7.1.2, we will prove that T has an axiom set $\Gamma \subseteq \Pi^0_1$. Assume that $\mathcal{A} \models T$

and $\mathcal{B} \models \mathrm{Th}(\mathcal{A}) \cap \Pi_1^0$. We will show that, under these conditions, there exists an L-structure $\mathcal{C}$ such that $\mathcal{A} \equiv \mathcal{C}$ and $\mathcal{B} \subseteq \mathcal{C}$. This will complete the proof, because then $\mathcal{C} \models T$ and — since T is preserved under substructures — also $\mathcal{B} \models T$.

In other words, we need to find a structure $(\mathcal{C}, b)_{b \in B}$ in the language L_B that is a model of the theory $\mathrm{Th}(\mathcal{A}) \cup \mathrm{D}(\mathcal{B})$. It suffices to verify that this theory in the language L_B is consistent. Otherwise, for some $n \geq 1$, there would exist formulas $\theta_1(\vec{b}), \ldots, \theta_n(\vec{b}) \in \mathrm{D}(\mathcal{B})$ such that $\theta_i(x_1, \ldots, x_k)$ are atomic or negatomic formulas in the language L, $\vec{b} = (b_1, \ldots, b_k) \in B^k$, and

$$\mathrm{Th}(\mathcal{A}) \vdash \neg\theta_1(\vec{b}) \vee \ldots \vee \neg\theta_n(\vec{b})$$

However, $\mathrm{Th}(\mathcal{A})$ is a theory in the language L, in which the constants $b_1, \ldots, b_k$ do not appear. Therefore, according to Lemma 4.7.1 on constants,

$$\mathrm{Th}(\mathcal{A}) \vdash (\forall \vec{x})\big(\neg\theta_1(\vec{x}) \vee \ldots \vee \neg\theta_n(\vec{x})\big)$$

where $\vec{x} = (x_1, \ldots, x_k)$. However, the last formula is a universal sentence satisfied in $\mathcal{A}$. Thus, it is also satisfied in $\mathcal{B}$. This contradicts the fact that

$$\mathcal{B} \models (\exists \vec{x})\big(\theta_1(\vec{x}) \wedge \ldots \wedge \theta_n(\vec{x})\big)$$

Abbreviated "vector" notation such as $\vec{x} = (x_1, \ldots, x_k)$ or $\vec{b} = (b_1, \ldots, b_k)$ for ordered k-tuples of variables or elements of sets will be used further without special notice.

7.3 Existential Theories and Superstructures

We say that a *theory T is preserved under extensions* if any extension of a model of the theory T is also a model of T.

7.3.1 Theorem. *Let T be a consistent theory. Then the following conditions are equivalent:*

(i) *T has a set of existential axioms.*

(ii) *T is preserved under extensions.*

Proof. Again, we focus only on the nontrivial implication (ii) $\Rightarrow$ (i). Suppose (ii) holds. Denote by Σ_1^0 as the set of all sentences of the language L equivalent to existential formulas. Clearly, Σ_1^0 is closed under finite disjunctions. Using the Axiomatization Lemma 7.1.2, we prove that T has a set of axioms $\Gamma \subseteq \Sigma_1^0$. to this end assume that $\mathcal{A} \models T$ as well as $\mathcal{B} \models \mathrm{Th}(\mathcal{A}) \cap \Sigma_1^0$. It follows that $\mathcal{A} \models \mathrm{Th}(\mathcal{B}) \cap \Pi_1^0$ (consider why). However, this is the same situation as in the proof of the previous theorem, only with the roles of the structures $\mathcal{A}$ and $\mathcal{B}$ swapped. Therefore, there exists a structure $\mathcal{C}$ such that $\mathcal{A} \subseteq \mathcal{C}$ and $\mathcal{B} \equiv \mathcal{C}$. Then — since T is preserved under extensions — we have that $\mathcal{C} \models T$, and consequently $\mathcal{B} \models T$.

7.4 Universal-Existential Theories and Chain Unions

Recall that an *ordered set* $(I; \leq)$ is a set $I \neq \emptyset$ with a binary relation $\leq$ satisfying the following conditions:

$$(\forall x)(x \leq x)$$
$$(\forall x, y)(x \leq y \ \wedge \ y \leq x \ \Rightarrow \ x = y)$$
$$(\forall x, y, z)(x \leq y \ \wedge \ y \leq z \ \Rightarrow \ x \leq z)$$
$$(\forall x, y)(x \leq y \ \vee \ y \leq x)$$

By a *chain of structures* over an ordered set $(I; \leq)$, we mean a system of structures $(\mathcal{A}_i)_{i \in I}$ of some first-order language L such that for any elements $i \leq j$ of the set I, the structure $\mathcal{A}_i$ is a substructure of the structure $\mathcal{A}_j$, i.e., $\mathcal{A}_i \subseteq \mathcal{A}_j$. By the *union of the chain* $(\mathcal{A}_i) i \in I$, we mean the L-structure $\bar{\mathcal{A}} = \bigcup_{i \in I} \mathcal{A}_i$ with the underlying set $\bar{A} = \bigcup_{i \in I} A_i$, and with the operation and relation symbols of the language L interpreted as follows:

For any n-ary operation symbol f or relation symbol r of the language L and elements $a_1, \ldots, a_n \in \bar{A}$, we define

$$f^{\bar{\mathcal{A}}}(a_1, \ldots, a_n) = f^{\mathcal{A}_i}(a_1, \ldots, a_n)$$
$$(a_1, \ldots, a_n) \in r^{\bar{\mathcal{A}}} \Leftrightarrow (a_1, \ldots, a_n) \in r^{\mathcal{A}_i}$$

where i is any element of the set I such that $a_1, \ldots, a_n \in A_i$.

(Realize that there always exists at least one such element $i \in I$, and if $i, j \in I$ are two such elements, then the interpretations of $f^{\bar{\mathcal{A}}}$ and $r^{\bar{\mathcal{A}}}$ do not depend on which one we use in their definition.)

It is evident that if the set $(I; \leq)$ has a greatest element m (specifically, if I is finite), then $\bigcup_{i \in I} \mathcal{A}_i = \mathcal{A}_m$, so the above construction can yield something new only if $(I; \leq)$ is infinite and has no greatest element. A typical example of chains of structures is the chain $(\mathcal{A}_n)_{n \in \mathbb{N}}$ over the set $(\mathbb{N}, \leq)$ of all natural numbers with the usual ordering. Such chains can also be written in the form

$$\mathcal{A}_0 \subseteq \mathcal{A}_1 \subseteq \ \ldots \ \subseteq \mathcal{A}_n \subseteq \mathcal{A}_{n+1} \ \subseteq \ldots$$

or alternatively in the form $(\mathcal{A}_n)_{n < \omega}$.

A simple proof of the following statement is left as an exercise for the reader.

7.4.1 Proposition. *Let $(\mathcal{A}_i) i \in I$ be a chain of structures over an ordered set $(I; \leq)$. Then for every $j \in I$, we have $\mathcal{A}_j \subseteq \bigcup_{i \in I} \mathcal{A}_i$, i.e., the union of the chain of structures is an extension of each of the structures in this chain. Moreover, if a structure $\mathcal{B}$ is an extension of each of the structures in the chain $(\mathcal{A}_i) i \in I$, then $\bigcup_{i \in I} \mathcal{A}_i \subseteq \mathcal{B}$; in other words, the union of the chain $(\mathcal{A}_i)_{i \in I}$ is the smallest structure that is an extension of each of the structures $\mathcal{A}_i$.*

A chain $(\mathcal{A}_i)_{i \in I}$ of structures over an ordered set $(I; \leq)$ is called an *elementary chain* if for all pairs of elements $i \leq j$ from the set I, we have $\mathcal{A}_i \prec \mathcal{A}_j$, i.e., $\mathcal{A}_i$ is an elementary substructure of the structure $\mathcal{A}_j$.

7.4.2 Theorem. *Let $(\mathcal{A}_i)i \in I$ be an elementary chain of structures over an ordered set $(I;\leq)$. Then for every $j \in I$, we have $\mathcal{A}_j \prec \bigcup_{i\in I} \mathcal{A}_i$, i.e., the union of an elementary chain of structures is an elementary extension of each of the structures in the chain.*

Proof. Let $\bar{\mathcal{A}} = \bigcup_{i\in I} \mathcal{A}_i$. We need to show that $\mathcal{A}_j \prec \bar{\mathcal{A}}$ for each structure $\mathcal{A}_j$ in the elementary chain $(\mathcal{A}_i)_{i\in I}$. This follows from the following statement:

For all L-formulas $\varphi(x_1, \ldots, x_n)$, for each $j \in I$, and for any elements $a_1, \ldots, a_n \in A_j$,

$$\mathcal{A}_j \vDash \varphi(a_1, \ldots, a_n) \;\Leftrightarrow\; \bar{\mathcal{A}} \vDash \varphi(a_1, \ldots, a_n)$$

The proof is by induction on the complexity of the formula φ. If φ is atomic, the condition is trivially satisfied because $\mathcal{A}_j \subseteq \bar{\mathcal{A}}$. Similarly, the validity of the given condition for formulas φ and ψ obviously implies its validity for the negation $\neg\varphi$ and for the conjunction $\varphi \wedge \psi$ (details are left to the reader). So let's assume that the condition holds for the formula $\varphi(x_0, x_1, \ldots, x_n)$ and verify its validity for the formula $(\exists x_0)\varphi(x_0, x_1, \ldots, x_n)$. Let $j \in I$ and $a_1, \ldots, a_n \in A_j$. Then the following conditions are equivalent:

(1) $\bar{\mathcal{A}} \vDash (\exists x_0)\varphi(x_0, a_1, \ldots, a_n)$

(2) there exists $a_0 \in \bar{A}$ such that $\bar{\mathcal{A}} \vDash \varphi(a_0, a_1, \ldots, a_n)$

(3) there exist $k \in I$ and $a_0 \in A_k$ such that $k \geq j$ and $\bar{\mathcal{A}} \vDash \varphi(a_0, a_1, \ldots, a_n)$

(Consider why we can assume without loss of generality that $k \geq j$ in (3).) For such $k \in I$ and $a_0 \in A_k$, the condition $\bar{\mathcal{A}} \vDash \varphi(a_0, a_1, \ldots, a_n)$, by the induction hypothesis, is equivalent to the condition $\mathcal{A}_k \vDash \varphi(a_0, a_1, \ldots, a_n)$, so condition (3) is successively equivalent to the conditions:

(4) there exist $k \in I$ and $a_0 \in A_k$ such that $k \geq j$ and $\mathcal{A}_k \vDash \varphi(a_0, a_1, \ldots, a_n)$

(5) there exists $k \in I$ such that $k \geq j$ and $\mathcal{A}_k \vDash (\exists x_0)\varphi(x_0, a_1, \ldots, a_n)$

Finally, since for $j \leq k$ we have $\mathcal{A}_j \prec \mathcal{A}_k$, condition (5) is equivalent to condition

(6) $\mathcal{A}_j \vDash (\exists x_0)\varphi(x_0, a_1, \ldots, a_n)$

which is what we wanted to arrive at.

7.4.3 Proposition. *Let $(\mathcal{A}_i)_{i\in I}$ be a chain of structures over an ordered set $(I;\leq)$, and $\varphi(x_1, \ldots, x_m, y_1, \ldots, y_n)$ be an open formula. Assume that*

$$\mathcal{A}_i \vDash (\forall\, \vec{x}\,)(\exists\, \vec{y}\,)\varphi(\vec{x}, \vec{y}\,)$$

holds in each structure $\mathcal{A}_i$. Then it also holds that

$$\bigcup_{i\in I} \mathcal{A}_i \vDash (\forall\, \vec{x}\,)(\exists\, \vec{y}\,)\varphi(\vec{x}, \vec{y}\,)$$

Proof. Under the given assumption, choose arbitrary elements $a_1, \ldots, a_m \in \bigcup_{i\in I} A_i$. We need to prove the existence of some elements $b_1, \ldots, b_n \in \bigcup_{i\in I} A_i$ such that

$$\bigcup_{i\in I} \mathcal{A}_i \vDash \varphi\big(\vec{a}, \vec{b}\,\big)$$

Given the conditions, there exists some $j \in I$ such that $a_1, \ldots, a_m \in A_j$. Since

$$\mathcal{A}_j \vDash (\forall\, \vec{x}\,)(\exists\, \vec{y}\,)\varphi(\vec{x}, \vec{y}\,)$$

there exist elements $b_1, \ldots, b_n \in A_j$ such that $\mathcal{A}_j \vDash \varphi(\vec{a}, \vec{b})$. Since $\mathcal{A}_j \subseteq \bigcup_{i\in I} \mathcal{A}_i$ and φ contains no quantifiers, according to Proposition 6.1.4 (a) we have

$$\bigcup_{i\in I} \mathcal{A}_i \vDash \varphi\big(\vec{a}, \vec{b}\,\big)$$

We say that a *theory T is preserved under chain unions* if the union of an arbitrary chain of models of the theory T is also a model of T. Preservation of a theory under chain unions over the ordered set $(\mathbb{N}; \leq)$ can be defined in a similar way.

A formula of the form $(\forall x_1, \ldots, x_m)(\exists y_1, \ldots, y_n)\varphi$, where φ is an open formula, is called a *universal-existential formula*.

7.4.4 Theorem. *Let T be a consistent theory. Then the following conditions are equivalent:*

(i) *T has a set of universal-existential axioms.*

(ii) *T is preserved under chain unions.*

(iii) *T is preserved under chain unions over $(\mathbb{N}; \leq)$.*

Proof. The validity of the implication (i) $\Rightarrow$ (ii) follows from the previous theorem, and the implication (ii) $\Rightarrow$ (iii) is trivial. It is thus sufficient to focus on the implication (iii) $\Rightarrow$ (i).

Assume (iii) holds. Let Π^0_2 denote the set of all sentences in the language L equivalent to universal-existential formulas. It is clear that Π^0_2 is closed under finite disjunctions. Using the Axiomatization Lemma 7.1.2, we will prove that T has a set of axioms $\Gamma \subseteq \Pi^0_2$. Let $\mathcal{A} \vDash T$ and assume

(0) $\mathcal{B} \vDash \mathrm{Th}(\mathcal{A}) \cap \Pi^0_2$

Under these conditions, we will construct two sequences of structures $(\mathcal{A}_n)_{n<\omega}$, $(\mathcal{B}_n)_{n<\omega}$ of the language L such that $\mathcal{A}_0 = \mathcal{A}$, $\mathcal{B}_0 = \mathcal{B}$, and, for every n, $\mathcal{A}_n \equiv \mathcal{A}_{n+1}$, hence also $\mathcal{A}_n \equiv \mathcal{A}$, as well as $\mathcal{B}_n \prec \mathcal{B}_{n+1}$, where the two sequences together form an alternating interleaved chain

$$\mathcal{B}_0 \subseteq \mathcal{A}_1 \subseteq \mathcal{B}_1 \subseteq \mathcal{A}_2 \subseteq \ldots \subseteq \mathcal{B}_n \subseteq \mathcal{A}_{n+1} \subseteq \mathcal{B}_{n+1} \subseteq \ldots$$

Let $\mathcal{C}$ denote the union of these structures. Then we have both

$$\mathcal{C} = \bigcup_{1\leq n<\omega} \mathcal{A}_n = \bigcup_{n<\omega} \mathcal{B}_n$$

Since the theory T is preserved under unions of chains over $(\mathbb{N}; \leq)$, and all $\mathcal{A}_n$ are models of T, it follows that $\mathcal{C} \vDash T$. Since the chain $(\mathcal{B}_n)_{n<\omega}$ is elementary, we get $\mathcal{B} \prec \mathcal{C}$, thus ultimately $\mathcal{B} \vDash T$. Therefore, by the Axiomatization Lemma 7.1.2, we conclude that T has a set of universal-existential axioms.

Now, we show that under the condition (0), the following statement holds:

(1) There exist structures $\mathcal{A}'$, $\mathcal{B}'$ such that $\mathcal{A} \equiv \mathcal{A}'$, $\mathcal{B} \prec \mathcal{B}'$, and $\mathcal{B} \subseteq \mathcal{A}' \subseteq \mathcal{B}'$.

Since every universal formula is also a universal-existential one, from condition (0) it follows that $\mathcal{B} \vDash \mathrm{Th}(\mathcal{A}) \cap \Pi^0_1$. As we saw in the proof of Theorem 7.2.1, under this condition, the theory $\mathrm{Th}(\mathcal{A}) \cup \mathrm{D}(\mathcal{B})$ is consistent and there exists a structure $\mathcal{A}'_B = (\mathcal{A}', b)_{b\in B}$, which is a model of the theory $\mathrm{Th}(\mathcal{A}) \cup \mathrm{D}(\mathcal{B})$ in the language L_B. Therefore, $\mathcal{A}' \equiv \mathcal{A}$ and $\mathcal{B} \subseteq \mathcal{A}'$.

The structure $\mathcal{B}'$ will be obtained as a restriction of the structure $(\mathcal{B}', b, a)_{b\in B, a\in A'}$ in the language $L_{B,A'} = (L_B)_{A'}$, which we require to be a model of the theory $\mathrm{Th}(\mathcal{B}_B) \cup \mathrm{D}(\mathcal{A}')$ in this language. It is sufficient to verify that this theory is consistent. Otherwise, by Lemma 7.1.1 on mutual compatibility, there would exist formulas $\theta_1(\vec{a}), \ldots, \theta_n(\vec{a}) \in \mathrm{D}(\mathcal{A}')$, where $\theta_i(x_1, \ldots, x_k)$ are atomic or negatomic formulas of the language L and $\vec{a} = (a_1, \ldots, a_k) \in (A')^k$, such that

$$\mathrm{Th}(\mathcal{B}_B) \vdash \neg\theta_1(\vec{a}\,) \vee \ldots \vee \neg\theta_n(\vec{a}\,)$$

Since $\mathrm{Th}(\mathcal{B}_B)$ is a theory in the language L_B, in which the constants $a_1, \ldots, a_k$ do not appear, by Lemma 4.7.1 on constants, it follows that

$$\mathrm{Th}(\mathcal{B}_B) \vdash (\forall\, \vec{x}\,)(\neg\theta_1(\vec{x}\,) \vee \ldots \vee \neg\theta_n(\vec{x}\,))$$

Since this is a formula in the language L, it means that it must be satisfied in $\mathcal{B}$. However, at the same time,

$$\mathcal{A}' \vDash (\exists\, \vec{x}\,)\big(\theta_1(\vec{x}\,) \wedge \ldots \wedge \theta_n(\vec{x}\,)\big)$$

and since $\mathcal{A}' \equiv \mathcal{A}$, this last formula must also hold in $\mathcal{A}$. Since it is existential, and thus even more so a universal-existential formula, it follows from assumption (0) that it is also satisfied in $\mathcal{B}$. This contradiction proves (1).

Finally, note that for the structures obtained in this way, condition (0) still holds, i.e., $\mathcal{B}' \vDash \mathrm{Th}(\mathcal{A}') \cap \Pi^0_2$, so the entire construction can be iterated. Thus, we consecutively obtain structures $\mathcal{A}_1 = \mathcal{A}'$, $\mathcal{B}_1 = \mathcal{B}'$ from the structures $\mathcal{A}_0 = \mathcal{A}$, $\mathcal{B}_0 = \mathcal{B}$, as well as structures $\mathcal{A}_{n+1} = \mathcal{A}'_n$, $\mathcal{B}_{n+1} = \mathcal{B}'_n$ from structures $\mathcal{A}_n$, $\mathcal{B}_n$ for $n \geq 1$.

7.4.5 Exercise. Let us consider groups as structures $(G; \cdot)$ with a single binary operation. Then, group theory is defined by two axioms: the associative law

$$(\forall\, x, y, z)\big(x \cdot (y \cdot z) = (x \cdot y) \cdot z\big)$$

and the axiom expressing the existence of an identity element as well as inverse elements

$$(\exists\, u)(\forall\, x)\big(x \cdot u = x = u \cdot x \;\wedge\; (\exists\, y)(x \cdot y = u = y \cdot x)\big)$$

which can be replaced by the axiom

$$(\forall\, x)(\exists\, y)(\forall\, z)\big(z \cdot (x \cdot y) = z = (y \cdot x) \cdot z\big)$$

Neither of these two axioms is universal-existential. Therefore, it may be surprising that group theory (in this language) is preserved under chain unions. In other

words, if $\big((G_i;\cdot)\big)_{i\in I}$ is a chain of groups over some ordered set $(I;\leq)$, then its union $\Big(\bigcup_{i\in I} G_i\,;\,\cdot\Big)$ is a group, again. Prove this independently.

From Theorem 7.4.4, it follows that group theory must have a set of universal-existential axioms in the language with a single binary operation symbol $\cdot$. Find such an axiomatization of group theory.

7.5 Algebraic Closure of a Field

The axioms of field theory as structures $(F;+,\cdot,0,1)$ in a language with two binary operations $+$ and $\cdot$ and two constants 0 and 1 are certainly well known to our readers, so we will not explicitly list them here. It suffices to recall that, in addition to the universal axioms (identities) expressing the commutative and associative laws for both operations, the roles of 0 and 1 as neutral elements of these operations, and the distributive law, there are also two universal-existential axioms

$$(\forall\, x)(\exists\, y)(x+y=0) \quad \text{and} \quad (\forall\, x)(\exists\, y)(x\neq 0 \;\Rightarrow\; x\cdot y=1)$$

ensuring the existence of inverse elements for both operations. From the theorem on the preservation of theories under chain unions, it follows that field theory is also preserved under chain unions, meaning that the union of any chain of fields $\big((F_i;+,\cdot,0,1)\big)_{i\in I}$ over an ordered set $(I;\leq)$ is again a field $\Big(\bigcup_{i\in I} F_i\,;\,+,\cdot,0,1\Big)$.

Chain unions are widely used in field theory. Their most important application is the construction of the *algebraic closure of a field*, which we will now introduce.

From now on, as is customary, we will denote the field $(F;+,\cdot,0,1)$ mostly just by F, that is, the same as its underlying set. We say that a field K is an *algebraic extension* of a field F if F is a subfield of K and every element of K is algebraic over F, i.e., it is a root of some polynomial $p(x)$ with coefficients from the field F.

The reader should remember from algebra lectures several results about algebraic extensions of fields, which we briefly summarize here. If $p(x)\in F[x]$ is an irreducible polynomial of degree ≥ 2 over the field F, then the quotient ring $F[x]/(p(x))$ of the polynomial ring $F[x]$ modulo the principal ideal $(p(x))$ generated by the polynomial $p(x)$ is again a field, which is moreover an extension of the field F. In this field, the polynomial $p(x)$ has a root (namely, the residue class $x+(p(x))$), so it is no longer irreducible. Such extensions of the field F are called its *simple extensions*. Clearly, every simple extension of the field F is algebraic. It is also easy to see that for any fields F, H, and K, the following holds: if H is an algebraic extension of the field F and K is an algebraic extension of the field H, then K is an algebraic extension of the field F.

A simple proof of the following auxiliary statement is left as an exercise for the reader.

7.5.1 Lema. *Let $(F_i)_{i\in I}$ be a chain of fields over an ordered set $(I;\leq)$, each of which is an algebraic extension of the field F. Then the union of this chain $\bigcup_{i\in I} F_i$ is also an algebraic extension of the field F.*

We say that a field K is *algebraically closed* if every polynomial $p \in K[x]$ of degree ≥ 2 has at least one root in K. It is easy to see that a field K is algebraically closed if and only if the only monic irreducible polynomials of degree ≥ 1 over K are the linear polynomials of the form $x-a$, where $a \in K$. It follows further that in an algebraically closed field, every polynomial of degree $n \geq 1$ has exactly n roots, counting each root with its multiplicity. A field K is called an *algebraic closure* of the field F if K is an algebraically closed field that is an algebraic extension of the field F.

7.5.2 Theorem. *For every field, its algebraic closure exists.*

Furthermore, the algebraic closure of any field is uniquely determined up to isomorphism. However, the proof of this result is beyond the scope of our course.

Proof. Let F be an arbitrary field. First, we construct an algebraic extension $\widehat{F}$ of the field F in which every irreducible polynomial $p(x) \in F[x]$ of degree ≥ 2 has a root. In other words, there does not exist a polynomial $p(x) \in F[x]$ of degree ≥ 2 that is irreducible over the field $\widehat{F}$. Let $(p_\alpha(x))_{\alpha<\rho}$ be an arbitrary enumeration of the set of all monic irreducible polynomials of degree ≥ 2 in the ring $F[x]$ by ordinal numbers less than some ordinal number ρ. By transfinite recursion over ordinals $\alpha < \rho$, we construct a transfinite sequence $(F_\alpha)_{\alpha<\rho}$ of extensions of the field F such that

$$F_0 = F$$

$$F_{\alpha+1} = \begin{cases} F_\alpha[x]/(p_\beta(x)), & \text{where } \beta < \rho \text{ is the first-ordinal such that the polynomial} \\ & p_\beta(x) \text{ is irreducible over the field } F_\alpha \\ F_\alpha, & \text{if no such ordinal } \beta \text{ exists} \end{cases}$$

$$F_\lambda = \bigcup_{\alpha<\lambda} F_\alpha$$

for any ordinal $\alpha < \rho$, or for any limit ordinal $\lambda < \rho$. Clearly, for any $\alpha \leq \beta < \rho$, the field F_β is an extension of the field F_α, so $(F_\alpha)_{\alpha<\rho}$ is a chain of fields over the ordered set $(\{\alpha\colon \alpha < \rho\}; \leq)$.

By transfinite induction, we prove that each of the fields F_α is an algebraic extension of the field F. For $\alpha = 0$, this is obvious. Suppose that $\alpha < \rho$ is an arbitrary ordinal and assume that F_α is an algebraic extension of the field F. Then $F_{\alpha+1}$ is either a simple (hence algebraic) extension $F_\alpha[x]/(p_\beta(x))$ of the field F_α, or it is the field F_α itself. In both cases, it is an algebraic extension of the field F. Finally, let $\lambda < \rho$ be a limit ordinal. Suppose that all fields F_α, for $\alpha < \lambda$, are algebraic extensions of the field F. Then F_λ, as the union of the chain $(F_\alpha)_{\alpha<\lambda}$, is also an algebraic extension of F by Lemma 7.5.1.

Now, we set $\widehat{F} = \bigcup_{\alpha<\rho} F_\alpha$. Then the field $\widehat{F}$, as the union of the chain $(F_\alpha)_{\alpha<\rho}$ of algebraic extensions of the field F, is itself an algebraic extension of the field F. Clearly, every irreducible polynomial of degree ≥ 2 over the field F has a root in some field F_α, and thus also in the field $\widehat{F}$.

Using the described construction, we now construct another chain $\left(F^{(n)}\right)_{n\in\mathbb{N}}$ of algebraic extensions of the field F over the ordered set $\mathbb{N}$ of all natural numbers such that

$$F^{(0)} = F \qquad \text{and} \qquad F^{(n+1)} = \widehat{F^{(n)}}$$

for $n \in \mathbb{N}$. Then the union of this chain $\widetilde{F} = \bigcup_{n\in\mathbb{N}} F^{(n)}$ is again an algebraic extension of the field F. We will prove that it is also an algebraically closed field. To do this, it suffices to verify that there do not exist monic irreducible polynomials $p(x) \in \widetilde{F}[x]$ of degree ≥ 2. Suppose that

$$p(x) = x^k + a_1 x^{k-1} + \ldots + a_{k-1}x + a_k \in \widetilde{F}[x]$$

is a monic irreducible polynomial of degree $k \geq 2$. Then there exists $n \in \mathbb{N}$ such that the field $F^{(n)}$ contains all its coefficients $a_1, \ldots, a_k$, so $p(x) \in F^{(n)}[x]$ is a monic irreducible polynomial over the field $F^{(n)}$. However, then $p(x)$ has a root in the field $F^{(n+1)} = \widehat{F^{(n)}}$, and therefore also in the field $\widetilde{F}$. But this means that — contrary to our initial assumption — the polynomial $p(x)$ is not irreducible over the field $\widetilde{F}$. Thus, we have proved that the field $\widetilde{F}$ is an algebraic closure of the field F.

7.6 Positive Theories and Homomorphic Images

We say that a *theory T is preserved under homomorphic images* if each homomorphic image of any model of the theory T is also a model of T.

7.6.1 Theorem. *Let T be a consistent theory. Then the following conditions are equivalent:*

(i) *T has a set of positive axioms.*

(ii) *T is preserved under homomorphic images.*

Proof. Again, we focus only on the implication (ii) $\Rightarrow$ (i). Assume that (ii) holds. For any extension L_C of the language L with new constants, let $\mathrm{Pos}(L_C)$ and $\mathrm{Neg}(L_C)$ denote the sets of all sentences of L_C equivalent to positive and negative formulas, respectively. Clearly, $\mathrm{Pos}(L)$ is closed under finite disjunctions. Using Axiomatization Lemma 7.1.2, we prove that T has a set of axioms $\Gamma \subseteq \mathrm{Pos}(L)$. To his end, take any $\mathcal{A} \models T$ and assume that

(0) $\mathcal{B} \models \mathrm{Th}(\mathcal{A}) \cap \mathrm{Pos}(L)$.

Under this assumption, we prove the following statements:

(1) There exists an elementary extension $\mathcal{B}' \succ \mathcal{B}$ and a homomorphism $h\colon \mathcal{A} \to \mathcal{B}'$ such that $(\mathcal{B}', h(a))_{a\in A} \models \mathrm{Th}(\mathcal{A}_A) \cap \mathrm{Pos}(L_A)$.

(2) There exists an elementary extension $\mathcal{A}' \succ \mathcal{A}$ and a mapping $g\colon B \to A'$ such that $(\mathcal{B}, b)_{b\in B} \models \mathrm{Th}(\mathcal{A}', g(b))_{b\in B} \cap \mathrm{Pos}(L_B)$.

(1): Consider the theory $\mathrm{Th}(\mathcal{B}_B) \cup (\mathrm{Th}(\mathcal{A}_A) \cap \mathrm{Pos}(L_A))$ in the language $L_{A,B}$. From assumption (0), we can conclude that it is consistent, using Lemma 7.1.1 on mutual compatibility and Lemma 4.7.1 on constants in a routine way. Let $(\mathcal{B}', h(a), b)_{a\in A, b\in B}$ be its model, where $\mathcal{B}'$ is a structure of the language L. Then $\mathcal{B} \prec \mathcal{B}'$, and $(\mathcal{B}', h(a))_{a\in A} \models \mathrm{Th}(\mathcal{A}_A) \cap \mathrm{Pos}(L_A)$, which, among other things, implies that $h\colon \mathcal{A} \to \mathcal{B}$ is a homomorphism.

(2): Similarly, based on assumption (0), we can prove the consistency of the theory $\mathrm{Th}(\mathcal{A}_A) \cup (\mathrm{Th}(\mathcal{B}_B) \cap \mathrm{Neg}(L_B))$ in the language $L_{A,B}$, using the same tools as above.

Let $(\mathcal{A}', a, g(b))_{a\in A, b\in B}$ be its model, where $\mathcal{A}'$ is a structure of the language L. Then $\mathcal{A} \prec \mathcal{A}'$, and for the mapping $g\colon B \to A'$ we have $(\mathcal{A}', g(b))_{b\in B} \models \mathrm{Th}(\mathcal{B}_B) \cap \mathrm{Neg}(L_B)$, hence $(\mathcal{B}, b)_{b\in B} \models \mathrm{Th}(\mathcal{A}', g(b))_{b\in B} \cap \mathrm{Pos}(L_B)$.

By alternating iterations of constructions (1) and (2), we now construct two sequences $(\mathcal{A}_n)_{n<\omega}$, $(\mathcal{B}_n)_{n<\omega}$ of L-structures, together with mappings $g_n\colon B_n \to A_{n+1}$ and homomorphisms $h_{n+1}\colon \mathcal{A}_{n+1} \to \mathcal{B}_{n+1}$ as follows:

0° Set $\mathcal{A}_0 = \mathcal{A}$, $\mathcal{B}_0 = \mathcal{B}$.

1° Construct $\mathcal{A}_{n+1}$ and $g_n\colon B_n \to A_{n+1}$ such that

$$(\mathcal{B}_n, b)_{b\in B_n} \models \mathrm{Th}(\mathcal{A}_{n+1}, g_n(b))_{b\in B_n} \cap \mathrm{Pos}(L_{B_n})$$

according to (2).

2° Construct $(\mathcal{B}_1, b)_{b\in B_0}$ and a homomorphism $h_1\colon (\mathcal{A}_1, g_0(b))_{b\in B_0} \to (\mathcal{B}_1, b)_{b\in B_0}$ such that

$$(\mathcal{B}_1, b, h_1(a))_{b\in B_0, a\in A_1} \models \mathrm{Th}(\mathcal{A}_1, g_0(b), a)_{b\in B_0, a\in A_1} \cap \mathrm{Pos}(L_{B_0, A_1})$$

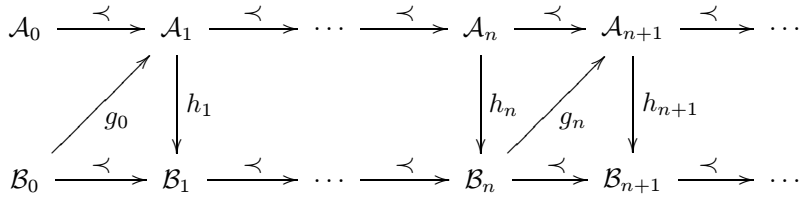
following (1).

3° Finally, for $n \geq 1$, construct $(\mathcal{B}_{n+1}, h_n(a), b)_{a\in A_n, b\in B_n}$ along with a homomorphism $h_{n+1}\colon (\mathcal{A}_{n+1}, a, g_n(b))_{a\in A_n, b\in B_n} \to (\mathcal{B}_{n+1}, h_n(a), b)_{a\in A_n, b\in B_n}$ such that

$$(\mathcal{B}_{n+1}, h_n(a), b, h_{n+1}(a'))_{a\in A_n, b\in B_n, a'\in A_{n+1}} \models \mathrm{Th}(\mathcal{A}_{n+1}, a, g_n(b), a')_{a\in A_n, b\in B_n, a'\in A_{n+1}} \cap \mathrm{Pos}(L_{A_n, B_n, A_{n+1}})$$

from the structures $(\mathcal{A}_{n+1}, a, g_n(b))_{a\in A_n, b\in B_n}$, $(\mathcal{B}_n, h_n(a), b)_{a\in A_n, b\in B_n}$ of the language L_{A_n, B_n}, following (1).

One should realize that namely meeting the stated conditions, which are assumptions of type (0), ensures that the next step can be performed anytime, so that the construction can continue indefinitely. As a result, we obtain two interlinked elementary chains:



Step 3° ensures that for $n \geq 1$ and any $a \in A_n$ we have $h_n(a) = h_{n+1}(a)$, i.e., $h_n = h_{n+1} \restriction A_n$. Similarly, steps 2° and 3° ensure that for every n and any $b \in B_n$, the following holds: $h_{n+1}(g_n(b)) = b$, therefore also $B_n \subseteq h_{n+1}(A_{n+1})$. We form the unions of elementary chains

$$\mathcal{A}_\omega = \bigcup_{n<\omega} \mathcal{A}_n \qquad \mathcal{B}_\omega = \bigcup_{n<\omega} \mathcal{B}_n$$

and set

$$h_\omega = \bigcup_{1 \le n < \omega} h_n$$

Then $\mathcal{A} \prec \mathcal{A}_\omega$, $\mathcal{B} \prec \mathcal{B}_\omega$, and $h_\omega \colon \mathcal{A}_\omega \to \mathcal{B}_\omega$ is clearly a surjective homomorphism. From the condition $\mathcal{A} \models T$, it follows that $\mathcal{A}_\omega \models T$, and since T is preserved under homomorphic images, also $\mathcal{B}_\omega \models T$, thus finally $\mathcal{B} \models T$. According to Axiomatization Lemma 7.1.2, T has a set of positive axioms.

7.6.2 Exercise. Under the assumption (0) from the proof of Theorem 7.6.1, prove that both the theories

$$\mathrm{Th}(\mathcal{B}_B) \cup (\mathrm{Th}(\mathcal{A}_A) \cap \mathrm{Pos}(L_A)) \quad \text{and} \quad \mathrm{Th}(\mathcal{A}_A) \cup (\mathrm{Th}(\mathcal{B}_B) \cap \mathrm{Neg}(L_B))$$

in the language $L_{A,B}$, appearing in (1) and (2), respectively, are consistent.

8 Ultraproducts and Axiomatic Classes

Our final chapter is devoted to model constructions based on the construction of the direct product. We will discuss the so-called *filtered products* and *filtered powers*. The most important case among them represent the so-called *ultraproducts*, which preserve all first-order properties, and *ultrapowers*, which allow us to construct elementary extensions of structures in a uniform and elegant way. We will also demonstrate some applications of these constructions, but the reader should be aware that this is only a small sample, and the possibilities of applications of ultraproducts are far from being exhausted by them.

8.1 Direct Product of Structures

The *direct* or *Cartesian product* of a system of sets $(A_i)_{i\in I}$ is defined as the set $\prod_{i\in I} A_i$ of all functions $\alpha\colon I \to \bigcup_{i\in I} A_i$ such that $\alpha(i) \in A_i$ for each $i \in I$. The statement that for any system $(A_i)_{i\in I}$ of *non-empty* sets A_i, their Cartesian product is also non-empty is equivalent to the axiom of choice, whose validity we will assume automatically.

8.1.1 Example. Consider that for a *finite* system $(A_i)_{i=1}^n$, there is a natural bijection between "two versions" of the Cartesian product $A_1 \times \ldots \times A_n$ and $\prod_{i=1}^n A_i$, and explicitly describe it.

The *direct* or *Cartesian power* of a set A with an index set I is defined as the direct product

$$A^I = \prod_{i\in I} A_i$$

where $A_i = A$ for each $i \in I$. If $I \neq \emptyset$, we have at our disposal the injective *diagonal mapping* $A \to A^I$, assigning to each element $a \in A$ the constant function $\overline{a} \in A^I$, i.e., $\overline{a}(i) = a$ for all $i \in I$.

Let $(\mathcal{A}_i)_{i\in I}$ be a system of first-order structures over the set I. The *direct product* of these structures is defined as the structure

$$\widetilde{\mathcal{A}} = \prod_{i\in I} \mathcal{A}_i$$

with the base set $\widetilde{A} = \prod_{i\in I} A_i$, where the operational and relational symbols are interpreted component-wise, i.e., for any n-ary functional symbol f, or n-ary relational symbol r of the language L and all $\alpha_1, \ldots, \alpha_n \in \widetilde{A}$, we have

$$f^{\widetilde{\mathcal{A}}}(\alpha_1, \ldots, \alpha_n)(i) = f^{\mathcal{A}_i}(\alpha_1(i), \ldots, \alpha_n(i)) \qquad \text{for each } i \in I$$

$$(\alpha_1, \ldots, \alpha_n) \in r^{\widetilde{\mathcal{A}}} \;\Leftrightarrow\; (\alpha_1(i), \ldots, \alpha_n(i)) \in r^{\mathcal{A}_i} \qquad \text{for each } i \in I$$

The above definitions naturally extend to arbitrary terms and atomic formulas. Verify for yourself that for any term $t(x_1,\dots,x_n)$ or atomic formula $\varphi(x_1,\dots,x_n)$ and elements $\alpha_1,\dots,\alpha_n \in \widetilde{A}$, the following holds:

$$t^{\widetilde{\mathcal{A}}}(\alpha_1,\dots,\alpha_n)(i) = t^{\mathcal{A}_i}(\alpha_1(i),\dots,\alpha_n(i)) \qquad \text{for each } i \in I$$
$$\widetilde{A} \vDash \varphi(\alpha_1,\dots,\alpha_n) \;\Leftrightarrow\; \mathcal{A}_i \vDash \varphi(\alpha_1(i),\dots,\alpha_n(i)) \qquad \text{for each } i \in I$$

For completeness, we also define the direct product of an empty system of structures (corresponding to the case $I = \emptyset$) as a single-element structure with operational symbols interpreted in the only possible way and with full relations.

If the index set I is clear from the context, we write simply $\prod A_i$ or $\prod \mathcal{A}_i$, respectively.

The *direct* or *Cartesian power* of a structure $\mathcal{A}$ with an index set I is defined as the direct product

$$\mathcal{A}^I = \prod_{i\in I} \mathcal{A}_i$$

where $\mathcal{A}_i = \mathcal{A}$ for each $i \in I$. Clearly, if $I \neq \emptyset$, then the diagonal mapping $A \to A^I$ is an embedding of the structure $\mathcal{A}$ into its direct power $\mathcal{A}^I$, i.e., $\mathcal{A} \hookrightarrow \mathcal{A}^I$.

For any formula $\varphi(x_1,\dots,x_n)$ of the language L, we define its *boolean truth value* on elements $\alpha_1,\dots,\alpha_n \in \prod A_i$ as the set

$$[\varphi(\alpha_1,\dots,\alpha_n)] = \big\{ i \in I \colon \mathcal{A}_i \vDash \varphi(\alpha_1(i),\dots,\alpha_n(i)) \big\}$$

Clearly, $[\varphi(\alpha_1,\dots,\alpha_n)] \subseteq I$, so the boolean truth value $[\varphi(\alpha_1,\dots,\alpha_n)]$ is an element of the Boolean algebra $\mathcal{P}(I)$ of all subsets of I.

Note that in the direct power, the boolean truth value on constant functions takes only two values.

8.1.2 Proposition. *Let I be a non-empty set, $\mathcal{A}$ be a structure, and $\varphi(x_1,\dots,x_n)$ be a formula. Then for all $a_1,\dots,a_n \in A$, we have*

$$[\varphi(\overline{a}_1,\dots,\overline{a}_n)] = \begin{cases} I & \text{if } \mathcal{A} \vDash \varphi(a_1,\dots,a_n) \\ \emptyset & \text{if } \mathcal{A} \nvDash \varphi(a_1,\dots,a_n) \end{cases}$$

The boolean truth value of formulas preserves the natural relationship between logical connectives and boolean operations in the Boolean algebra $\mathcal{P}(I)$. We leave the simple proof of the following proposition as an exercise for the reader.

8.1.3 Proposition. *Let $(\mathcal{A}_i)_{i\in I}$ be a system of first-order structures and $\varphi(x_1,\dots,x_n)$, $\psi(x_1,\dots,x_n)$ be formulas. Then for any $\alpha_1,\dots,\alpha_n \in \prod A_i$, we have*

$$[\neg\varphi(\alpha_1,\dots,\alpha_n)] = I \smallsetminus [\varphi(\alpha_1,\dots,\alpha_n)]$$
$$[(\varphi \wedge \psi)(\alpha_1,\dots,\alpha_n)] = [\varphi(\alpha_1,\dots,\alpha_n)] \cap [\psi(\alpha_1,\dots,\alpha_n)]$$
$$[(\varphi \vee \psi)(\alpha_1,\dots,\alpha_n)] = [\varphi(\alpha_1,\dots,\alpha_n)] \cup [\psi(\alpha_1,\dots,\alpha_n)]$$

For the boolean truth value of existentially quantified formulas, the so-called *maximality principle* holds.

8.1.4 Proposition. *Let $(\mathcal{A}_i)_{i\in I}$ be a system of structures and $\varphi(x, x_1, \dots, x_n)$ be a formula. Then for any $\alpha_1, \dots, \alpha_n \in \prod A_i$, there exists $\alpha_0 \in \prod A_i$ such that for each $\alpha \in \prod A_i$, we have*

$$[\varphi(\alpha, \alpha_1, \dots, \alpha_n)] \subseteq [\varphi(\alpha_0, \alpha_1, \dots, \alpha_n)] = [(\exists\, x)\varphi(x, \alpha_1, \dots, \alpha_n)]$$

Proof. Let us denote

$$J = [(\exists\, x)\varphi(x, \alpha_1, \dots, \alpha_n)] = \big\{i \in I\colon (\exists\, a_i \in A_i)\big(\mathcal{A}_i \vDash \varphi(a_i, \alpha_1(i), \dots, \alpha_n(i))\big)\big\}$$

Using the axiom of choice, we define a function $\alpha_0 \in \prod A_i$ as follows:

$$\alpha_0(i) = \begin{cases} a_i & \text{where } a_i \in A_i \text{ satisfies } \varphi(a_i, \alpha_1(i), \dots, \alpha_n(i)), \text{ if } i \in J \\ a_i & \text{where } a_i \text{ is any element of the set } A_i, \text{ if } i \notin J \end{cases}$$

Verification that the function α_0 satisfies all required conditions is left to the reader.

The following two propositions show that the boolean truth value satisfies conditions analogous to the axioms of equality. Using these axioms, they can indeed be easily verified — try it on your own.

8.1.5 Proposition. *Let $(\mathcal{A}_i)_{i\in I}$ be a system of structures. Then for any $\alpha, \beta, \gamma \in \prod A_i$, we have*

$$[\alpha = \alpha] = I$$
$$[\alpha = \beta] = [\beta = \alpha]$$
$$[\alpha = \beta] \cap [\beta = \gamma] \subseteq [\alpha = \gamma]$$

8.1.6 Proposition. *Let $(\mathcal{A}_i)_{i\in I}$ be a system of structures, $t(x_1, \dots, x_n)$ be a term, and $\varphi(x_1, \dots, x_n)$ be a formula of the language L. Then for any $\alpha_1, \beta_1, \dots, \alpha_n, \beta_n \in \prod A_i$, we have*

$$[\alpha_1 = \beta_1] \cap \dots \cap [\alpha_n = \beta_n] \subseteq [t(\alpha_1, \dots, \alpha_n) = t(\beta_1, \dots, \beta_n)]$$
$$[\alpha_1 = \beta_1] \cap \dots \cap [\alpha_n = \beta_n] \cap [\varphi(\alpha_1, \dots, \alpha_n)] \subseteq [\varphi(\beta_1, \dots, \beta_n)]$$

Note that for an atomic formula $\varphi(x_1, \dots, x_n)$, we have

$$\prod_{i\in I} \mathcal{A}_i \vDash \varphi(\alpha_1, \dots, \alpha_n) \;\Leftrightarrow\; [\varphi(\alpha_1, \dots, \alpha_n)] = I$$

which means that for it to hold, it must "hold everywhere". On the other hand, for its negation $\neg\varphi(x_1, \dots, x_n)$, we have

$$\prod_{i\in I} \mathcal{A}_i \vDash \neg\varphi(\alpha_1, \dots, \alpha_n) \;\Leftrightarrow\; [\neg\varphi(\alpha_1, \dots, \alpha_n)] \neq \emptyset$$

so for it to hold, it suffices that it holds in at least one of the structures $\mathcal{A}_i$. In other words, our definition of the structure $\prod \mathcal{A}_i$ has a consequence that, for the satisfaction of formulas of different syntactic form, different criteria for their boolean truth value hold. In the next three sections, we will attempt to navigate this situation further.

8.2 Filters and Ultrafilters

A *filter* on the set I is any nonempty set $\mathcal{D} \subseteq \mathcal{P}(I)$ such that, for any $X, Y \subseteq I$, we have

(a) if $X \in \mathcal{D}$ and $X \subseteq Y$, then $Y \in \mathcal{D}$;

(b) if $X, Y \in \mathcal{D}$, then $X \cap Y \in \mathcal{D}$.

Instead of the term *filter*, the term *dual ideal* is sometimes used. Conditions (a) and (b) imply that for any $X, Y \subseteq I$, we have $X \cap Y \in \mathcal{D}$ if and only if both $X \in \mathcal{D}$ and $Y \in \mathcal{D}$.

Clearly, every filter contains the entire set I, and the single-element set $\mathcal{D} = \{I\}$ is the smallest of all filters on I — we call it the *trivial filter*. The system $\mathcal{P}(I)$ of all subsets of I also forms a filter on I — we call it the *improper filter*. All other filters are called *proper filters*. Clearly, a filter $\mathcal{D}$ is proper if and only if $\emptyset \notin \mathcal{D}$. Some authors understand the term *filter* to mean only a proper filter. More generally, for any set $J \subseteq I$,

$$\mathcal{P}_J(I) = \{X \subseteq I : J \subseteq X\}$$

is a filter on I; such filters are called *principal filters*. Clearly, on a finite set I, all filters are principal. However, on an infinite set, there also exist *non-principal filters*, i.e., those that do not have a smallest element. An example is the so-called *Fréchet filter*

$$\mathcal{F}(I) = \{X \subseteq I : \text{set } I \setminus X \text{ is finite}\}$$

Clearly, for a finite set I, the Fréchet filter $\mathcal{F}(I)$ coincides with the improper filter $\mathcal{P}(I)$. However, if I is infinite, then $\mathcal{F}(I)$ is a nontrivial, proper, and non-principal filter on I.

Additional examples of filters are provided by topological and measurable spaces. In a topological space $(X, \mathcal{T})$, for each point $x \in X$, the set

$$\mathcal{V}(x) = \{V \subseteq X : (\exists U \in \mathcal{T})(x \in U \subseteq V)\}$$

forms a filter on X, called the *neighborhood filter* of the point x.

Similarly, if $\mathbf{X} = (X, \mathcal{B}, \mu)$ is a measurable space with a complete measure μ (i.e., for any $A, B \subseteq X$, from $A \subseteq B \in \mathcal{B}$ and $\mu(B) = 0$, it follows that $A \in \mathcal{B}$ and $\mu(A) = 0$), then the set

$$\mathcal{Z}(\mathbf{X}) = \{A \subseteq X : \mu(X \setminus A) = 0\}$$

forms a filter on X.

Intuitively, we view a (proper) filter $\mathcal{D}$ on a set I as a system of its subsets that are, "large" in some sense, and thus the corresponding filter "captures" them. There can be many criteria of "largeness" — each (proper) filter represents one of the possibilities. However, for such a criterion to be "reasonable", it must certainly satisfy condition (a); condition (b) expresses an idealizing requirement that the complement of a "large" set must be "small", and the union of two "small" sets should be a "small" set, again.

A filter $\mathcal{D}$ on a set I is called an *ultrafilter* if $\mathcal{D}$ is a proper filter and for any set $X \subseteq I$, either $X \in \mathcal{D}$ or $I \setminus X \in \mathcal{D}$ (for a proper filter, only one of these two possibilities can occur). Clearly, for any element $i \in I$, the principal filter

$$\mathcal{P}_{\{i\}}(I) = \{X \subseteq I : i \in X\}$$

is an ultrafilter on I. However, we are primarily interested in non-principal ultrafilters — such ultrafilters can exist only on infinite sets. Clearly, they must be ultrafilters that extend the Fréchet filter $\mathcal{F}(I)$.

Think independently about the fact that a filter $\mathcal{D}$ on a set I is an ultrafilter if and only if it is a maximal proper filter on I. From the Axiom of Choice, more precisely from its equivalent formulation in the form of Zorn's Lemma or Hausdorff's maximality principle, the following result follows. A system of subsets $\mathcal{S} \subseteq \mathcal{P}(I)$ of the set I is called *centered* if for any finite number of sets $X_1, \ldots, X_n \in \mathcal{S}$, we have $X_1 \cap \ldots \cap X_n \neq \emptyset$.

8.2.1 Proposition. *Let $\mathcal{S}$ be any centered system of subsets of the set I. Then there exists an ultrafilter $\mathcal{E}$ on I such that $\mathcal{S} \subseteq \mathcal{E}$. In particular, for any proper filter $\mathcal{D}$ on the set I, there exists an ultrafilter $\mathcal{E}$ on I such that $\mathcal{D} \subseteq \mathcal{E}$.*

8.3 Filtered Products

The basic idea behind of the filtered product construction consists in a certain relaxation of the requirements for the equality of functions in the direct product and of their belonging to relations corresponding to relational symbols of the given language. Functions from the direct product that are equal on a "large" subset of the index set are identified, and to include an n-tuple of functions in the corresponding relation, it suffices that the individual components of the elements of this n-tuple belong to the respective relations in the individual factors for a "large" set of indices. A more precise description of the entire construction follows.

Let $(A_i)_{i\in I}$ be a system of sets and let $\mathcal{D}$ be a filter on the set I. For any functions $\alpha, \beta \in \prod A_i$, we define

$$\alpha \equiv_{\mathcal{D}} \beta \;\Leftrightarrow\; [\alpha = \beta] \in \mathcal{D}$$

From the properties of the filter and boolean truth values, it immediately follows that $\equiv_{\mathcal{D}}$ is an equivalence relation on the set $\prod_{i\in I} A_i$. The quotient set $\prod_{i\in I} A_i / \equiv_{\mathcal{D}}$ will be denoted as

$$\prod_{i\in I} A_i \Big/ \mathcal{D} \quad \text{or only} \quad \prod A_i \Big/ \mathcal{D}$$

and called the *filtered* or *reduced product* of the system of sets $(A_i)_{i\in I}$ with respect to the filter $\mathcal{D}$. The equivalence class of a function $\alpha \in \prod A_i$ is denoted $\alpha^{\mathcal{D}}$. Clearly, the assignement $\alpha \mapsto \alpha^{\mathcal{D}}$ defines a canonical surjective mapping $\prod_{i\in I} A_i \to \prod_{i\in I} A_i/\mathcal{D}$ from the direct product $\prod A_i$ onto the filtered product $\prod A_i/\mathcal{D}$.

If $(\mathcal{A}_i)_{i\in I}$ is a system of first-order structures, then the *filtered* or *reduced product* of this system is defined as the structure

$$\widetilde{\mathcal{A}}/\mathcal{D} = \prod_{i\in I} \mathcal{A}_i \Big/ \mathcal{D}$$

with the base set $\prod_{i\in I} A_i/\mathcal{D}$, where the specific symbols of the respective language L are interpreted as follows:

$$f^{\widetilde{\mathcal{A}}/\mathcal{D}}\big(\alpha_1^{\mathcal{D}},\dots,\alpha_n^{\mathcal{D}}\big) = f^{\widetilde{\mathcal{A}}}(\alpha_1,\dots,\alpha_n)^{\mathcal{D}}$$
$$\big(\alpha_1^{\mathcal{D}},\dots,\alpha_n^{\mathcal{D}}\big) \in r^{\widetilde{\mathcal{A}}/\mathcal{D}} \;\Leftrightarrow\; [r(\alpha_1,\dots,\alpha_n)] \in \mathcal{D}$$

for any n-ary functional symbol f or n-ary relational symbol r and all $\alpha_1,\dots,\alpha_n \in \prod A_i$.

The properties of filters and boolean truth values ensure that these definitions are correct, i.e., they do not depend on the individual representatives α_k of the equivalence classes $\alpha_k^{\mathcal{D}}$ for $k = 1,\dots,n$. (Consider this independently.) The canonical mapping $\prod A_i \to \prod A_i/\mathcal{D}$ given by the assignment $\alpha \mapsto \alpha^{\mathcal{D}}$ is then a homomorphism from the direct product $\prod \mathcal{A}_i$ onto the reduced product $\prod \mathcal{A}_i/\mathcal{D}$.

If $J \subseteq I$ and $\mathcal{D} = \mathcal{P}_J(I)$ is a principal filter, then the assignment $\alpha \mapsto \alpha \restriction J$, which maps each function $\alpha \in \prod_{i\in I} A_i$ to its restriction $\alpha \restriction J \in \prod_{i\in J} A_i$, has the property that

$$\alpha \restriction J = \beta \restriction J \;\Leftrightarrow\; J \subseteq [\alpha = \beta] \;\Leftrightarrow\; \alpha \equiv_{\mathcal{D}} \beta$$

for any $\alpha, \beta \in \prod A_i$. Thus, the assignment $\alpha^{\mathcal{D}} \mapsto \alpha \restriction J$ defines a bijective mapping $\prod_{i\in I} A_i/\mathcal{D} \to \prod_{i\in J} A_i$. The reader should independently verify that this is even an isomorphism of the filtered product $\prod_{i\in I} \mathcal{A}_i\Big/\mathcal{D}$ onto the direct product $\prod_{i\in J} \mathcal{A}_i$. In some special cases, we obtain

$$\prod_{i\in I} \mathcal{A}_i\Big/\mathcal{D} \cong \prod_{i\in I} \mathcal{A}_i \quad \text{for the trivial filter } \mathcal{D} = \{I\}$$
$$\prod_{i\in I} \mathcal{A}_i\Big/\mathcal{D} \cong \mathcal{A}_j \quad \text{for the principal ultrafilter } \mathcal{D} = \{X \subseteq I\colon j \in X\}$$
$$\prod_{i\in I} \mathcal{A}_i\Big/\mathcal{D} \cong \prod_{i\in \emptyset} \mathcal{A}_i \quad \text{for the improper filter } \mathcal{D} = \mathcal{P}(I)$$

In any case, it is clear that reduced products can bring something new compared to direct products only in the case of non-principal filters.

A *filtered* or *reduced power* of a structure $\mathcal{A}$ with an index set I under the filter $\mathcal{D}$ on I is defined as the filtered product

$$\mathcal{A}^I/\mathcal{D} = \prod_{i\in I} \mathcal{A}_i\Big/\mathcal{D}$$

where $\mathcal{A}_i = \mathcal{A}$ for each $i \in I$. The assignment $a \mapsto \overline{a}^{\mathcal{D}}$, which maps each element $a \in A$ to the equivalence class $\overline{a}^{\mathcal{D}} \in A^I/\mathcal{D}$ of the constant function $\overline{a} \in A^I$, is called the *diagonal mapping* of the set A into its filtered power $A^I/\mathcal{D}$. From the properties of filters and boolean truth values, the following observation immediately follows. The proof is left to the reader.

8.3.1 Proposition. *Let $\mathcal{A}$ be a structure and $\mathcal{D}$ be a proper filter on a nonempty set I. Then the diagonal mapping $A \to A^I/\mathcal{D}$ is an embedding of the structure $\mathcal{A}$ into its filtered power $A^I/\mathcal{D}$, i.e., $\mathcal{A} \hookrightarrow \mathcal{A}^I/\mathcal{D}$.*

8.3.2 Example. (a) The filtered power $\mathbb{R}^{\mathbb{N}}/\mathcal{F}$ with respect to the Fréchet filter $\mathcal{F} = \mathcal{F}(\mathbb{N})$ consists of equivalence classes $\alpha^{\mathcal{F}}$ of all real sequences $\alpha\colon \mathbb{N} \to \mathbb{R}$, where sequences α, β represent the same equivalence class if and only if they are equal "almost everywhere", i.e., everywhere except for elements $n \in J$ of some finite set $J \subseteq \mathbb{N}$.

(b) Let $\mathbf{X} = (X, \mathcal{B}, \mu)$ be a measurable space with a complete measure μ, and let $\mathcal{Z} = \mathcal{Z}(\mathbf{X})$ be the filter on X consisting of all sets $A \subseteq X$ whose complement has measure 0. Then, the filtered power $\mathbb{R}^X/\mathcal{Z}$ consists of equivalence classes $\alpha^{\mathcal{Z}}$ of all real functions $\alpha : X \to \mathbb{R}$, where functions α, β represent the same equivalence class if and only if they are equal "almost everywhere", i.e., everywhere except for elements $x \in B$ of some set $B \in \mathcal{B}$ of measure 0.

8.3.3 Exercise. (Frayne-Morel-Scott)

(a) Let $K \neq \emptyset$, $(I_k)_{k\in K}$ be a system of nonempty pairwise disjoint sets, and $I = \bigcup_{k\in K} I_k$. Further, let $(A_i)_{i\in I}$ be a system of nonempty sets. The mapping h from the direct product $\prod_{i\in I} A_i$ to the iterated direct product $\prod_{k\in K}\Big(\prod_{i\in I_k} A_i\Big)$ is defined by

$$h(\alpha)(k) = \alpha \restriction I_k$$

for $\alpha \in \prod_{i\in I} A_i$, $k \in K$. Prove that h is a bijection. Assuming that $(\mathcal{A}_i)_{i\in I}$ is a system of structures of a first-order language L, prove that

$$h\colon \prod_{i\in I} \mathcal{A}_i \longrightarrow \prod_{k\in K}\Big(\prod_{i\in I_k} \mathcal{A}_i\Big)$$

is an isomorphism of L-structures.

(b) Let $\mathcal{E}$ be a filter on the set K, and for each $k \in K$, let $\mathcal{D}_k$ be a filter on the set I_k. Define

$$\mathcal{D} = \big\{X \subseteq I\colon \{k \in K\colon X \cap I_k \in \mathcal{D}_k\} \in \mathcal{E}\big\}$$

Prove that $\mathcal{D}$ is a filter on the set I.

(c) Prove that $\mathcal{D}$ is an ultrafilter if and only if $\mathcal{E}$ is an ultrafilter and

$$\{k \in K\colon \mathcal{D}_k \text{ je ultrafilter}\} \in \mathcal{E}$$

(d) Using the mapping h from (a), construct a natural bijection between the filtered product $\prod_{i\in I} A_i/\mathcal{D}$ and the iterated filtered product

$$\prod_{k\in K}\Big(\prod_{i\in I_k} A_i\Big/\mathcal{D}_k\Big)\Big/\mathcal{E}$$

(e) Prove that the bijection described in (d) is an isomorphism of L-structures

$$\prod_{i\in I} \mathcal{A}_i\Big/\mathcal{D} \longrightarrow \prod_{k\in K}\Big(\prod_{i\in I_k} \mathcal{A}_i\Big/\mathcal{D}_k\Big)\Big/\mathcal{E}$$

8.4 Ultraproducts

The *ultraproduct* of a system $(\mathcal{A}_i)_{i\in I}$ of first-order structures is the reduced product

$$\prod_{i\in I}\mathcal{A}_i\Big/\mathcal{D}$$

of this system with respect to some ultrafilter $\mathcal{D}$ on the set I. A reduced power $\mathcal{A}^I/\mathcal{D}$ of a structure $\mathcal{A}$ with respect to an ultrafilter $\mathcal{D}$ is called an *ultrapower* of $\mathcal{A}$.

Since for the principal ultrafilter $\mathcal{D} = \mathcal{P}_{\{j\}}(I)$, we have $\prod \mathcal{A}_i/\mathcal{D} \cong \mathcal{A}_j$, we will mainly focus on ultraproducts with respect to non-principal ultrafilters. Such ultrafilters can only exist on infinite sets. Clearly, an ultrafilter $\mathcal{D}$ on an infinite set I is nonprincipal if and only if it extends the Fréchet filter, i.e., $\mathcal{F}(I) \subseteq \mathcal{D}$.

The satisfaction of formulas in ultraproducts can be fully characterized in terms of their boolean truth values.

8.4.1 Łoś' Theorem. *Let $(\mathcal{A}_i)_{i\in I}$ be a system of structures, and let $\mathcal{D}$ be an ultrafilter on the set I. Then, for any formula $\varphi(x_1,\dots,x_n)$ and elements $\alpha_1,\dots,\alpha_n \in \prod A_i$, we have*

$$\prod_{i\in I}\mathcal{A}_i\Big/\mathcal{D} \vDash \varphi\big(\alpha_1^{\mathcal{D}},\dots,\alpha_n^{\mathcal{D}}\big) \;\Leftrightarrow\; [\varphi(\alpha_1,\dots,\alpha_n)] \in \mathcal{D}$$

In other words, an n-tuple of elements $\alpha_1^{\mathcal{D}},\dots,\alpha_n^{\mathcal{D}} \in \prod A_i/\mathcal{D}$ satisfies the formula φ in the ultraproduct $\prod_{i\in I}\mathcal{A}_i/\mathcal{D}$ if and only if the boolean truth value $[\varphi(\alpha_1,\dots,\alpha_n)]$ is "large" in the sense of the ultrafilter $\mathcal{D}$, meaning that the n-tuple of functions $\alpha_1,\dots,\alpha_n \in \prod A_i$ satisfies the formula φ componentwise for a "large" subset of indices $i \in I$.

Proof. The proof can be carried out simply by induction on the complexity of the formula φ. If φ is atomic, then the validity of the stated equivalence follows from the definition of operations and relations in filtered products. Assume, therefore, that the given equivalence holds for the formulas $\varphi(x_1,\dots,x_n)$, $\psi(x_1,\dots,x_n)$, and for any elements $\alpha_1,\dots,\alpha_n \in \prod A_i$. Using this assumption, we verify its validity for the formulas $\neg\varphi$ and $\varphi\wedge\psi$.

Under these assumptions, for any $\alpha_1,\dots,\alpha_n \in \prod A_i$, the following conditions are equivalent:

$$\prod\mathcal{A}_i\Big/\mathcal{D} \vDash \neg\varphi\big(\alpha_1^{\mathcal{D}},\dots,\alpha_n^{\mathcal{D}}\big)$$
$$\prod\mathcal{A}_i\Big/\mathcal{D} \nvDash \varphi\big(\alpha_1^{\mathcal{D}},\dots,\alpha_n^{\mathcal{D}}\big)$$
$$[\varphi(\alpha_1,\dots,\alpha_n)] \notin \mathcal{D}$$
$$[\neg\varphi(\alpha_1,\dots,\alpha_n)] = I \smallsetminus [\varphi(\alpha_1,\dots,\alpha_n)] \in \mathcal{D}$$

Here, the equivalence of the conditions in the second and third lines is guaranteed by the induction hypothesis, and the equivalence of the conditions in the third and fourth lines follows from the properties of boolean truth values and ultrafilters.

Similarly, for any $\alpha_1, \dots, \alpha_n \in \prod A_i$, the following conditions are equivalent:

$$\prod \mathcal{A}_i \Big/ \mathcal{D} \vDash (\varphi \wedge \psi)\big(\alpha_1^{\mathcal{D}}, \dots, \alpha_n^{\mathcal{D}}\big)$$
$$\prod \mathcal{A}_i \Big/ \mathcal{D} \vDash \varphi\big(\alpha_1^{\mathcal{D}}, \dots, \alpha_n^{\mathcal{D}}\big) \;\wedge\; \prod \mathcal{A}_i \Big/ \mathcal{D} \vDash \psi\big(\alpha_1^{\mathcal{D}}, \dots, \alpha_n^{\mathcal{D}}\big)$$
$$[\varphi(\alpha_1, \dots, \alpha_n)] \in \mathcal{D} \;\wedge\; [\psi(\alpha_1, \dots, \alpha_n)] \in \mathcal{D}$$
$$[(\varphi \wedge \psi)(\alpha_1, \dots, \alpha_n)] = [\varphi(\alpha_1, \dots, \alpha_n)] \cap [\psi(\alpha_1, \dots, \alpha_n)] \in \mathcal{D}$$

Here, the equivalence of the conditions in the second and third lines is guaranteed by the induction hypothesis, and the equivalence of the conditions in the third and fourth lines follows from the properties of boolean truth values and arbitrary filters.

Further, assume that the given equivalence holds for the formula $\varphi(x, x_1, \dots, x_n)$ and any elements $\alpha, \alpha_1, \dots, \alpha_n \in \prod A_i$. Using this assumption, we verify its validity for the formula $(\exists\, x)\varphi(x, x_1, \dots, x_n)$. Under these assumptions, for any $\alpha_1, \dots, \alpha_n \in \prod A_i$, the following conditions are equivalent:

$$\prod \mathcal{A}_i \Big/ \mathcal{D} \vDash (\exists\, x)\varphi\big(x, \alpha_1^{\mathcal{D}}, \dots, \alpha_n^{\mathcal{D}}\big)$$
$$\Big(\exists\, \alpha \in \prod A_i\Big)\Big(\prod \mathcal{A}_i \Big/ \mathcal{D} \vDash \varphi\big(\alpha^{\mathcal{D}}, \alpha_1^{\mathcal{D}}, \dots, \alpha_n^{\mathcal{D}}\big)\Big)$$
$$\Big(\exists\, \alpha \in \prod A_i\Big)\big([\varphi(\alpha, \alpha_1, \dots, \alpha_n)] \in \mathcal{D}\big)$$
$$[(\exists\, x)\varphi(x, \alpha_1, \dots, \alpha_n)] \in \mathcal{D}$$

Here, the equivalence of the conditions in the second and third lines is guaranteed by the induction hypothesis, and the equivalence of the conditions in the third and fourth lines follows from the maximum principle for boolean truth values and the properties of arbitrary filters.

Note that the inductive steps for conjunction and existential quantification would work for any filter $\mathcal{D}$; the condition that $\mathcal{D}$ is an ultrafilter was needed only in the inductive step for negation.

An immediate consequence of Łoś's theorem is the fact that the ultrapower of models of any first-order theory is itself a model of that theory.

8.4.2 Corollary. *Let T be a first-order theory, $\mathcal{D}$ be an ultrafilter on a set I, and $(\mathcal{A}_i)_{i\in I}$ be a system of structures such that $\mathcal{A}_i \vDash T$ for each $i \in I$. Then*

$$\prod_{i\in I} \mathcal{A}_i \Big/ \mathcal{D} \vDash T$$

8.4.3 Corollary. *Let $\mathcal{A}$ be a first-order structure and $\mathcal{D}$ be an ultrafilter on a set I. Then the diagonal mapping $a \mapsto \overline{a}^{\mathcal{D}}$ is an elementary embedding $\mathcal{A} \xrightarrow{\prec} \mathcal{A}^I/\mathcal{D}$ of the structure $\mathcal{A}$ into its ultrapower $\mathcal{A}^I/\mathcal{D}$.*

Proof. It suffices to verify that for any formula $\varphi(x_1, \dots, x_n)$ and arbitrary elements $a_1, \dots, a_n \in A$, the condition $\mathcal{A} \vDash \varphi(a_1, \dots, a_n)$ implies that $\mathcal{A}^I/\mathcal{D} \vDash \varphi\big(\overline{a}_1^{\mathcal{D}}, \dots, \overline{a}_n^{\mathcal{D}}\big)$. Under this assumption, however, we have $[\varphi(\overline{a}_1, \dots, \overline{a}_n)] = I \in \mathcal{D}$, so the desired conclusion follows from Łoś' theorem.

8.4.4 Exercise. Let $\mathcal{A} = (A; \dots)$ be a structure of infinite cardinality $|A| = \alpha$, such that $\alpha^{\aleph_0} = \alpha$ (e.g., this is the case of the power of the continuum $\alpha = 2^{\aleph_0}$). Let further I be an infinite countable set (e.g., $I = \mathbb{N}$) and $\mathcal{D}$ be a nonprincipal ultrafilter on I. Prove that the ultrapower $\mathcal{A}^I/\mathcal{D}$ is a proper elementary extension of the structure $\mathcal{A}$ with the same cardinality $\left|A^I/\mathcal{D}\right| = \alpha$.

8.5 Nonstandard Analysis

The construction of ultrapowers makes it possible to extend the domain of real numbers into a numerical system that includes "infinitesimally small" and "infinitely large" quantities and to use them in developing infinitesimal calculus, i.e., differential and integral calculus, in a form close to its historical origin, as conceived by its founders, Isaac Newton and Gottfried Wilhelm Leibniz. The mathematical discipline that, using methods of mathematical logic and model theory, studies various classical mathematical structures through their elementary extensions, where they are enriched with different types of ideal elements, is called *nonstandard analysis*. Its foundations were laid in the 1960s by Abraham Robinson. "Nonstandard methods" are currently used in various areas of mathematics, such as measure theory and probability, topology, functional analysis, ergodic theory, dynamical systems, etc., so the significance of nonstandard analysis extends far beyond the rehabilitation of the original infinitesimal calculus. In this brief introduction, however, we will outline only a few of its fundamental ideas related to the infinitesimal calculus of real functions of one real variable.

For concreteness, let us choose the index set I to be the set of all positive integers, i.e., $I = 1, 2, 3, \dots$; however, any infinite set could serve this role. Further, let $\mathcal{D}$ be any non-principal ultrafilter on the set I; this means that $\mathcal{F}(I) \subseteq \mathcal{D}$. For any structure $\mathcal{A}$ of an arbitrary first-order language L, we denote its ultrapower $\mathcal{A}^I/\mathcal{D}$ as ${}^*\mathcal{A}$. Specifically, for any set A, we define ${}^*A = A^I/\mathcal{D}$, and we identify each element $a \in A$ with its image $\overline{a}^{\mathcal{D}}$ in the diagonal embedding $A \to {}^*A$.

The set ${}^*\mathbb{R} = \mathbb{R}^I/\mathcal{D}$ is called the set of all *hyperreal numbers*. Similarly, one can define *hypernatural, hyperinteger, hyperrational,* and *hypercomplex numbers*. We agree that $L = (F, R, \nu)$ is a first-order language whose specific symbols include the usual symbols for addition $+$ and multiplication $\cdot$, the constants 0 and 1, the ordering relations $<$ and $\leq$, as well as the standard symbols for all functions, operations, constants, and relations commonly used in mathematical analysis, such as the absolute value function $|x|$, the exponential function e^x, sine $\sin x$, cosine $\cos x$, division operation x/y, exponentiation x^y, constants π, e, unary predicates $\mathbb{Z}(x)$ "to be an integer", $\mathbb{Q}(x)$ "to be a rational number", $\mathbb{R}^+(x)$ "to be a positive (real) number", and so on. Then, $\mathcal{R}$ denotes the L-structure with the underlying set $\mathbb{R}$, where all specific symbols of the language L are interpreted in the natural (i.e., standard) way. The ultrapower

$$ {}^*\mathcal{R} = \mathcal{R}^I/\mathcal{D} $$

is called the *nonstandard extension* of the structure $\mathcal{R}$. While the interpretations of the specific symbols of the language L in the structure $\mathcal{R}$ will be denoted by the corresponding symbol itself, the interpretation of a symbol s in the structure ${}^*\mathcal{R}$ will

be denoted by ${}^*s$. When no confusion arises, we will omit this asterisk, especially for symbols with established meanings, such as $+$, $\cdot$, 0, 1, π, e, $|x|$, $\sin x$, $\cos x$, $<$, etc. For any function $f\colon P \to \mathbb{R}$ that we intend to study, we assume that f and its domain $P \subseteq \mathbb{R}$ "have names" expressed by specific symbols, or at least terms or formulas of the language L. To such a function then corresponds a function ${}^*f\colon {}^*P \to {}^*\mathbb{R}$ that satisfies the conditions $P \subseteq {}^*P \subseteq {}^*\mathbb{R}$ and ${}^*f(x) = f(x)$ for $x \in P$. Since $\mathcal{R} \prec {}^*\mathcal{R}$, ${}^*f$ has, in certain well-defined sense, "the same properties" as the original function f.

Restricting the structures $\mathcal{R}$ and ${}^*\mathcal{R}$ to the language with specific symbols $+$, $\cdot$, 0, 1, and $<$, we obtain the ordered fields $\mathcal{R}_0 = (\mathbb{R}; +, \cdot, 0, 1, <)$ and ${}^*\mathcal{R}_0 = ({}^*\mathbb{R}; +, \cdot, 0, 1, <)$, where $\mathcal{R}_0 \prec {}^*\mathcal{R}_0$ still holds. The ordered field ${}^*\mathcal{R}_0$, however, contains "infinitely large" and "infinitesimally small" numbers (quantities). We say that a hyperreal number $x \in {}^*\mathbb{R}$ is *finitely large*, or alternatively *bounded* or simply *finite*, if there exists a positive real number $r \in \mathbb{R}$ such that $|x| < r$. Otherwise, we say that x is *infinitely large*, or alternatively *unbounded* or simply *infinite*. We say that a hyperreal number $x \in {}^*\mathbb{R}$ is *infinitesimally small*, or simply *infinitesimal*, if for every positive real number $r \in \mathbb{R}$, we have $|x| < r$. It is easy to see that the set

$$\mathbb{F}^*\mathbb{R} = \bigl\{x \in {}^*\mathbb{R}\colon (\exists\, r \in \mathbb{R}^+)(|x| < r)\bigr\}$$

of all finite hyperreal numbers forms a *convex subring* [6] of the ordered field ${}^*\mathcal{R}_0$. The set

$$\mathbb{I}^*\mathbb{R} = \bigl\{x \in {}^*\mathbb{R}\colon (\forall\, r \in \mathbb{R}^+)(|x| < r)\bigr\}$$

of all infinitesimally small hyperreal numbers forms a *convex ideal* of the ordered ring $(\mathbb{F}^*\mathbb{R}; +, \cdot, 0, 1, <)$.

The reader should verify that for the sequence $\alpha = (1, 2, 3, \dots) \in \mathbb{R}^I$, i.e., $\alpha(i) = i$ for $i \in I$, we have $\alpha^{\mathcal{D}} \in {}^*\mathbb{R} \setminus \mathbb{F}^*\mathbb{R}$, which means that $\alpha^{\mathcal{D}}$ is a positive infinitely large hyperreal number. On the other hand, for the sequence $\beta = (1, 1/2, 1/3, \dots) \in \mathbb{R}^I$, i.e., $\beta(i) = 1/i$ for $i \in I$, we have $0 \neq \beta^{\mathcal{D}} \in \mathbb{I}^*\mathbb{R}$, meaning that $\beta^{\mathcal{D}}$ is a positive infinitesimally small hyperreal number. In general, a hyperreal number x is infinitely large if and only if its reciprocal $1/x$ is infinitesimally small.

Since the set $\mathbb{I}^*\mathbb{R}$ of infinitesimally small numbers is also a subgroup of the abelian group $({}^*\mathbb{R}; +, 0)$, the relation

$$x \approx y \;\Leftrightarrow\; x - y \in \mathbb{I}^*\mathbb{R}$$

defines an equivalence relation on the set ${}^*\mathbb{R}$, called the *relation of infinitesimal nearness*, or *indiscernibility equivalence*. The set

$$\mu(x) = x + \mathbb{I}^*\mathbb{R} = \{y \in {}^*\mathbb{R}\colon y \approx x\}$$

of all hyperreal numbers infinitely close to the number (point) $x \in {}^*\mathbb{R}$ is called, following Leibniz, the *monad* of the number (point) x. Apparently, $\mu(0) = \mathbb{I}^*\mathbb{R}$ and each monad $\mu(x)$ is a convex subset of the set ${}^*\mathbb{R}$.

[6] A subset X of a partially ordered set $(A; <)$ is called *convex* if for all $x, y, z \in A$, the condition $(x, z \in X \wedge x < y < z)$ implies $y \in X$.

For every *finite* hyperreal number x, there exists exactly one (standard) real number r such that $x \approx r$; this real number is called the *standard part* or *shadow* of the number x and is denoted as

$$r = \operatorname{st} x = {}^\circ x$$

For a reader with some knowledge of topology, we add that for $x \in \mathbb{F}^*\mathbb{R}$ determined by a sequence $\xi \in \mathbb{R}^I$, i.e., $x = \xi^{\mathcal{D}}$, the equality ${}^\circ x = r$ holds if and only if the sequence ξ converges to the point r *with respect to the ultrafilter* $\mathcal{D}$.[7] From this, it follows that the convergence

$$\lim_{i \to \infty} \xi(i) = r$$

in the "usual" sense (i.e., according to the Fréchet filter $\mathcal{F}(I)$) is a sufficient (though not necessary) condition for the equality ${}^\circ x = r$.

From our considerations, it follows that the quotient set $\mathbb{F}^*\mathbb{R}/\mathbb{I}^*\mathbb{R} = \mathbb{F}^*\mathbb{R}/{\approx}$ consists of the monads $\mu(x)$ of bounded hyperreal numbers $x \in \mathbb{F}^*\mathbb{R}$. By taking the quotient of the ordered ring $(\mathbb{F}^*\mathbb{R}; +, \cdot, 0, 1, <)$ with respect to its convex ideal $\mathbb{I}^*\mathbb{R}$, we again obtain an ordered ring $(\mathbb{F}^*\mathbb{R}/\mathbb{I}^*\mathbb{R}; +, \cdot, 0, 1, <)$. Specifically, for the ordering in this ring, the following holds:

$$\mu(x) < \mu(y) \;\Leftrightarrow\; (x < y \wedge x \not\approx y)$$
$$\mu(x) \leq \mu(y) \;\Leftrightarrow\; (x < y \vee x \approx y)$$

Moreover, by assigning $\mu(x) \mapsto \operatorname{st} x$, a well-defined isomorphism of ordered rings is established: $(\mathbb{F}^*\mathbb{R}/\mathbb{I}^*\mathbb{R}; +, \cdot, 0, 1, <) \cong (\mathbb{R}; +, \cdot, 0, 1, <)$. From this, it follows, among other things, that $(\mathbb{F}^*\mathbb{R}/\mathbb{I}^*\mathbb{R}; +, \cdot, 0, 1, <)$ is an ordered field.

8.5.1 Exercise. Similarly, we can define the sets of all finite hyperrational numbers $\mathbb{F}^*\mathbb{Q}$ and of all infinitely small hyperrational numbers $\mathbb{I}^*\mathbb{Q}$, and form the quotient ordered ring $(\mathbb{F}^*\mathbb{Q}/\mathbb{I}^*\mathbb{Q}; +, \cdot, 0, 1, <)$. Consider which of the two possibilities

$$(\mathbb{F}^*\mathbb{Q}/\mathbb{I}^*\mathbb{Q}; +, \cdot, 0, 1, <) \cong (\mathbb{Q}; +, \cdot, 0, 1, <)$$

or

$$(\mathbb{F}^*\mathbb{Q}/\mathbb{I}^*\mathbb{Q}; +, \cdot, 0, 1, <) \cong (\mathbb{R}; +, \cdot, 0, 1, <)$$

actually takes place, and explain why.

The fact that the ordered fields $\mathcal{R}_0 = (\mathbb{R}; +, \cdot, 0, 1, <)$ and ${}^*\mathcal{R}_0 = ({}^*\mathbb{R}; +, \cdot, 0, 1, <)$ are elementarily equivalent, which follows from the elementary inclusion $\mathcal{R}_0 \prec {}^*\mathcal{R}$, merely means that $\mathcal{R}_0$ and ${}^*\mathcal{R}_0$ share the same properties expressible in the language of ordered fields. However, regarding other properties, they may differ significantly. For example, in $\mathcal{R}_0$, the least upper bound property holds, meaning that every bounded set $X \subseteq \mathbb{R}$ has a supremum in $\mathbb{R}$. On the other hand, the sets $\mathbb{I}^*\mathbb{R} \subseteq {}^*\mathbb{R}$ and $\mathbb{F}^*\mathbb{R} \subseteq {}^*\mathbb{R}$ are both bounded in ${}^*\mathbb{R}$ (the former by any positive number $r \in \mathbb{R}$, the latter by any infinite positive number $z \in {}^*\mathbb{R}$). However, it is easy to verify that

[7] A sequence $\alpha\colon I \to \mathbb{R}$ converges to a real number a with respect to a filter $\mathcal{D}$ on the set I, if for every real number $\varepsilon > 0$ there exists a set $X \in \mathcal{D}$ such that $|\alpha(k) - a| < \varepsilon$ for every $k \in X$.

neither of them has a supremum in ${}^*\mathbb{R}$. The fact that the structure ${}^*\mathcal{R}_0$ is an elementary extension of the structure $\mathcal{R}_0$ implies the existence of a supremum only for all bounded sets of the form $X = \{x \in {}^*\mathbb{R} \colon \varphi(x, a_1, \ldots, a_n)\}$ where $\varphi(x, x_1, \ldots, x_n)$ is a formula in the language of ordered fields and $a_1, \ldots, a_n \in {}^*\mathbb{R}$. Consequently, we see that the sets $\mathbb{I}{}^*\mathbb{R}$, $\mathbb{F}{}^*\mathbb{R}$ cannot be expressed in this form. Similar considerations apply to the richer language L and the L-structures $\mathcal{R} \prec {}^*\mathcal{R}$.

The following three theorems, which we state without proofs, characterize three fundamental concepts of mathematical analysis (continuity, differentiation, and definite integration) using infinitesimal and infinitely large quantities. At the same time, they can be understood as *definitions* of these concepts. However, in that case, it would be appropriate to prove that the corresponding standard definitions of these concepts are equivalent to them. We leave it to the reader's personal judgment whether they find the standard or nonstandard formulations more intuitively appealing.

8.5.2 Theorem. *Let $P \subseteq \mathbb{R}$ and $f \colon P \to \mathbb{R}$ be any function named in the language L. Then*

(a) *f is continuous at the point $a \in P$ if and only if for all $x \in {}^*P$, it holds that $x \approx a \Rightarrow {}^*f(x) \approx f(a)$;*

(b) *f is continuous on the set P if and only if for all $a \in P$, $x \in {}^*P$, it holds that $x \approx a \Rightarrow {}^*f(x) \approx f(a)$;*

(c) *f is uniformly continuous on the set P if and only if for all $x, y \in {}^*P$, it holds that $x \approx y \Rightarrow {}^*f(x) \approx {}^*f(y)$.*

We recall that a point $a \in P$ of a set $P \subseteq \mathbb{R}$ is called its *interior point* if there exists a positive number $\varepsilon \in \mathbb{R}$ such that $(a - \varepsilon, a + \varepsilon) \subseteq P$. It can be shown that $a \in P$ is an interior point of this set if and only if $\mu(a) \subseteq {}^*P$.

8.5.3 Theorem. *Let $P \subseteq \mathbb{R}$, $f \colon P \to \mathbb{R}$ be any function named in the language L, and $a \in P$ be an interior point of the set P. Then the function f has a finite derivative at the point a if and only if there exists a real number r such that for every infinitesimally small number $d \neq 0$, the following holds:*

$$\frac{{}^*f(a + d) - f(a)}{d} \approx r$$

In such a case, we write

$$f'(a) = \frac{\mathrm{d}f(a)}{\mathrm{d}x} = r = \mathrm{st}\left(\frac{{}^*f(a + d) - f(a)}{d}\right)$$

and this number is called the *derivative of the function f at the point a*.

The Riemann definite integral can be defined using "infinite integral sums" corresponding to *infinitesimal partitions* of a given interval. A rigorous introduction of such sums would require additional finer considerations, which we will omit this time. However, the fundamental ideas on which the following concepts and the associated

construction are based are quite intuitive. It suffices to note that a set or a sequence is called *hyperfinite* if it can be meaningfully assigned some hypernatural number $n \in {}^*\mathbb{N}$ as the number of its elements or members. If $n \in {}^*\mathbb{N} \setminus \mathbb{N}$, then this set or sequence is infinite in the standard sense. The numbers forming a hyperfinite sequence $(c_1, \ldots, c_n) \in {}^*\mathbb{R}^n$ can also be summed, and their sum $\sum_{i=1}^{n} c_n$ can be handled in a well-defined sense just like a finite sum.

Let $a < b$ be (standard) real numbers. A *partition of the interval* $[a, b]$ is a hyperfinite sequence $(c_0, c_1, \ldots, c_n)$, where $n \in {}^*\mathbb{N}$, such that $a = c_0 < c_1 < \ldots < c_n = b$. We say that $(c_0, c_1, \ldots, c_n)$ is an *infinitesimal partition* of the interval $[a, b]$ if $c_{k-1} \approx c_k$ for every $1 \leq k \leq n$. Clearly, if $(c_0, c_1, \ldots, c_n)$ is an infinitesimal partition of the interval $[a, b]$, then $n \in {}^*\mathbb{N} \setminus \mathbb{N}$, meaning n is an infinitely large hypernatural number. The *integral sum* corresponding to the partition $(c_0, c_1, \ldots, c_n)$ is defined as

$$\sum_{k=1}^{n} {}^*f(x_k)(c_k - c_{k-1})$$

where $(x_1, \ldots, x_n)$ is a hyperfinite sequence of selected points $x_k \in [c_{k-1}, c_k]$ from the subintervals $[c_{k-1}, c_k] \subseteq [a, b]$.

8.5.4 Theorem. *Let $a < b$ be real numbers, and let $f\colon [a, b] \to \mathbb{R}$ be any function named in the language L. Then the function f is Riemann integrable on the interval $[a, b]$ if and only if there exists a real number S such that for all infinitesimal partitions $(c_0, c_1, \ldots, c_n)$ of the interval $[a, b]$ and chosen points $x_k \in [c_{k-1}, c_k]$ for $1 \leq k \leq n$, the following holds:*

$$\sum_{k=1}^{n} {}^*f(x_k)(c_k - c_{k-1}) \approx S$$

In such a case, we write

$$\int_a^b f(x)\,\mathrm{d}x = S = \mathrm{st}\left(\sum_{k=1}^{n} {}^*f(x_k)(c_k - c_{k-1})\right)$$

and this number is called the *definite integral of the function f on the interval* $[a, b]$. Specifically, if $n \in {}^*\mathbb{N} \setminus \mathbb{N}$, $d = (b - a)/n \approx 0$, and $c_k = a + kd$ for $0 \leq k \leq n$, then $(c_0, c_1, \ldots, c_n)$ is a *uniform* infinitesimal partition of the interval $[a, b]$. In this case, for a Riemann integrable function $f\colon [a, b] \to \mathbb{R}$ and any choice of points $x_k \in [c_{k-1}, c_k]$, we have

$$\int_a^b f(x)\,\mathrm{d}x = S = \mathrm{st}\left(\sum_{k=1}^{n} {}^*f(x_k)\,d\right) = (b - a)\,\mathrm{st}\left(\frac{1}{n}\sum_{k=1}^{n} {}^*f(x_k)\right)$$

which means that the standard part of the arithmetic mean $\frac{1}{n}\sum_{k=1}^{n} {}^*f(x_k)$ represents the *average* or the *mean value* of the function f on the interval $[a, b]$.

Finally, let us note something that some attentive readers may have already noticed on their own. The fact that we constructed the structure ${}^*\mathcal{R}$ over the set of hyperreal numbers ${}^*\mathbb{R}$ as an ultrapower ${}^*\mathcal{R} = \mathcal{R}^I/\mathcal{D}$ of the structure $\mathcal{R}$ over the set of real

numbers $\mathbb{R}$ played no essential role in our considerations. The crucial aspect was only that the structure ${}^*\mathcal{R}$ is an elementary extension of the structure $\mathcal{R}$, along with the intuition and appealing transparency associated with the ideas of infinitesimally small and infinitely large numbers. The ultrapower construction merely provided an initial foundation by essentially "legalizing" them within the universe of sets, which we have come to regard as the realm of the decisive part of modern mathematics.

8.6 The Compactness Theorem in the Language of Ultraproducts

Our original proof of the Compactness Theorem 4.9.1 was based on the Completeness Theorem 4.8.6 and it was non-constructive in the sense that it provided no method for constructing a model of a first-order theory T from models of its finite subtheories. The following theorem gives us some insight into such a construction. However, it is important to realize that non-principal ultrafilters cannot be explicitly described, and the axioms of set theory guarantee only their existence. Therefore, the "constructiveness" of the following theorem remains largely illusory.

8.6.1 Compactness Theorem. *Let Σ be a set of sentences in the language L, closed under finite conjunctions, and for each $\sigma \in \Sigma$, let $\mathcal{A}_\sigma$ be a structure in the language L such that $\mathcal{A}_\sigma \vDash \sigma$. Then there exists an ultrafilter $\mathcal{D}$ on the set Σ such that*

$$\prod_{\sigma\in\Sigma} \mathcal{A}_\sigma \Big/ \mathcal{D} \vDash \Sigma$$

Proof. For each $\sigma \in \Sigma$, we denote $J_\sigma = \{\varrho \in \Sigma : \mathcal{A}_\varrho \vDash \sigma\}$. Clearly, the set $J_{\sigma_1} \cap \cdots \cap J_{\sigma_n}$ contains the formula $\sigma_1 \wedge \cdots \wedge, \sigma_n$, so $\{J_\sigma : \sigma \in \Sigma\}$ is a centered system of subsets of the set Σ. Therefore, there exists an ultrafilter $\mathcal{D}$ on Σ such that $J_\sigma \in \mathcal{D}$ for each $\sigma \in \Sigma$. We will prove that

$$\prod_{\varrho\in\Sigma} \mathcal{A}_\varrho \Big/ \mathcal{D} \vDash \sigma$$

for any $\sigma \in \Sigma$. It suffices to realize that

$$[\sigma] = \{\varrho \in \Sigma \colon \mathcal{A}_\varrho \vDash \sigma\} = J_\sigma \in \mathcal{D}$$

The required conclusion follows from Łoś' theorem 8.4.1.

8.7 Elementary Equivalence and Ultraproducts

The just-proven version of the Compactness Theorem enables to characterize the relationship of elementary equivalence of structures through the elementary embedding of one of them into an ultrapower of the other.

8.7.1 Theorem. *Let $\mathcal{A}$, $\mathcal{B}$ be structures of the language L. Then $\mathcal{A} \equiv \mathcal{B}$ if and only if there exists an ultrapower $\mathcal{A}^I/\mathcal{D}$ of $\mathcal{A}$ and an elementary embedding $h \colon \mathcal{B} \xrightarrow{\prec} \mathcal{A}^I/\mathcal{D}$.*

Proof. Let $\mathcal{A} \equiv \mathcal{B}$. Put $\Sigma = \mathrm{Th}(\mathcal{B}_B)$. Clearly, Σ is a set of sentences of the language L_B closed under finite conjunctions. Any $\sigma \in \Sigma$ has the form $\varphi(b_1, \dots, b_n)$ for an appropriate formula $\varphi(x_1, \dots, x_n)$ of the language L and $b_1, \dots, b_n \in B$. Then $\mathcal{B} \vDash \varphi(b_1, \dots, b_n)$, and since $\mathcal{A} \equiv \mathcal{B}$, also

$$\mathcal{A} \vDash (\exists\, x_1, \dots, x_n)\varphi(x_1, \dots, x_n)$$

Let $g_\sigma \colon B \to A$ be a mapping such that for the structure $\mathcal{A}_\sigma = \big(\mathcal{A}, g_\sigma(b)\big)_{b \in B}$ of the language L_B, we have

$$\mathcal{A} \vDash \varphi\big(g_\sigma(b_1), \dots, g_\sigma(b_n)\big)$$

i.e., $\mathcal{A}_\sigma \vDash \sigma$. By the Compactness Theorem 8.6.1, there exists an ultrafilter $\mathcal{D}$ on Σ such that

$$\prod_{\sigma \in \Sigma} \mathcal{A}_\sigma \Big/ \mathcal{D} \vDash \Sigma$$

Then, the restriction of this ultraproduct to the language L is the ultrapower $\mathcal{A}^\Sigma/\mathcal{D}$, and the mapping $h \colon B \to A^\Sigma/\mathcal{D}$, given by $h(b)(\sigma) = g_\sigma(b)$ for $b \in B$, $\sigma \in \Sigma$, is an elementary embedding $\mathcal{B} \xrightarrow{\prec} \mathcal{A}^\Sigma/\mathcal{D}$.

The reverse implication is trivial.

Note that using techniques beyond the scope of our course, one can prove a significantly stronger result.

8.7.2 Keisler-Shelah Theorem. *Let $\mathcal{A}$, $\mathcal{B}$ be first-order structures. Then $\mathcal{A} \equiv \mathcal{B}$ if and only if there exist a set I and an ultrafilter $\mathcal{D}$ on I such that $\mathcal{A}^I/\mathcal{D} \cong \mathcal{B}^I/\mathcal{D}$.*

8.8 Characterization of Axiomatic and Finitely Axiomatizable Classes

We denote the class of all structures of the language L by $\mathrm{Mod}(L)$. If T is a theory in the language L, then we denote the class of all its models by $\mathrm{Mod}(T)$, that is,

$$\mathrm{Mod}(T) = \{\mathcal{A} \in \mathrm{Mod}(L) \colon \mathcal{A} \vDash T\}$$

We say that a class $\mathbf{K} \subseteq \mathrm{Mod}(L)$ of structures of the language L is an *axiomatic class*, if there exists a theory T in the language L such that $\mathbf{K} = \mathrm{Mod}(T)$. We say that a class $\mathbf{K}$ of structures of the language L is *finitely axiomatizable*, if there exists a finite theory in the language L such that $\mathbf{K} = \mathrm{Mod}(T)$.

The *theory of the class* $\mathbf{K} \subseteq \mathrm{Mod}(L)$ is defined as the set $\mathrm{Th}(\mathbf{K})$ of all sentences of the language L satisfied in every structure $\mathcal{A} \in \mathbf{K}$, i.e.,

$$\mathrm{Th}(\mathbf{K}) = \big\{\varphi \in \mathrm{Form}(L) \colon \varphi \text{ is closed a } (\forall\, \mathcal{A} \in \mathbf{K})(\mathcal{A} \vDash \varphi)\big\}$$

The following two theorems characterize axiomatic and finitely axiomatizable classes in terms of their closure properties with respect to isomorphism, elementary equivalence, elementary substructures, and ultraproducts.

8.8.1 Theorem. *Let* $\mathbf{K}$ *be an arbitrary class of structures in the language* L*. The following conditions are equivalent:*

(i) $\mathbf{K}$ *is an axiomatic class.*

(ii) $\mathbf{K}$ *is closed under isomorphism, elementary substructures, and ultraproducts.*

(iii) $\mathbf{K}$ *is closed under elementary equivalence and ultraproducts.*

Proof. (i) $\Rightarrow$ (ii) is trivial, (ii) $\Rightarrow$ (iii) follows from Theorem 8.7.1. It remains to prove (iii) $\Rightarrow$ (i).

Assume that $\mathbf{K}$ is closed under elementary equivalence and ultraproducts. Let $T = \mathrm{Th}(\mathbf{K})$. Clearly, $\mathbf{K} \subseteq \mathrm{Mod}(T)$. We prove the reverse inclusion. Let $\mathcal{B} \in \mathrm{Mod}(T)$. Then $\Sigma = \mathrm{Th}(\mathcal{B})$ is a set of sentences in the language L that is closed under finite conjunctions. Also, for every $\sigma \in \Sigma$, there exists some $\mathcal{A}_\sigma \in \mathbf{K}$ such that $\mathcal{A}_\sigma \vDash \sigma$. Otherwise, we would have $\neg\sigma \in T$, and thus $\mathcal{B} \models \neg\sigma$. By the Compactness Theorem 8.6.1, there exists an ultrafilter $\mathcal{D}$ on the set Σ such that $\prod_{\sigma\in\Sigma} \mathcal{A}_\sigma/\mathcal{D} \vDash \Sigma$, hence $\prod \mathcal{A}_\sigma/\mathcal{D} \equiv \mathcal{B}$. Therefore, $\mathcal{B} \in \mathbf{K}$.

8.8.2 Theorem. *Let* $\mathbf{K}$ *be an arbitrary class of structures in the language* L*. The following conditions are equivalent:*

(i) $\mathbf{K}$ *is a finitely axiomatizable class.*

(ii) *Both* $\mathbf{K}$ *and* $\mathrm{Mod}(L) \setminus \mathbf{K}$ *are axiomatic classes.*

(iii) $\mathbf{K}$ *is an axiomatic class and the class* $\mathrm{Mod}(L) \setminus \mathbf{K}$ *is closed under ultraproducts.*

Proof. (i) $\Rightarrow$ (ii) is almost obvious. If $\mathbf{K}$ is a finitely axiomatizable class, then there exists a closed L-formula σ (the universal closure of the conjunction of all axioms defining the class $\mathbf{K}$) such that $\mathbf{K} = \mathrm{Mod}(\sigma)$. Then for the complement $\mathbf{K}' = \mathrm{Mod}(L) \setminus \mathbf{K}$ in the class of all L-structures, we have $\mathbf{K}' = \mathrm{Mod}(\neg\sigma)$, meaning that $\mathbf{K}'$ is also a finitely axiomatizable class.

As (ii) $\Rightarrow$ (iii) is trivial, it remains to prove (iii) $\Rightarrow$ (i). Let $\Sigma = \mathrm{Th}(\mathbf{K})$. Clearly, Σ is a set of sentences in the language L that is closed under finite conjunctions. Since $\mathbf{K}$ is an axiomatic class, we have $\mathbf{K} = \mathrm{Mod}(\Sigma)$. We show that there exists some $\sigma \in \Sigma$ such that $\mathbf{K} = \mathrm{Mod}(\sigma)$. Otherwise, for every $\sigma \in \Sigma$, there would exist a structure $\mathcal{A}_\sigma \in \mathbf{K}'$ such that $\mathcal{A}_\sigma \vDash \sigma$. By the Compactness Theorem 8.6.1, there exists an ultrafilter $\mathcal{D}$ on the set Σ such that $\prod_{\sigma\in\Sigma} \mathcal{A}_\sigma/\mathcal{D} \vDash \Sigma$, i.e., $\prod \mathcal{A}_\sigma/\mathcal{D} \in \mathbf{K}$. Since $\mathbf{K}'$ is closed with respect to ultraproducts, we have also $\prod \mathcal{A}_\sigma/\mathcal{D} \in \mathbf{K}'$, which is a contradiction.

8.8.3 Exercise. Let S, T be theories (whose axioms are closed formulas) in the language L, and let $\mathbf{J}$, $\mathbf{K}$ be classes of L-structures.

(a) Prove the following relations:

$$S \subseteq T \;\Rightarrow\; \mathrm{Mod}(T) \subseteq \mathrm{Mod}(S) \qquad T \subseteq \mathrm{Th}(\mathrm{Mod}\, T)$$
$$\mathbf{J} \subseteq \mathbf{K} \;\Rightarrow\; \mathrm{Th}(\mathbf{K}) \subseteq \mathrm{Th}(\mathbf{J}) \qquad \mathbf{K} \subseteq \mathrm{Mod}(\mathrm{Th}\, \mathbf{K})$$

(b) Derive from (a) the equalities

$$\mathrm{Mod}(\mathrm{Th}(\mathrm{Mod}\, T)) = \mathrm{Mod}(T) \qquad\qquad \mathrm{Th}(\mathrm{Mod}(\mathrm{Th}\, \mathbf{K})) = \mathrm{Th}(\mathbf{K})$$

(c) Prove that the set $\mathrm{Th}(\mathrm{Mod}\, T)$ consists precisely of all closed L-formulas φ such that $T \vdash \varphi$.

(d) Prove that $\mathrm{Mod}(\mathrm{Th}\, \mathbf{K})$ is the smallest axiomatic class $\mathbf{M} \subseteq \mathrm{Mod}(L)$ such that $\mathbf{K} \subseteq \mathbf{M}$.

(e) Prove that the class $\mathrm{Mod}(\mathrm{Th}\, \mathbf{K})$ consists precisely of all L-structures $\mathcal{B}$ that are isomorphic to elementary substructures of ultraproducts of all possible systems $(\mathcal{A}_i)_{i \in I}$ of structures $\mathcal{A}_i \in \mathbf{K}$.

Selected Bibliography

Barwise, J. (ed.) [1977], *Handbook of Mathematical Logic*, North Holland, Amsterdam – New York – Oxford.

Chang, C. C., Keisler, H. J. [1990], *Model Theory*, Studies in Logic and the Foundations of Mathematics, Elsevier, North Holland, Amsterdam, (3^{rd} ed.).

Dauben, J. W. [1979], *Georg Cantor: His Mathematics and Philosophy of the Infinite*, Harvard University Press, Cambridge, Mass.

Davis, M. (ed.) [1965], *The Undecidable. Basic Papers on Undecidable Propositions, Unsolvable Problems and Computable Functions*, Raven Press, Hewlett, N. Y.

Davis, M. [1977], *Applied Nonstandard Analysis*, John Wiley & Sons, New York – London.

Davis, M. [2018], *Pragmatic Platonism, Mathematics and the Infinite*, in Hellman, G., Cook, R. T. (eds.), *Hillary Putnam on Mathematical Logic*, Springer-Verlag, Cham (Switzerland), pages 145 – 159;
`https://www.researchgate.net/publication/329449494%_Pragmatic_Platonism_Mathematics_and_the_Infinite`

Davis, M. [2005], *What Did Gödel Believe and When Did He Believe It?* Bull. Symbolic Logic **11**, 194 – 206;
`https://www.jstor.org/stable/1556749?seq=1%#metadata_info_tab_contents`

Ebbinghaus, H.-D., Flum, J. [1999], *Finite Model Theory*, Perspectives in Mathematical Logic, Springer, Berlin – Heidelberg – New York (2^{nd} ed.).

Enderton, H. B. [1977], *Elements of Set Theory*, Aacademic Press, New York.

Enderton, H. B. [2001], *A Mathematical Introduction to Logic*, Academic Press, New York (2^{nd} ed.).

Feferman, S. [2006], *The Impact of the Incompleteness Theorems on Mathematics*, Notices Amer. Math. Soc. **53** (April), 434 – 439;
`https://www.ams.org/notices/200604/fea-feferman.pdf`

Franzen, T. [2005], *Gödel's Theorem: An Incomplete Guide to Its Use and Abuse*, A K Peters, Wellesley, Mass.

Gaifman, H., *What Gödel's Incompleteness Result Does and Does Not Show*;
`https://pdfs.semanticscholar.org/1efb/5896951d8c20984f558c1e4ab4d6d029143d.pdf`

Goldblatt, R. [1998], *Lectures on the Hyperreals. An Introduction to Nonstandard Analysis*, Springer-Verlag, New York.

Hills, M., Loeser, F. [2019], *A First Journey through Logic*, Student Mathematical Library vol. 89, Amer. Math. Soc., Providence, Rhode Island.

Hodges, W. [1993], *Model Theory*, Cambridge University Press, Cambridge – New York.

Hodges, W. [1997], *A Shorter Model Theory*, Cambridge University Press, Cambridge – New York.

Jech, T. [1978], *Set Theory*, Academic Press, New York.

Kneale, W., Kneale, M. [1984], *The Development of Logic*, Clarendon Press, Oxford.

Kossak, R. [2024], *Mathematical Logic. On Numbers, Sets, Structures and Symmetry*, Springer Graduate Texts in Philosophy, vol. 4, Springer, Cham, Switzerland (2^{nd} ed.).

Manin, Yu. I. [2010], *A Course in Mathematical Logic for Mathematicians*, Graduate Texts in Mathematics, Springer, New York – Heidelberg – London (2^{nd} ed. with collaboration by B. Zilber).

Marker, D. [2002], *Model Theory, An Introduction*, Graduate Texts in Mathematics, Springer, New York – London.

Mendelson, E. [1964], *Introduction to Mathematical Logic*, D. Van Nostrand, Princeton, New Jersey.

Raatikainen, P. [2020], *Gödel's Incompleteness Theorems*, Stanford Encyclopedia of Philosophy;
`https://plato.stanford.edu/entries/goedel-incompleteness/`

Restall, G. [2005], *Logic, An Introduction*, Routlege, Taylor & Francis Group, London – New York.

Robinson, A. [1963], *Introduction to Model Theory and to the Metamathematics of Algebra*, North Holland, Amsterdam.

Robinson, A. [1966], *Non-Standard Analysis*, North Holland, Amsterdam.

Shoenfield, J. R. [1967], *Mathematical Logic*, Addison-Wesley, Reading, Mass.; reprinted by Assoc. Symb. Logic, Urbana, Illinois and A K Peters, Natick, Mass. 2001.

Smith, P. [2013], *An Introduction to Gödel's Theorems*, Cambridge University Press, Cambridge – New York.

Smullyan, R. M. [1992], *Gödel's Incompleteness Theorems*, Oxford University Press, Oxford – New York.

Smullyan, R. M. [2014], *A Beginner's Guide to Mathematical Logic*, Dover Publications, New York.

Smullyan, R. M. [2017], *A Beginner's Further Guide to Mathematical Logic*, World Scientific, Singapore.

Tarski, A. [1965], *Introduction to Logic and to the Methodology of Deductive Sciences*, Oxford University Press (3^{rd} ed.).

van Heijenoort, J. (ed.) [1967], *From Frege to Gödel. A Source Book in Mathematical Logic, 1879 – 1931*, Harvard University Press, Cambridge, Mass.

Wang, Hao [1987], *Reflections on Kurt Gödel*, MIT Press, Cambridge, Mass.

Wang, Hao [1996], *A Logical Journey: From Gödel to Philosophy*, MIT Press, Cambridge, Mass.

Zach, R. [2003], *Hilbert's Program*, Stanford Encyclopedia of Philosophy; `http://plato.stanford.edu/archives/fall2003/entries/hilbert-program/(2003)`

List of Symbols

Index of Subjects

Index of Names

Pavol Zlatoš

Lectures on Mathematical Logic

Published by Comenius University Bratislava, 2025
Graphic design: Mária Benešová
Cover design: Pavol Zlatoš, Marián Sekela

Manuscript 178 pages, 12.66 author's sheets, first edition
printed by Publishing Centre of Comenius University Bratislava

ISBN 978-80-223-6123-1 (print)
ISBN 978-80-223-6124-8 (online)