Propositional Calculus

Propositional Calculus, also called *Sentential Calculus* 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 platform for the exposition of some topics and issues which will reappear later on within the Predicate Calculus in a more advanced form.

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 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, ... , 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, ... , p_n and the way how A is formed, regardless of the meaning and content of the propositions p_1, ... , p_n.*

In other words, the truth value of A can be *computed* from the truth values of its components p_1, ... , 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, ... , q', q'', ...
- 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.

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° an 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 some property it is enough to show that the set of all propositional forms having this property satisfies the above conditions 1° and 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 reduce the number of logical connectives for

which the step 2° has to be performed to some minimal list. We will return to this point in the next paragraph.

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 *intepretation* 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

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

It follows that we could have used just the binary 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 $\vee$

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

for any propositional forms $A, B, C \in \mathrm{VF}(P)$. This allows us to omit the superfluous parenthesis in conjunctions and alternatives of an arbitrary 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

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

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 $\neg$, $\wedge$ and $\vee$, $\Rightarrow$ 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.

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 above: 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 an arbitrary propositional form A belongs to any of the four classes defined above can be decided algorithmically using the method of *truth 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.

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.

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.

Example. Using the truth 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.

Exercise. 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.

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 $F_A(e_1, \ldots, e_n) = A^I$ 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 $F = F_A$.

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).

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* or *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 tables, for any theory T with just finitely many specific axioms and each propositional form $B \in \mathrm{VF}(P)$.

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 $I(B) = 0$. If this is the case, then we say that (the validity of) B in T was refuted by a counterexample. However, in order to confirm that $T \vDash B$, the definition requires of us to determine the truth value $I(B)$ for infinitely many interpretations of T, which seems to be 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 these axioms are true and checking the validity of the statement in each of these 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.

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 Appendix to this Chapter.

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:

Deduction rule Modus Ponens

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

Exercise. (a) Show that all the logical axioms are tautologies and explain their intuitive meaning.
(b) Show that the inference 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 (MP) (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.

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 the Appendix. Both of 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" rules of inference. On the other hand, the *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.

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 just 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)

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).

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 false conclusions from the axioms of T which would be a disaster witnessing a

collapse of our axiomatization. Therefore it is of crucial importance that we have the following

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, the rule (MP) is correct.

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 contain some quantification, they interfere 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 formalized axiomatization of Propositional Calculus is in good accord with the logical structure of deduction and inference within our natural language.

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

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 $T \vdash B$.*

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 first 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 its certain fairly natural fragment not admitting any axiomatization satisfying both the Soundness and the Completeness Theorem.

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.

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 (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

$$\textit{If } T \cup \{A\} \vdash B \textit{ then } T \vdash A \Rightarrow B$$

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 the Exercise (c) it follows that every propositional form B is provable in an inconsistent theory T.

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 (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.

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 (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 $T \cup \{\neg A\} \vdash B$.

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 $T \cup \{A_i\} \vdash B$ for each $i = 1, \ldots, n$.

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., $I(A^I) = 1$.

Lemma on Interpretation. [A. Church] *Let $p_1, \dots, p_n \in P$ and $A \in \mathrm{VF}(p_1, \dots, p_n)$. Then for any interpretation $I\colon \mathrm{VF}(P) \to \{0,1\}$ we have*

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

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 $B \in \mathrm{VF}(p_1, \dots, p_n)$.

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

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 $\{p_1^I, \dots, p_n^I\} \vdash B^I$, i.e., $\{p_1^I, \dots, p_n^I\} \vdash B$. According to Exercise (b) we have $\vdash B \Rightarrow \neg\neg B$, and by (MP) we get $\{p_1^I, \dots, p_n^I\} \vdash \neg\neg B$, i.e., $\{p_1^I, \dots, p_n^I\} \vdash A^I$.

(c) Let $A \approx (B \Rightarrow C)$ and for B, C the conclusion is true. Then $B, C \in \mathrm{VF}(p_1, \dots, p_n)$. 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, $\{p_1^I, \dots, p_n^I\} \vdash \neg B$. According to Exercise (c) we have $\vdash \neg B \Rightarrow (B \Rightarrow C)$ and by (MP) we get $\{p_1^I, \dots, p_n^I\} \vdash B \Rightarrow C$, i.e., $\{p_1^I, \dots, p_n^I\} \vdash A^I$.

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, $\{p_1^I, \dots, p_n^I\} \vdash C$. (LAx 1) gives $\vdash C \Rightarrow (B \Rightarrow C)$, and by (MP) we get $\{p_1^I, \dots, p_n^I\} \vdash B \Rightarrow C$, i.e., $\{p_1^I, \dots, p_n^I\} \vdash A^I$.

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, $\{p_1^I, \dots, p_n^I\} \vdash B$ and $\{p_1^I, \dots, p_n^I\} \vdash \neg C$. Exercise (f) gives $\vdash B \Rightarrow (\neg C \Rightarrow \neg(B \Rightarrow C))$. Using (MP) twice we get $\{p_1^I, \dots, p_n^I\} \vdash \neg(B \Rightarrow C)$, i.e., $\{p_1^I, \dots, p_n^I\} \vdash A^I$.

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

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

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 (cf. Exercise...)

A special case of the Completeness Theorem deals with the provability of tautologies.

Completeness Theorem for Tautologies. [E. Post] *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,

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

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

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

According to Corollary on Proof by Distinct Cases this implies

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

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.

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 $\mathrm{VF}(P) = \{A_0, A_1, \ldots, A_n, \ldots\}$. Now we define a sequence of theories $T_0 \subseteq T_1 \subseteq \ldots \subseteq T_n \subseteq T_{n+1} \subseteq \ldots$ recursively:

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

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 as an exercise — see also the proof of the Compactness Theorem). 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.

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 $T_{n+1} = T_n \cup \{\neg A_n\}$.

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.

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.

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 way 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.

Appendix

Axiomatization of Propositional Calculus Using Four Logical Connectives

For completeness sake we include the axiomatization of Propostional 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.

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) $A \vee \neg A$

Deduction rule Modus Ponens

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

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

First Order Logic or Predicate Calculus

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.

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.

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, ... , y', y'', ...
- 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)

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 (base set or carrier) 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.

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-terms 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*.

Example. In the formula φ

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

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

Sentences or *closed formulas* 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 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:

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

For $\varphi(x_1, \ldots, x_n, y_1, \ldots, y_m)$ and $a_1, \ldots, a_n \in A$ we write $\mathcal{A} \vDash \varphi(a_1, \ldots, a_n, y_1, \ldots, y_m)$ if and only if $\mathcal{A} \vDash \varphi(a_1, \ldots, a_n, b_1, \ldots, b_m)$ for all $b_1, \ldots, b_m \in A$, i.e., if and only if $\mathcal{A} \vDash (\forall\, y_1, \ldots, y_m)\varphi(a_1, \ldots, a_n, y_1, \ldots, y_m)$. In particular, $\mathcal{A} \vDash \varphi(x_1, \ldots, x_n)$ means that $\mathcal{A} \vDash \varphi(a_1, \ldots, a_n)$ for all $a_1, \ldots, a_n \in A$, i.e., $\mathcal{A} \vDash (\forall\, x_1, \ldots, x_n)\varphi(x_1, \ldots, x_n)$.

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* $T \subseteq \mathrm{Form}(L)$. 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 grasp the Universe of Sets.

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 $\big(G; \cdot, e, {}^{-1}\big)$ 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.

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

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

stating 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 and the remainders modulo $n \geq 2$ form commutative rings with unit $(\mathbb{Z};\, +, \cdot, 0, 1)$ and $(\mathbb{Z}_n;\, +, \cdot, 0, 1)$, respectively. 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 $\mathrm{i}^2 = \mathrm{j}^2 = \mathrm{k}^2 = \mathrm{i}\,\mathrm{j}\,\mathrm{k} = -1$. 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.

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

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

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.

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 Total Order* or the *Theory of Linear 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.

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 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 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 order relation $X \leq Y$ in $\mathcal{P}(I)$ 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.

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.

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 (nonstrict) order relation on it as follows:

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

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.

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):

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

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 $\leq$ defined by

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

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

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

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.

Set Theory. Most versions of the 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 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. The Scheme of Replacement generalizes the Scheme of Comprehension by guaranteeing even

the sethood of certain 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 $x \in X$.

Axiomatization of First Order Logic and the Soundness Theorem

Similarly to 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*.

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. 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".

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.

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)

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

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 $(\forall x)\varphi$)

Exercise 1. (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 $\varphi_0, \varphi_1, \dots, \varphi_n$ of formulas such that every 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 $\varphi_0, \varphi_1, \dots, \varphi_n$ 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.

Exercise 2. 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) $(x_1 = y_1 \wedge \dots \wedge x_n = y_n) \Rightarrow t(x_1, \dots, x_n) = t(y_1, \dots, y_n)$ (for any term t)
(d) $(x_1 = y_1 \wedge \dots \wedge x_n = y_n) \Rightarrow (\varphi(x_1, \dots, x_n) \Leftrightarrow \varphi(y_1, \dots, y_n))$
(for any formula φ such that all the substitutions y_i/x_i, in φ are admissible)

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 $(\forall x)\varphi_i$. 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 1.)

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

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

At this moment, the reader should return to the remarks following the demonstration of the Soundness Theorem and the first formulation of the Completeness Theorem in Propositional Calculus and realize that they equally apply to their first order versions.

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.

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 $T \vDash \mathcal{A}$ 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 $T \nvdash_{\text{fin}} \varphi$ for any sound provability relation $\vdash_{\text{fin}}$.

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.

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 $(\mathrm{Q}_1\, x_1)\ldots(\mathrm{Q}_n\, x_n)\psi$ where $\mathrm{Q}_1, \ldots, \mathrm{Q}_n$ 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:

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

In each case decide which of the implications "left $\Rightarrow$ right", "right $\Rightarrow$ left" remain logically valid even without the assumption that x has no free occurrence in ψ, and give examples that in the remaining cases this assumption cannot be omitted.

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

(d) Replace any case of the logical equivalences $\theta \equiv \chi$ in (a) and (b) 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).

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

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 the corresponding implication in the demonstration of the Deduction Theorem 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 proving the implication $\varphi \Rightarrow \psi$ in T by proving 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 implication 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.

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 $x = z$. 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 in the same way as in Propositional Calculus.

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).*

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 $T \cup \{\neg\varphi\} \vdash \psi$.*

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

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 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).

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

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.

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.

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.

Dense Linear Order. A linearly ordered set $(A; <)$ is called *dense* if it 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. Three more complete extension of the Theory of Dense Linear Order can obtained by the variation of the condition of the existence of endpoints in the obvious way.

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

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

using the previously defined symbol $\leq$. Then $\mathcal{B}$ is called *atomic* if for every nonzero element of B there is an atom contained in it, i.e., $\mathcal{B}$ 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 nontriviality 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 by a pair of integer invariants; however, this description is already beyond the scope of our exposition.

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 addition (and the

successor operation) of natural numbers. Its language contains just the constant symbols 0 and 1 and the operation symbol +; its axioms are obtained from the axioms of Peano Arithmetic 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 $\cdot$.

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.

Theorem 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, ... , 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 $T \vdash \varphi(c_1, \ldots, c_k)$. Let's realize that T is a theory in the language L not containing the constants c_1, ... , 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 every 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 several times did so in the part devoted to various examples of first order theories, without paying special attention to this point. In particular, 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 of 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 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 $L' \supseteq L$. 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 L-formulas in T.

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.

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 the Theorem 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 not contribute to the reader's understanding the issue.

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 $\mathcal{P}(X)$.

(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

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

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 $\sim_T$.

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

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

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*

$$K/\!\sim = \{\widetilde{t} : t \in K\},$$

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 in a single element all pairwise equivalent elements of K, i.e., considering the equivalence relation $\sim$ as a "new equality" on K. The last two *compatibility conditions* express the fact that both the operations and the relations in $\mathcal{K}$ preserve 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 functional 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 block $f(t_1, \dots, t_n)^{\sim} \in M$ 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 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 $\widetilde{t}_i$.

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.

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, 1·1, 0+(1·1) 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 $1 \sim_{\mathrm{PA}} t$. Similarly, the natural number 2 denotes the equivalence block of the constant term $1+1$ or of any constant term provably equivalent 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 $\mathcal{M}(T) \models T$, 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 $\mathcal{M}(T) \models T$ 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 $(\exists\, x)\varphi(x)$ 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.

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

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

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 $r(t_1, \dots, t_n)$ where t, s and $t_1, \dots, t_n$ 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 $\mathcal{M}(T) \vDash (\exists\, x)\varphi$.

Exercise. Assume that S is a contradictory theory in a first order language L. 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* (one of the higher axioms of Set Theory) it is possible to prove the following theorem. The interested reader will find its proof in the Appendix. 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.

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.

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 $\mathcal{M}(\widehat{T})$ of the theory $\widehat{T}$ is a model of $\widehat{T}$, i.e., $\mathcal{M}(\widehat{T}) \vDash \widehat{T}$. Since $T \subseteq \widehat{T}$, we have $\mathcal{M}(\widehat{T}) \vDash T$, hence $\mathcal{M}(\widehat{T})$ is a model of T, as well.

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

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. 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 the Theorem on Proof by Contradiction. Then, according to Gödel's Completeness Theorem, $T \cup \{\neg\psi\}$ has some model $\mathcal{A}$. Then $\mathcal{A} \vDash T$ 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 $T \vdash \psi$.

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.

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.

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 of 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 model $\mathcal{A} = (A; \dots)$ of the theory T such that $|A| \geq n$.

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:

$$(\exists\, x_1, \dots, x_n)\Big(\bigwedge_{1 \leq i < j \leq n} x_i \neq x_j \Big)$$

Then, for any L-structure $\mathcal{A} = (A; \dots)$, 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 $\sigma_2, \dots, \sigma_n$ 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 models. This is to say 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.

Exercise. (a) Show that if a first order theory T in the language of the Theory of Unitary Rings 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 a first order theory T in the language of the Theory of Unitary Rings 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 a sentence φ in the language of the Theory of Unitary Rings 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 non-standard models of Peano Arithmetic.

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 the extended language (recall that, for every natural number n, we denote by n also the constant term $(\dots(0+1)+\dots+1)+1$ 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 $q^{\mathcal{M}_n}$ as the natural number $n+1$, we obtain the model

$$\mathcal{M}_n = (\mathbb{N}; +, \cdot, 0, 1, n+1) \vDash \mathrm{PA} \cup S_n$$

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.

Cardinality of Models and Skolem's Paradox

A detailed inspection of the proof of Gödel's Completeness Theorem (both in Section ... as well as in the Appendix) would show 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 $\mathcal{M}(T) = (M; \dots)$ of any theory T in the language L is a quotient $M = K/\!\sim_T$, therefore $|M| \leq |K|$. Thus the canonical structure $\mathcal{M}(T) = (M; \dots)$ of any theory in a first order 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 T^+ of T, has the same cardinality $\|L\|$. Thus the new language L^+ has the same cardinality as the original language L. Putting things together, we see that the canonical structure $\mathcal{M}(T^+) = (M; \dots)$ still has the cardinality $|M| \leq \|L\|$. As a consequence, we obtain the following strengthening of Gödel's Completeness Theorem.

Theorem. *Let T be a consistent theory in a first order language L of cardinality $\|L\| = \alpha$. Then T has a model $\mathcal{M} = (M; \dots)$ of cardinality $|M| \leq \alpha$.*

Corollary 1. *Every consistent first order theory in a countable language has a countable model, i.e., a model $\mathcal{M} = (M; \dots)$ 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:

Corollary 2. *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, 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, lack an absolute character.

Appendix

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 $m < a$. A subset $C \subseteq A$ is called a *chain* if $a \leq b$ or $b \leq a$ holds for any $a, b \in C$, i.e., if C is totally ordered by the relation $\leq$.

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 $F \in \mathcal{T}$ for any finite set $F \subseteq T$. We say that a system $\mathcal{T} \subseteq \mathcal{P}(X)$ is *inductive* if for any chain $\mathcal{C} \subseteq \mathcal{T}$ also its union $\bigcup \mathcal{C}$ belongs to $\mathcal{T}$.

We record without proof the following two consequences of the Axiom of Choice with the remark that anyone of them is in fact equivalent to it.

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 $T \subseteq M$.*

Zorn-Kuratowski Lemma. *Let X be any set and $\mathcal{T} \subseteq \mathcal{P}(X)$ be an inductive system 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 first order theories in L. As already noticed in the proof of the Compactness Theorem, a theory is consistent if and only if every its finite subtheory $T_0 \subseteq T$ is consistent. In other words, the system $\mathcal{T} \subseteq \mathcal{P}(\Phi)$ of all consistent theories in L has finite character.

The other way round, given a chain $\mathcal{C}$ of consistent theories in L, it can easily be seen that its union $U = \bigcup \mathcal{C}$ is a consistent theory, again. Indeed, if U were contradictory then there would be some sentence φ and a proof $\varphi_0, \varphi_1, \ldots, \varphi_m$ of φ in U, as well as a proof $\psi_0, \psi_1, \ldots, \psi_n$ of its negation $\neg\varphi$ in U. Let $\chi_1, \ldots, \chi_k$ be all the axioms of U occurring in any of these proofs. Then there are theories $T_1, \ldots, T_k \in \mathcal{C}$ such that $\chi_i \in T_i$ for each $i = 1, \ldots, k$. Since $\mathcal{C}$ is a chain, there is some p such that $1 \leq p \leq k$ and $T_i \subseteq T_p$ for any $i = 1, \ldots, k$. Then $\{\chi_1, \ldots, \chi_k\} \subseteq T_p$, hence both $\varphi_0, \varphi_1, \ldots, \varphi_m$ and $\psi_0, \psi_1, \ldots, \psi_n$ are already proofs in T_p, so that T_p is contradictory. However, this is impossible, as $T_p \in \mathcal{C}$ and all the elements of $\mathcal{C}$ are consistent theories. Summing up, the system of all consistent theories $\mathcal{T} \subseteq \mathcal{P}(\Phi)$ is inductive.

Now, both the Teichmüller-Tukey Lemma as well as the Zorn-Kuratowski Lemma imply that every consistent theory T in a first order language L can be extended to a maximal consistent theory $M \supseteq T$ in L.

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.

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.*

Remark. Although (some equivalent alternative formulations of) the Axiom of Choice played 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.

Next we show a result on Henkin extensions of consistent theories.

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 sentence $(\exists\, x)\varphi(x)$ with a new witnessing constant d_φ and to extend 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 of constant symbols $D_n = \{d_\varphi \colon \varphi \in \Phi_n\}$, the language L_{n+1} and the set of witnessing axioms $W_n = \{(\exists\, x)\varphi(x) \Rightarrow \varphi(d_\varphi) \colon \varphi \in \Phi_n\}$ 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, 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 Completeness Theorem for Propositional Calculus, 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 propositional axioms of Predicate Calculus. Then, by Modus Ponens, we have both

$$T \cup \Theta \vdash \neg(\exists\, x)\varphi(x) \Rightarrow \psi \qquad \text{and} \qquad T \cup \Theta \vdash \varphi(d) \Rightarrow \psi$$

Since the symbol d doesn't occur in any of the specific axioms of the theory $T \cup \Theta$, applying the Theorem on Constants to the latter relation 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 on the prenex normal form,

$$T \cup \Theta \vdash (\exists\, x)\varphi(x) \Rightarrow \psi$$

Since for any propositional forms A, B the propositional form

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

is a tautology, hence provable just from the logical axioms(cf. Exercise... (g)), we have

$$T \cup \Theta \vdash ((\exists\, x)\varphi(x) \Rightarrow \psi) \Rightarrow ((\neg(\exists\, x)\varphi(x) \Rightarrow \psi) \Rightarrow \psi)$$

Applying Modus Ponens twice we get $T \cup \Theta \vdash \psi$. Since the set Θ consists of $k-1$ witnessing axioms, only, we are done.

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 simpler way than was the demonstration of conservativeness of the extension T_H.)

Combining the two recently established theorems, we can finally prove the announced result.

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.

Selected Bibliography

B. Balcar, P. Štěpánek, *Teorie množin*, Academia, Praha 2000 (2. vyd.).

L. Bukovský, *Množiny a všeličo okolo nich*, Alfa, Bratislava 1985.

C. C. Chang, H. J. Keisler, *Model Theory*, Studies in Logic and the Foundations of Mathematics, Elsevier, North Holland, Amsterdam 1990 (3rd ed.).

H.-D. Ebbinghaus, J. Flum, *Finite Model Theory*, Perspectives in Mathematical Logic, Springer, Berlin-Heidelberg-New York 1999 (2nd ed.).

H. B. Enderton, *A Mathematical Introduction to Logic*, Hartcourt/Academic Press, San Diego-New York-Boston-London-Toronto-Sydney-Tokyo 2001 (2nd ed.).

T. Franzen, *Gödel's Theorem: An Incomplete Guide to Its Use and Abuse*, A K Peters, Wellesley, Mass. 2005.

M. Hills, F. Loeser, *A First Journey through Logic*, Student Mathematical Library vol. 89, Amer. Math. Soc. 2019.

W. Hodges, *Model Theory*, Cambridge University Press, Cambridge-New York 1993.

W. Hodges, *A Shorter Model Theory*, Cambridge University Press, Cambridge-New York 1997.

J. Kolář, O. Štepánková, M. Chytil, *Logika, algebry a grafy* (Kap. 2. Výroková logika, Kap. 3. Predikátová logika), SNTL, Praha 1989

V. Kolman, V. Punčochář, *Formy jazyka: Úvod do logiky a její filosofie*, Filosofia, Praha 2015.

Yu. I. Manin, *A Course in Mathematical Logic for Mathematicians*, Graduate Texts in Mathematics, Springer, New York-Dordrecht-Heidelberg-London 2010 (2nd ed. with collaboration by B. Zilber).

D. Marker, *Model Theory, An Introduction*, Graduate Texts in Mathematics, Springer, New York-London 2002.

E. Mendelson, *Introduction to Mathematical Logic*, D. Van Nostrand, Princeton, New Jersey 1964.

G. Restal, *Logic, An Introduction*, Routlege, Taylor & Francis Group, London-New York 2005.

J. R. Shoenfield, *Mathematical Logic*, Addison-Wesley, Reading, Mass. 1967, reprinted by Assoc. Symb. Logic, Urbana, Illinois and A K Peters, Natick, Mass. 2001.

P. Smith, *An Introduction to Gödel's Theorems*, Cambridge University Press, Cambridge-New York 2013.

R. M. Smullyan, *Gödel's Incompleteness Theorems*, Oxford University Press, Oxford-New York 1992.

A. Sochor, *Klasická matematická logika*, Univerzita Karlova, Nakladatelství Karolinum, Praha 2001.

A. Sochor, *Logika pro všechny ochotné myslet*, Univerzita Karlova, Praha 2011.

P. Štěpánek, *Matematická logika*, skriptá MFF UK, SPN, Praha 1982.

V. Švejdar, *Logika – neúplnost, složitost, nutnost*, Academia, Praha 2002.

Hao Wang, *A Logical Journey: From Gödel to Philosophy*, MIT Press, Cambridge, Mass., London, England 1996.

P. Zlatoš, *Ani matematika si nemôže byť istá sama sebou: Úvahy o množinách, nekonečne, paradoxoch a Gödelových vetách*, IRIS, Bratislava 1995.

P. Zlatoš, *O dobrom usporiadaní a axióme výberu*, Obzory matematiky, fyziky a informatiky, **58** (1999), 1–9 (1. časť), **1** (2000), 14–24 (2. časť).

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 relations to our brains and 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)

Liar's Paradox and the Paradoxes of Russell and Berry

Let us recall the famous ancient *Liar's Paradox*, also known as the *Epimenidus Paradox*, usually ascribed to Eubilides of Miletus. In it the Cretan Epimenidus 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 Epimenidus 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 Epimenidus has told us a lie. This is not a contradiction, however, the fact that pronouncing a false sentence can guarantee the existence of a person not lying all the time is still fairly paradoxical. The strong version of the *Liar's 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's Paradox*.

In everyday life we need not to worry too much about the Liar's 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 "naïve" 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's 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's Paradox inside of this theory.

Liar's 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.

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 $\mathbb{N} \setminus A$ has to belong to the set A, as well. However, this is a contradiction, since $A \cap (\mathbb{N} \setminus A) = \emptyset$.

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 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 described 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.

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)

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, closely related to Liar's 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 (born 1906 in Brno) 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, 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.

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:

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:*

$P(m, n, k)$	$T \vdash P(m, n, k)$	Δ_k is a proof of the sentence $\varphi_n(m)$ in T
$R(m, n, k)$	$T \vdash R(m, n, k)$	Δ_k is a proof of the sentence $\neg\varphi_n(m)$ in T
$\neg P(m, n, k)$	$T \nvdash P(m, n, k)$	$T \vdash \neg P(m, n, k)$
$\neg R(m, n, k)$	$T \nvdash R(m, n, k)$	$T \vdash \neg R(m, n, k)$

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:

$$\varphi_g(g)\colon \quad \textit{"The sentence } \varphi_g(g) \textit{ is not provable in } T."$$

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 entitle us to state the semantic version of Gödel's First Incompleteness Theorem.

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, e.g., 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 Gödel's First 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 semantic belief is even stronger than its syntactic counterpart, namely the weaker belief in the consistency of PA, of which we still lack a direct and immediate evidence. Anyway, it will be interesting to see what we can infer from this weaker syntactic assumption.

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 two different 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... . 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... . 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.

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).

Later on B. 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.

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 $(\exists\, u \leq k)R(r,r,u)$ 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, ... , Δ_{l-1}, there is a proof of the sentence $\varphi_r(r)$ in T, contradicting the consistency of T, again.

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 there is a sentence ψ in the language L such that both ψ and $\neg\psi$ are provable in T. Accordingly, the consistency statement can be formulated in following fairly suggestive way:

$$\mathrm{Cons}_1(T)\colon \neg(\exists\, x, y, z, w)(P(x, y, z) \wedge R(x, y, w))$$

excluding the existence of any sentence of the form $\varphi_n(m)$ such that both $\varphi_n(m)$, $\neg\varphi_m(m)$ 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.

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

$$\neg(\exists\, x,y,z,w)(P(x,y,z) \wedge R(x,y,w)) \Rightarrow \neg(\exists\, z)P(g,g,z)$$

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 $R(z,y,z)$, 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:

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 the 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 Theorem on Proof by Contradiction, its extension $T \cup \{\neg\,\mathrm{Cons}(T)\}$ is consistent, as well. However, as T is consistent, the 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 is. 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 $R(x,y,z)$.

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. 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.

Exercise. (a) Show that the Rosser formula $\varphi_r(r)$ 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 S. 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).

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 $i = 1, \ldots, 3$. 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 get 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, 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 PA 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 particular proofs of the individual 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 PA based on the removal of the ω-incompleteness phenomenon was proposed by S. Feferman in 1962. However, in order 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 $\widetilde{T}$.

The Theorems of Tarski and Church-Turing

To complete the picture we formulate two further incompleteness results by A. Tarski, and A. Church and A. 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.

Tarski's Theorem on Undefinability of Truth. *Let T be any arithmetical first order 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, which is of semantic nature, 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 T it is necessary to extend T to a first order 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.

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 related 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.

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 in T of the sentence $\varphi_n(m)$ 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 relational symbol or at least two operation symbols is a "first order tautology", i.e., whether it is satisfied in all L-structures, (or, which is the same, whether is provable just from the logical axioms). Thus there is a striking difference between the First Order Logic and the Propositional Calculus in which the tautologies can be effectively recognized by the truth table algorithm.

Compared with Tarski's Theorem, Church-Turing Theorem is of syntactic character. 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.

Goodstein Sequences:

An Example of a True Arithmetical Statement Unprovable in PA

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

$$\begin{aligned} G(m,0) &= 2^4 + 2^3 + 2^2 + 1 = 2^{2^2} + 2^{2+1} + 2^2 + 1 \\ G(m,1) &= 3^{3^3} + 3^{3+1} + 3^3 + 1 - 1 = 3^{3^3} + 3^{3+1} + 3^3 = 7\,625\,597\,485\,095 \\ G(m,2) &= 4^{4^4} + 4^{4+1} + 4^4 - 1 = 4^{4^4} + 4^{4+1} + 3 \cdot 4^3 + 3 \cdot 4^2 + 3 \cdot 4 + 3 \approx 1.340 \cdot 10^{154} \\ G(m,3) &= 5^{5^5} + 5^{5+1} + 3 \cdot 5^3 + 3 \cdot 5^2 + 3 \cdot 5 + 3 - 1 \\ &= 5^{5^5} + 5^{5+1} + 3 \cdot 5^3 + 3 \cdot 5^2 + 3 \cdot 5 + 2 \sim 10^{2\,200} \\ G(m,4) &= 6^{6^6} + 6^{6+1} + 3 \cdot 6^3 + 3 \cdot 6^2 + 3 \cdot 6 + 2 - 1 \\ &= 6^{6^6} + 6^{6+1} + 3 \cdot 6^3 + 3 \cdot 6^2 + 3 \cdot 6 + 1 \sim 10^{36\,305} \\ G(m,5) &= 7^{7^7} + 7^{7+1} + 3 \cdot 7^3 + 3 \cdot 7^2 + 3 \cdot 7 + 1 - 1 \\ &= 7^{7^7} + 7^{7+1} + 3 \cdot 7^3 + 3 \cdot 7^2 + 3 \cdot 7 \sim 10^{696\,000} \end{aligned}$$

. . . . . .

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 is highly surprising and unexpected.

Goodstein's Theorem [1944]**.** *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

$$\begin{array}{lll} G(3,0) = 3 = 2 + 1 & G(3,1) = 3 + 1 - 1 = 3 & G(3,2) = 4 - 1 = 3 \\ G(3,3) = 3 - 1 = 2 & G(3,4) = 2 - 1 = 1 & G(3,5) = 0 = G(3,n) \text{ for } n > 5 \end{array}$$

For $m = 4$ the first n such that $G(4, n) = 0$ equals the immense value $3 \cdot \left(2^{402\,653\,211} - 1\right)$.

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 $G(m,n) \le \Gamma(m,n)$ 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

$$G(m,0) = 2^{2^2} + 2^{2+1} + 2^2 + 1 < \omega^{\omega^\omega} + \omega^{\omega+1} + \omega^\omega + 1 = \Gamma(m,0)$$

$$G(m,1) = 3^{3^3} + 3^{3+1} + 3^3 < \omega^{\omega^\omega} + \omega^{\omega+1} + \omega^\omega = \Gamma(m,1)$$

$$G(m,2) = 4^{4^4} + 4^{4+1} + 3{\cdot}4^3 + 3{\cdot}4^2 + 3{\cdot}4 + 3 < \omega^{\omega^\omega} + \omega^{\omega+1} + 3{\cdot}\omega^3 + 3{\cdot}\omega^2 + 3{\cdot}\omega + 3 = \Gamma(m,2)$$

$$G(m,3) = 5^{5^5} + 5^{5+1} + 3{\cdot}5^3 + 3{\cdot}5^2 + 3{\cdot}5 + 2 < \omega^{\omega^\omega} + \omega^{\omega+1} + 3{\cdot}\omega^3 + 3{\cdot}\omega^2 + 3{\cdot}\omega + 2 = \Gamma(m,3)$$

$$G(m,4) = 6^{6^6} + 6^{6+1} + 3{\cdot}6^3 + 3{\cdot}6^2 + 3{\cdot}6 + 1 < \omega^{\omega^\omega} + \omega^{\omega+1} + 3{\cdot}\omega^3 + 3{\cdot}\omega^2 + 3{\cdot}\omega + 1 = \Gamma(m,4)$$

$$G(m,5) = 7^{7^7} + 7^{7+1} + 3{\cdot}7^3 + 3{\cdot}7^2 + 3{\cdot}7 < \omega^{\omega^\omega} + \omega^{\omega+1} + 3{\cdot}\omega^3 + 3{\cdot}\omega^2 + 3{\cdot}\omega = \Gamma(m,5)$$

$\ldots\ldots$

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 J. Paris and L. Kirby, Goodstein's Theorem cannot be proved just by means of the Peano Arithmetic alone.

Paris-Kirby Theorem [1982]. *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, for any fixed $m \in \mathbb{N}$, the existential statement $(\exists\, y)(G(m, y) = 0)$ is provable in PA. 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 within PA it is impossible to extract any general common idea out of those particular proofs and convert them into a single proof of the universal-existential sentence $(\forall\, x)(\exists\, y)(G(x, y) = 0)$.

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.[1]

[1] Freely quoted according to Hao Wang [9].

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?

Suggestions for Further Reading

[1] Martin Davis, *Pragmatic Platonism, Mathematics and the Infinite*; https://www.researchgate.net/publication/329449494

[2] Martin Davis, *What Did Gödel Believe and When Did He Believe It?* Bull. Symbolic Logic **11** (2005), 194-206; https://www.jstor.org/stable/1556749?seq=1

[3] Solomon Feferman, *The Impact of the Incompleteness Theorems on Mathematics*, Notices Amer. Math. Soc. **53** (April 2006), 434–439; https://www.ams.org/notices/200604/fea-feferman.pdf

[4] Torkel Franzén, *Gödel's Theorem: An Incomplete Guide to its Use and Abuse*, A. K. Peters, Wellesley, 2005.

[5] Haim Gaifman, *What Gödel's Incompleteness Result Does and Does Not Show*; https://pdfs.semanticscholar.org/1efb/5896951d8c20984f558c1e4ab4d6d029143d.pdf

[6] Panu Raatikainen, *Gödel's Incompleteness Theorems*, Stanford Encyclopedia of Philosophy; https://plato.stanford.edu/entries/goedel-incompleteness/

[7] Peter Smith, *An Introduction to Gödel's Theorems*, Cambridge University Press, Cambridge, 2007.

[8] Raymond Smullyan, *Gödel's Incompleteness Theorems*, Oxford University Press, Oxford, 1992.

[9] Hao Wang, *A Logical Journey, From Gödel to Philosophy*, MIT Press, Cambridge, Massachusetts–London, England, 1996.

[10] Richard Zach, *Hilbert's Program*, Stanford Encyclopedia of Philosophy; http://plato.stanford.edu/archives/fall2003/entries/hilbert-program/(2003)

[11] Pavol Zlatoš, *Ani matematika si nemôže byť istá sama sebou, Úvahy o množinách, nekonečne, paradoxoch a Gödelových vetách*, IRIS, Bratislava, 1995; http://thales.doa.fmph.uniba.sk/zlatos/animat/animat.pdf