numbers is empty, and hence cannot be a member of itself (otherwise, it would not be empty). Now, call a set x *good* if x is not a member of itself and let C be the collection of all sets which are good. Is C, as a set, good or not? If C is good, then C is *not* a member of itself, but since C contains all sets which are good, C *is* a member of C, a contradiction. Otherwise, if C is a member of itself, then C must be good, again a contradiction. In order to avoid this paradox we have to exclude the collection C from being a set, but then, we have to give reasons why certain collections are sets and others are not. The axiomatic way to do this is described by Zermelo as follows: *Starting with the historically grown Set Theory, one has to search for the principles required for the foundations of this mathematical discipline. In solving the problem we must, on the one hand, restrict these principles sufficiently to exclude all contradictions and, on the other hand, take them sufficiently wide to retain all the features of this theory*.

The principles, which are called axioms, will tell us how to get new sets from already existing ones. In fact, most of the axioms of Set Theory are constructive to some extent, *i.e.*, they tell us how new sets are *constructed* from already existing ones and what elements they contain.

However, before we state the axioms of Set Theory we would like to introduce informally the formal language in which these axioms will be formulated.

First-Order Logic in a Nutshell

First-Order Logic is the system of Symbolic Logic concerned not only to represent the logical relations between sentences or propositions as wholes (like *Propositional Logic*), but also to consider their internal structure in terms of subject and predicate. First-Order Logic can be consider as a kind of language which is distinguished from higher-order languages in that it does not allow quantification over subsets of the domain of discourse or other objects of higher type. Nevertheless, First-Order Logic is strong enough to formalise all of Set Theory and thereby virtually all of Mathematics. In other words, First-Order Logic is an abstract language that in one particular case is the language of Group Theory, and in another case is the language of Set Theory.

The goal of this brief introduction to First-Order Logic is to illustrate and summarise some of the basic concepts of this language and to show how it is applied to fields like Group Theory and Peano Arithmetic (two theories which will accompany us for a while).

Syntax: Formulae, Formal Proofs, and Consistency

Like any other written language, First-Order Logic is based on an *alphabet*, which consists of the following *symbols*:

(a) **Variables** such as $v_0, v_1, x, y, \dots$, which are place holders for objects of the *domain* under consideration (which can for example be the elements of a group, natural numbers, or sets).

(b) **Logical operators** which are "$\neg$" (*not*), "$\wedge$" (*and*), "$\vee$" (*or*), "$\rightarrow$" (*implies*), and "$\leftrightarrow$" (*if and only if*, abbreviated *iff*).

(c) **Logical quantifiers** which are the *existential quantifier* "$\exists$" (*there is* or *there exists*) and the *universal quantifier* "$\forall$" (*for all* or *for each*), where quantification is restricted to objects only and not to formulae or sets of objects (but the objects themselves may be sets).

(d) **Equality symbol** "$=$", which stands for the particular binary *equality relation*.

(e) **Constant symbols** like the number $\mathtt{0}$ in Peano Arithmetic, or the neutral element $\mathtt{e}$ in Group Theory. Constant symbols stand for fixed individual objects in the domain.

(f) **Function symbols** such as $\circ$ (the operation in Group Theory), or $+$, $\cdot$, $\mathtt{s}$ (the operations in Peano Arithmetic). Function symbols stand for fixed functions taking objects as arguments and returning objects as values. With each function symbol we associate a positive natural number, its co-called "arity" (*e.g.*, "$\circ$" is a 2-ary or binary function, and the successor operation "$\mathtt{s}$" is a 1-ary or unary function).

(g) **Relation symbols** or **predicate constants** (such as $\in$ in Set Theory) stand for fixed relations between (or properties of) objects in the domain. Again we associate an "arity" with each relation symbol (*e.g.*, "$\in$" is a binary relation).

The symbols in (a)–(d) form the core of the alphabet and are called **logical symbols**. The symbols in (e)–(g) depend on the specific topic we are investigating and are called **non-logical symbols**. The set of non-logical symbols which are used in order to formalise a certain mathematical theory is called the **language** of this theory, denoted by $\mathscr{L}$, and *formulae* which are formulated in a language $\mathscr{L}$ are usually called $\mathscr{L}$-formulae. For example if we investigate groups, then the only non-logical symbols we use are "$\mathtt{e}$" and "$\circ$", thus, $\mathscr{L} = \{\mathtt{e}, \circ\}$ is the language of Group Theory.

A first step towards a proper language is to build words (*i.e.*, *terms*) with these symbols.

Terms:

(T1) Each variable is a term.

(T2) Each constant symbol is a term.

(T3) If $t_1, \ldots, t_n$ are terms and F is an n-ary function symbol, then $Ft_1 \cdots t_n$ is a term.

It is convenient to use auxiliary symbols like brackets in order to make terms, relations, and other expressions easier to read. For example we usually write $F(t_1, \ldots, t_n)$ rather than $Ft_1 \cdots t_n$.

To some extent, terms correspond to words, since they denote objects of the domain under consideration. Like real words, they are not statements and cannot express or describe possible relations between objects. So, the next step is to build sentences (*i.e.*, *formulae*) with these terms.

Formulae:

(F1) If t_1 and t_2 are terms, then $t_1 = t_2$ is a formula.
(F2) If $t_1, \dots, t_n$ are terms and R is an n-ary relation symbol, then $Rt_1 \cdots t_n$ is a formula.
(F3) If φ is a formula, then $\neg\varphi$ is a formula.
(F4) If φ and ψ are formulae, then $(\varphi \wedge \psi)$, $(\varphi \vee \psi)$, $(\varphi \rightarrow \psi)$, and $(\varphi \leftrightarrow \psi)$ are formulae. (To avoid the use of brackets one could write these formulae for example in *Polish notation*, *i.e.*, $\wedge\varphi\psi$, $\vee\varphi\psi$, *et cetera*.)
(F5) If φ is a formula and x a variable, then $\exists x\varphi$ and $\forall x\varphi$ are formulae.

Formulae of the form (F1) or (F2) are the most basic expressions we have, and since every formula is a logical connection or a quantification of these formulae, they are called **atomic formulae**.

For binary relations R it is convenient to write xRy instead of $R(x, y)$. For example we write $x \in y$ instead of $\in(x, y)$, and we write $x \notin y$ rather than $\neg(x \in y)$.

If a formula φ is of the form $\exists x\psi$ or of the form $\forall x\psi$ (for some formula ψ) and x occurs in ψ, then we say that x is in the *range* of a logical quantifier. A variable x occurring at a particular place in a formula φ is either in the range of a logical quantifier or it is not in the range of any logical quantifier. In the former case this particular instance of the variable x is **bound** in φ, and in the latter case it is **free** in φ. Notice that it is possible that a certain variable occurs in a given formula bound as well as free (*e.g.*, in $\exists z(x = z) \wedge \forall x(x = y)$, the variable x is both bound and free, whereas z is just bound and y is just free). However, one can always rename the bound variables occurring in a given formula φ such that each variable in φ is either bound or free. For formulae φ, the set of variables occurring free in φ is denoted by $\text{free}(\varphi)$. A formula φ is a **sentence** if it contains no free variables (*i.e.*, $\text{free}(\varphi) = \emptyset$). For example $\forall x(x = x)$ is a sentence but $(x = x)$ is not.

Sometimes it is useful to indicate explicitly which variables occur free in a given formula φ, and for this we usually write $\varphi(x_1, \dots, x_n)$ to indicate that $\{x_1, \dots, x_n\} \subseteq \text{free}(\varphi)$.

If $\varphi(x)$ is a formula (*i.e.*, $x \in \text{free}(\varphi)$), and t a term, then $\varphi(x/t)$ is the formula we get after replacing all *free* instances of x by t. A so-called **substitution** $\varphi(x/t)$ is **admissible** *iff* no free occurrence of x in φ is in the range of a quantifier that binds any variable contained in t (*i.e.*, for each variable v appearing in t, no place where x occurs free in φ is in the range of "$\exists v$" or "$\forall v$").

So far we have letters, and we can build words and sentences. However, these sentences are just strings of symbols without any inherent meaning. Later we shall interpret formulae in the intuitively natural way by giving the symbols the intended meaning (*e.g.*, "$\wedge$" meaning "and", "$\forall x$" meaning "for all x", *et cetera*). But before we shall do so, let us stay a little bit longer on the syntactical side—nevertheless, one should consider the formulae also from a semantical point of view.

Below we shall label certain formulae or types of formula as **axioms**, which are used in connection with *inference rules* in order to derive further formulae. From a semantical point of view we can think of axioms as "true" statements from which we deduce or prove further results. We distinguish two types of axiom, namely *logical axioms* and *non-logical axioms* (which will be discussed later). A **logical axiom** is a sentence or formula φ which is universally valid (*i.e.*, φ is true in any possible universe, no matter how the variables, constants, *et cetera*, occurring in φ are interpreted). Usually one takes as logical axioms some minimal set of formulae that is sufficient for deriving all universally valid formulae (such a set is given below).

If a symbol is involved in an axiom which stands for an arbitrary relation, function, or even for a first-order formula, then we usually consider the statement as an **axiom schema** rather than a single axiom, since each instance of the symbol represents a single axiom. The following list of axiom schemata is a system of logical axioms.

Let φ, φ_1, φ_2, and ψ, be arbitrary first-order formulae:

L_1: $\varphi \to (\psi \to \varphi)$,
L_2: $(\psi \to (\varphi_1 \to \varphi_2)) \to ((\psi \to \varphi_1) \to (\psi \to \varphi_2))$,
L_3: $(\varphi \wedge \psi) \to \varphi$,
L_4: $(\varphi \wedge \psi) \to \psi$,
L_5: $\varphi \to (\psi \to (\psi \wedge \varphi))$,
L_6: $\varphi \to (\varphi \vee \psi)$,
L_7: $\psi \to (\varphi \vee \psi)$,
L_8: $(\varphi_1 \to \varphi_3) \to ((\varphi_2 \to \varphi_3) \to ((\varphi_1 \vee \varphi_2) \to \varphi_3))$,
L_9: $(\varphi \to \psi) \to ((\varphi \to \neg\psi) \to \neg\varphi)$,
L_{10}: $\neg\varphi \to (\varphi \to \psi)$,
L_{11}: $\varphi \vee \neg\varphi$.

If t is a term and the substitution $\varphi(x/t)$ is admissible, then:

L_{12}: $\forall x \varphi(x) \to \varphi(t)$,
L_{13}: $\varphi(t) \to \exists x \varphi(x)$.

If ψ is a formula such that $x \notin \mathrm{free}(\psi)$, then:

L_{14}: $\forall x(\psi \to \varphi(x)) \to (\psi \to \forall x \varphi(x))$,
L_{15}: $\forall x(\varphi(x) \to \psi) \to (\exists x \varphi(x) \to \psi)$.

What is not covered yet is the symbol "$=$", so, let us have a closer look at the binary equality relation. The defining properties of equality can already be found in Book VII, Chapter 1 of Aristotle's *Topics* [2], where one of the rules to decide whether two things are the same is as follows: ... *you should look at every possible predicate of each of the two terms and at the things of which they are predicated and see whether there is any discrepancy anywhere. For anything which is predicated of the one ought also to be predicated of the other, and of anything of which the one is a predicate the other also ought to be a predicate*.

In our formal system, the binary equality relation is defined by the following three axioms.

If $t, t_1, \ldots, t_n, t'_1, \ldots, t'_n$ are any terms, R an n-ary relation symbol (*e.g.*, the binary relation symbol "$=$"), and F an n-ary function symbol, then:

L_{16}: $t = t$,
L_{17}: $(t_1 = t'_1 \wedge \ldots \wedge t_n = t'_n) \to (R(t_1, \ldots, t_n) \to R(t'_1, \ldots, t'_n))$,
L_{18}: $(t_1 = t'_1 \wedge \ldots \wedge t_n = t'_n) \to (F(t_1, \ldots, t_n) = F(t'_1, \ldots, t'_n))$.

Finally, we define the logical operator "$\leftrightarrow$" by stipulating

$$\varphi \leftrightarrow \psi \quad \iff \quad (\varphi \to \psi) \land (\psi \to \varphi),$$

i.e., $\varphi \leftrightarrow \psi$ is just an abbreviation for $(\varphi \to \psi) \land (\psi \to \varphi)$.

This completes the list of our logical axioms. In addition to these axioms, we are allowed to state arbitrarily many theory-specific assumptions, so-called **non-logical axioms**. Such axioms are for example the three axioms of *Group Theory*, denoted GT, or the axioms of *Peano Arithmetic*, denoted PA.

GT: The language of Group Theory is $\mathscr{L}_{\mathsf{GT}} = \{\mathtt{e}, \circ\}$, where "$\mathtt{e}$" is a constant symbol and "$\circ$" is a binary function symbol.

GT_0: $\forall x \forall y \forall z (x \circ (y \circ z) = (x \circ y) \circ z)$ (*i.e.*, "$\circ$" is *associative*),
GT_1: $\forall x (\mathtt{e} \circ x = x)$ (*i.e.*, "$\mathtt{e}$" is a *left-neutral* element),
GT_2: $\forall x \exists y (y \circ x = \mathtt{e})$ (*i.e.*, every element has a *left-inverse*).

PA: The language of Peano Arithmetic is $\mathscr{L}_{\mathsf{PA}} = \{\mathtt{0}, \mathtt{s}, +, \cdot\}$, where "$\mathtt{0}$" is a constant symbol, "$\mathtt{s}$" is a unary function symbol, and "$+$" and "$\cdot$" are binary function symbols.

PA_1: $\forall x (\mathtt{s}(x) \neq \mathtt{0})$,
PA_2: $\forall x \forall y (\mathtt{s}(x) = \mathtt{s}(y) \to x = y)$,
PA_3: $\forall x (x + \mathtt{0} = x)$,
PA_4: $\forall x \forall y (x + \mathtt{s}(y) = \mathtt{s}(x + y))$,
PA_5: $\forall x (x \cdot \mathtt{0} = \mathtt{0})$,
PA_6: $\forall x \forall y (x \cdot \mathtt{s}(y) = (x \cdot y) + x)$.

If φ is any $\mathscr{L}_{\mathsf{PA}}$-formula with $x \in \mathrm{free}(\varphi)$, then:

PA_7: $(\varphi(\mathtt{0}) \land \forall x (\varphi(x) \to \varphi(\mathtt{s}(x)))) \to \forall x \varphi(x)$.

It is often convenient to add certain *defined symbols* to a given language so that the expressions get shorter or at least are easier to read. For example in Peano Arithmetic—which is an axiomatic system for the natural numbers—we usually replace the expression $\mathtt{s}(\mathtt{0})$ with $\mathtt{1}$ and consequently $\mathtt{s}(x)$ by $x + \mathtt{1}$. Probably, we would like to introduce an ordering "$<$" on the natural numbers. We can do this by stipulating

$$\mathtt{1} := \mathtt{s}(\mathtt{0}), \quad x < y \quad \iff \quad \exists z ((x + z) + \mathtt{1} = y).$$

We usually use "$:=$" to define constants or functions, and "$\iff$" to define relations. Obviously, all that can be expressed in the language $\mathscr{L}_{\mathsf{PA}} \cup \{\mathtt{1}, <\}$ can also be expressed in $\mathscr{L}_{\mathsf{PA}}$.

So far we have a set of logical and non-logical axioms in a certain language and can define, if we wish, as many new constants, functions, and relations as we like. However, we are still not able to deduce anything from the given axioms, since we have neither *inference rules* nor the notion of *formal proof*.

Surprisingly, just two **inference rules** are sufficient, namely:

$$\text{Modus Ponens:}\quad \frac{\varphi \to \psi,\ \varphi}{\psi} \quad \text{and} \quad \text{Generalisation:}\quad \frac{\varphi}{\forall x \varphi}.$$

In the former case we say that ψ is obtained from $\varphi \to \psi$ and φ by Modus Ponens, and in the latter case we say that $\forall x \varphi$ (where x can be any variable) is obtained from φ by Generalisation.

Using these two inference rules, we are able to define the notion of **formal proof**: Let T be a possibly empty set of non-logical axioms (usually sentences), formulated in a certain language $\mathscr{L}$. An $\mathscr{L}$-formula ψ is **provable** from T (or provable in T), denoted $\mathsf{T} \vdash \psi$, if there is a *finite* sequence $\varphi_1, \dots, \varphi_n$ of $\mathscr{L}$-formulae such that φ_n is equal to ψ (*i.e.*, the formulae φ_n and ψ are identical), and for all i with $1 \le i \le n$ we have:

- φ_i is a logical axiom, or
- $\varphi_i \in \mathsf{T}$, or
- there are $j, k < i$ such that φ_j is equal to the formula $\varphi_k \to \varphi_i$, or
- there is a $j < i$ such that φ_i is equal to the formula $\forall x\, \varphi_j$.

If a formula ψ is not provable in T, *i.e.*, if there is no formal proof for ψ which uses just formulae from T, then we write $\mathsf{T} \nvdash \psi$.

Formal proofs, even of very simple statements, can get quite long and tricky. So, before we give an example of a formal proof, let us state a theorem which allows us to simplify formal proofs:

THEOREM 3.1 (DEDUCTION THEOREM). *If* $\{\psi_1, \dots, \psi_n\} \cup \{\varphi_1, \dots, \varphi_k\} \vdash \varphi$, *where* Generalisation *is not applied to the free variables of the formulae* $\varphi_1, \dots, \varphi_k$ *(e.g., if these formulae are sentences), then*

$$\{\psi_1, \dots, \psi_n\} \vdash (\varphi_1 \wedge \dots \wedge \varphi_k) \to \varphi.$$

Now, as an example of a formal proof let us show the equality relation is symmetric. We first work with $\mathsf{T}_{x=y}$, consisting only of the formula $x = y$, and show that $\mathsf{T}_{x=y} \vdash y = x$, in other words we show that $\{x = y\} \vdash y = x$:

φ_1:	$(x = y \wedge x = x) \to (x = x \to y = x)$	instance of L_{17}
φ_2:	$(x = y \wedge x = x) \to x = x$	instance of L_4
φ_3:	$\varphi_1 \to (\varphi_2 \to ((x = y \wedge x = x) \to y = x))$	instance of L_2
φ_4:	$\varphi_2 \to ((x = y \wedge x = x) \to y = x)$	from φ_3 and φ_1 by Modus Ponens
φ_5:	$(x = y \wedge x = x) \to y = x$	from φ_4 and φ_2 by Modus Ponens
φ_6:	$x = x$	instance of L_{16}
φ_7:	$x = y$	$(x = y) \in \mathsf{T}_{x=y}$
φ_8:	$x = x \to (x = y \to (x = y \wedge x = x))$	instance of L_5
φ_9:	$x = y \to (x = y \wedge x = x)$	from φ_8 and φ_6 by Modus Ponens
φ_{10}:	$x = y \wedge x = x$	from φ_9 and φ_7 by Modus Ponens
φ_{11}:	$y = x$	from φ_5 and φ_9 by Modus Ponens

Thus, we have $\{x = y\} \vdash y = x$, and by the Deduction Theorem 3.1 we see that $\vdash x = y \to y = x$, and finally, by Generalisation we get

$$\vdash \forall x \forall y (x = y \to y = x).$$

We leave it as an exercise to the reader to show that the equality relation is also transitive. Therefore, since the equality relation is by definition reflexive, it is an equivalence relation.

Furthermore, we say that two formulae φ and ψ are **equivalent**, denoted $\varphi \equiv \psi$, if $\vdash \varphi \leftrightarrow \psi$. In other words, if $\varphi \equiv \psi$, then—from a logical point of view—φ and ψ state exactly the same, and therefore we could call $\varphi \leftrightarrow \psi$ a tautology, which means *saying the same thing twice*. However, in Logic, a formula φ is a **tautology** if $\vdash \varphi$. Thus, the formulae φ and ψ are equivalent if and only if $\varphi \leftrightarrow \psi$ is a tautology.

A few examples:

- $\varphi \vee \psi \equiv \psi \vee \varphi$, $\varphi \wedge \psi \equiv \psi \wedge \varphi$, which shows that "$\vee$" and "$\wedge$" are commutative (up to equivalence). Moreover, "$\vee$" and "$\wedge$" are (up to equivalence) also associative—a fact which we tacitly used already.
- $\neg\neg\varphi \equiv \varphi$, $(\varphi \vee \psi) \equiv \neg(\neg\varphi \wedge \neg\psi)$, which shows for example how "$\vee$" can be replaced by "$\neg$" and "$\wedge$".
- $(\varphi \to \psi) \equiv (\neg\varphi \vee \psi)$, which shows how the logical operator "$\to$" can be replaced by "$\neg$" and "$\vee$".
- $\forall x \varphi \equiv \neg\exists x \neg\varphi$, which shows how "$\forall$" can be replaced by "$\neg$" and "$\exists$".

Thus, some of the logical operators are redundant and we could work for example with just "$\neg$", "$\wedge$", and "$\exists$". However, it is more convenient to use all of them.

Let T be a set of $\mathscr{L}$-formulae. We say that T is **consistent**, denoted $\mathrm{Con}(\mathsf{T})$, if there is *no* $\mathscr{L}$-formula φ such that $\mathsf{T} \vdash (\varphi \wedge \neg\varphi)$, otherwise T is called **inconsistent**, denoted $\neg\,\mathrm{Con}(\mathsf{T})$.

PROPOSITION 3.2. *Let T be a set of $\mathscr{L}$-formulae.*

(a) *If $\neg\,\mathrm{Con}(\mathsf{T})$, then for all $\mathscr{L}$-formulae ψ we have $\mathsf{T} \vdash \psi$.*
(b) *If $\mathrm{Con}(\mathsf{T})$ and $\mathsf{T} \vdash \varphi$ for some $\mathscr{L}$-formula φ, then $\mathsf{T} \nvdash \neg\varphi$.*

Proof. (a) Let ψ be any $\mathscr{L}$-formula and assume that $\mathsf{T} \vdash (\varphi \wedge \neg\varphi)$ for some $\mathscr{L}$-formula φ. Then $\mathsf{T} \vdash \psi$:

φ_1:	$\varphi \wedge \neg\varphi$	provable from T by assumption
φ_2:	$(\varphi \wedge \neg\varphi) \to \varphi$	instance of L_3
φ_3:	φ	from φ_2 and φ_1 by Modus Ponens
φ_4:	$(\varphi \wedge \neg\varphi) \to \neg\varphi$	instance of L_4
φ_5:	$\neg\varphi$	from φ_4 and φ_1 by Modus Ponens
φ_6:	$\neg\varphi \to (\varphi \to \psi)$	instance of L_{10}
φ_7:	$\varphi \to \psi$	from φ_6 and φ_5 by Modus Ponens
φ_8:	ψ	from φ_7 and φ_3 by Modus Ponens

(b) Assume that $\mathsf{T} \vdash \varphi$ and $\mathsf{T} \vdash \neg\varphi$. Then $\mathsf{T} \vdash (\varphi \wedge \neg\varphi)$, *i.e.*, $\neg\,\mathrm{Con}(\mathsf{T})$:

φ_1:	φ	provable from T by assumption
φ_2:	$\neg\varphi$	provable from T by assumption
φ_3:	$\varphi \to (\neg\varphi \to (\varphi \wedge \neg\varphi))$	instance of L_5
φ_4:	$\neg\varphi \to (\varphi \wedge \neg\varphi)$	from φ_3 and φ_1 by Modus Ponens
φ_5:	$\varphi \wedge \neg\varphi$	from φ_4 and φ_2 by Modus Ponens

$\dashv$

Notice that PROPOSITION 3.2(a) implies that from an inconsistent set of axioms T one can prove everything and T would be completely useless. So, if we design a set of axioms T, we have to make sure that T is consistent. However, as we shall see later, in many cases this task is impossible.

Semantics: Models, Completeness, and Independence

Let T be any set of $\mathscr{L}$-formulae (for some language $\mathscr{L}$). There are two different ways to approach T, namely the *syntactical* and the *semantical* way. The above presented syntactical approach considers the set T just as a set of well-formed formulae—regardless of their intended sense or meaning—from which we can prove some other formulae.

On the other hand, we can consider T also from a semantical point of view by interpreting the symbols of the language $\mathscr{L}$ in a reasonable way, and then seeking for a *model* in which all formulae of T are true. To be more precise, we first have to define how models are built and what "true" means:

Let $\mathscr{L}$ be an arbitrary but fixed language. An **$\mathscr{L}$-structure** $\mathfrak{A}$ consists of a (non-empty) set or collection A, called the **domain** of $\mathfrak{A}$, together with a mapping which assigns to each constant symbol $c \in \mathscr{L}$ an element $c^{\mathfrak{A}}$ of A, to each n-ary relation symbol $R \in \mathscr{L}$ a set of n-tuples $R^{\mathfrak{A}}$ of elements of A, and to each n-ary function symbol $F \in \mathscr{L}$ a function $F^{\mathfrak{A}}$ from n-tuples of A to A. Further, the interpretation of variables is given by a so-called assignment: An **assignment** in an $\mathscr{L}$-structure $\mathfrak{A}$ is a mapping j which assigns to each variable an element of the domain A. Finally, an **$\mathscr{L}$-interpretation $\mathbf{I}$** is a pair $(\mathfrak{A}, j)$ consisting of an $\mathscr{L}$-structure $\mathfrak{A}$ and an assignment j in $\mathfrak{A}$. For a variable x, an element $a \in A$, and an assignment j in $\mathfrak{A}$ we define the assignment $j\frac{a}{x}$ by stipulating

$$j\tfrac{a}{x}(y) = \begin{cases} a & \text{if } y = x, \\ j(y) & \text{otherwise.} \end{cases}$$

Further, for an interpretation $\mathbf{I} = (\mathfrak{A}, j)$ let $\mathbf{I}\frac{a}{x} := (\mathfrak{A}, j\frac{a}{x})$.

We associate with every interpretation $\mathbf{I} = (\mathfrak{A}, j)$ and every term t an element $\mathbf{I}(t)$ from the domain A as follows:

- For a variable x let $\mathbf{I}(x) := j(x)$.
- For a constant symbol $c \in \mathscr{L}$ let $\mathbf{I}(c) := c^{\mathfrak{A}}$.

- For an n-ary function symbol $F \in \mathscr{L}$ and terms $t_1, \dots, t_n$ let

$$\mathbf{I}\big(F(t_1, \dots, t_n)\big) := F^{\mathfrak{A}}\big(\mathbf{I}(t_1), \dots, \mathbf{I}(t_n)\big).$$

Now, we are able to define precisely the notion of a formula φ being true under an interpretation $\mathbf{I} = (\mathfrak{A}, j)$, in which case we write $\mathbf{I} \vDash \varphi$ and say that φ holds in $\mathbf{I}$. The definition is by induction on the complexity of the formula φ (where it is enough to consider formulae containing—besides terms and relations—just the logical operators "$\neg$" and "$\wedge$", and the logical quantifier "$\exists$"):

- If φ is of the form $t_1 = t_2$, then

$$\mathbf{I} \vDash t_1 = t_2 \quad \Longleftrightarrow \quad \mathbf{I}(t_1) \text{ is the same element as } \mathbf{I}(t_2).$$

- If φ is of the form $R(t_1, \dots, t_n)$, then

$$\mathbf{I} \vDash R(t_1, \dots, t_n) \quad \Longleftrightarrow \quad \big(\mathbf{I}(t_1), \dots, \mathbf{I}(t_n)\big) \text{ belongs to } R^{\mathfrak{A}}.$$

- If φ is of the form $\neg\psi$, then

$$\mathbf{I} \vDash \neg\psi \quad \Longleftrightarrow \quad \text{it is not the case that } \mathbf{I} \vDash \psi.$$

- If φ is of the form $\exists x\,\psi$, then

$$\mathbf{I} \vDash \exists x\,\psi \quad \Longleftrightarrow \quad \text{there is an element } a \in A \text{ such that } \mathbf{I}\tfrac{a}{x} \vDash \psi.$$

- If φ is of the form $\psi_1 \wedge \psi_2$, then

$$\mathbf{I} \vDash \psi_1 \wedge \psi_2 \quad \Longleftrightarrow \quad \mathbf{I} \vDash \psi_1 \text{ and } \mathbf{I} \vDash \psi_2.$$

Notice that since the domain of $\mathbf{I}$ is non-empty we always have $\mathbf{I} \vDash \exists x(x = x)$.

Now, let T be an arbitrary set of $\mathscr{L}$-formulae. Then an $\mathscr{L}$-structure $\mathfrak{A}$ is a **model** of T if for every assignment j in $\mathfrak{A}$ and for each formula $\varphi \in \mathsf{T}$ we have $(\mathfrak{A}, j) \vDash \varphi$, *i.e.*, φ holds in the $\mathscr{L}$-interpretation $\mathbf{I} = (\mathfrak{A}, j)$. We usually denote models by bold letters like $\mathbf{M}$, $\mathbf{N}$, $\mathbf{V}$, *et cetera*. Instead of saying "$\mathbf{M}$ is a model of T" we just write $\mathbf{M} \vDash \mathsf{T}$. If φ fails in $\mathbf{M}$, then we write $\mathbf{M} \nvDash \varphi$, which is equivalent to $\mathbf{M} \vDash \neg\varphi$ (this is because for any $\mathscr{L}$-formula φ we have *either* $\mathbf{M} \vDash \varphi$ *or* $\mathbf{M} \vDash \neg\varphi$).

For example S_7 (*i.e.*, the set of all permutations of seven different items) is a model of GT, where the interpretation of the binary operation is composition and the neutral element is interpreted as the identity permutation. In this case, the elements of the domain of S_7 can be real and can even be heard, namely when the seven items are seven bells and a peal of for example Stedman Triples consisting of all 5040 permutations of the seven bells is rung—which happens quite often, since Stedman Triples are very popular with change-ringers. However, the objects of models of mathematical theories usually do not belong to our physical world and are not more real than for example the *number zero* or the *empty set*.

The following two theorems, which we state without proofs, are the main connections between the syntactical and the semantical approach to first-order theories. On the one hand, the SOUNDNESS THEOREM 3.3 just tells us that our deduction system is sound, *i.e.*, if a sentence φ is provable from T then φ is true in each model of T. On the other hand, GÖDEL'S COMPLETENESS THEOREM 3.4 tells us that our deduction system is even complete, *i.e.*, every sentence which is true in all models of T is provable from T. As a consequence we find that $\mathsf{T} \vdash \varphi$ if and only if φ is true in each model of T. In particular, if T is empty, this implies that every tautology (*i.e.*, universally valid formula) is provable.

THEOREM 3.3 (SOUNDNESS THEOREM). *Let T be a set of $\mathscr{L}$-sentences and let φ be any $\mathscr{L}$-sentence. If $\mathsf{T} \vdash \varphi$, then in any model $\mathbf{M}$ such that $\mathbf{M} \vDash \mathsf{T}$ we have $\mathbf{M} \vDash \varphi$.*

THEOREM 3.4 (GÖDEL'S COMPLETENESS THEOREM). *Let T be a set of $\mathscr{L}$-sentences and let φ be any $\mathscr{L}$-sentence. Then $\mathsf{T} \vdash \varphi$ or there is a model $\mathbf{M}$ such that $\mathbf{M} \vDash \mathsf{T} \cup \{\neg\varphi\}$. In other words, if for every model $\mathbf{M} \vDash \mathsf{T}$ we have $\mathbf{M} \vDash \varphi$, then $\mathsf{T} \vdash \varphi$. (Notice that this does not imply the existence of a model of T.)*

One of the main consequences of GÖDEL'S COMPLETENESS THEOREM 3.4 is that formal proofs—which are usually quite long and involved—can be replaced by informal ones: Let T be a consistent set of $\mathscr{L}$-formulae and let φ be any $\mathscr{L}$-sentence. Then, by GÖDEL'S COMPLETENESS THEOREM 3.4, in order to show that $\mathsf{T} \vdash \varphi$ it is enough to show that $\mathbf{M} \vDash \varphi$ whenever $\mathbf{M} \vDash \mathsf{T}$. In fact, we would take an arbitrary model $\mathbf{M}$ of T and show that $\mathbf{M} \vDash \varphi$.

As an example let us show that $\mathsf{GT} \vdash (y \circ x = \mathrm{e}) \to (x \circ y = \mathrm{e})$: Firstly, let $\mathbf{G}$ be a model of GT, with domain G, and let x and y be any elements of G. By GT_2 we know that every element of G has a left-inverse. In particular, y has a left-inverse, say $\bar{y}$, and we have $\bar{y} \circ y = \mathrm{e}$. By GT_1 we have $x \circ y = (\bar{y} \circ y) \circ (x \circ y)$, and by GT_0 we get $(\bar{y} \circ y) \circ (x \circ y) = \bar{y} \circ ((y \circ x) \circ y)$. Now, if $y \circ x = \mathrm{e}$, then we have $x \circ y = \bar{y} \circ y$ and consequently we get $x \circ y = \mathrm{e}$. Notice that we tacitly used that the equality relation is symmetric and transitive.

We leave it as an exercise to the reader to find the corresponding formal proof of this basic result in Group Theory. In a similar way one can show that every left-neutral element is also a right-neutral element (called *neutral element*) and that there is just one neutral element in a group.

The following result, which is a consequence of GÖDEL'S COMPLETENESS THEOREM 3.4, shows that *every* consistent set of formulae has a model.

PROPOSITION 3.5. *Let T be any set of $\mathscr{L}$-formulae. Then* $\mathrm{Con}(\mathsf{T})$ *if and only if T has a model.*

Proof. ($\Rightarrow$) If T has no model, then, by GÖDEL'S COMPLETENESS THEOREM 3.4, for *every* $\mathscr{L}$-formula ψ we have $\mathsf{T} \vdash \psi$ (otherwise, there would be a model of $\mathsf{T} \cup \{\neg\psi\}$, and in particular for T). So, for ψ being $\varphi \wedge \neg\varphi$ we get $\mathsf{T} \vdash (\varphi \wedge \neg\varphi)$, hence T is inconsistent.

($\Leftarrow$) If T is inconsistent, then, by PROPOSITION 3.2(a), for every $\mathscr{L}$-formula ψ we have $\mathsf{T} \vdash \psi$, in particular, $\mathsf{T} \vdash \varphi \wedge \neg\varphi$. Now, the SOUNDNESS THEOREM 3.3 implies that in all models $\mathbf{M} \vDash \mathsf{T}$ we have $\mathbf{M} \vDash \varphi \wedge \neg\varphi$; thus, there are no models of T. $\dashv$

A set of *sentences* T is usually called a **theory**. A consistent theory T (in a certain language $\mathscr{L}$) is said to be **complete** if for every $\mathscr{L}$-sentence φ, *either* $\mathsf{T} \vdash \varphi$ *or* $\mathsf{T} \vdash \neg\varphi$. If T is not complete, we say that T is **incomplete**.

The following result is an immediate consequence of PROPOSITION 3.5.

COROLLARY 3.6. *Every consistent theory is contained in a complete theory.*

Proof. Let T be a theory in the language $\mathscr{L}$. If T is consistent, then it has a model, say $\mathbf{M}$. Now let $\bar{\mathsf{T}}$ be the set of all $\mathscr{L}$-sentences φ such that $\mathbf{M} \vDash \varphi$. Obviously, $\bar{\mathsf{T}}$ is a complete theory which contains T. $\dashv$

Let T be a set of $\mathscr{L}$-formulae and let φ be any $\mathscr{L}$-formula not contained in T. φ is said to be **consistent relative to** T (or that φ is **consistent with** T) if $\mathrm{Con}(\mathsf{T})$ implies $\mathrm{Con}(\mathsf{T} \cup \{\varphi\})$ (later we usually write $\mathsf{T} + \varphi$ instead of $\mathsf{T} \cup \{\varphi\}$). If both φ and $\neg\varphi$ are consistent with T, then φ is said to be **independent** of T. In other words, if $\mathrm{Con}(\mathsf{T})$, then φ is independent of T if *neither* $\mathsf{T} \vdash \varphi$ *nor* $\mathsf{T} \vdash \neg\varphi$. By GÖDEL'S COMPLETENESS THEOREM 3.4 we see that if $\mathrm{Con}(\mathsf{T})$ and φ is independent of T, then there are models $\mathbf{M}_1$ and $\mathbf{M}_2$ of T such that $\mathbf{M}_1 \vDash \varphi$ and $\mathbf{M}_2 \vDash \neg\varphi$. A typical example of a statement which is independent of GT is $\forall x \forall y (x \circ y = y \circ x)$ (*i.e.*, the binary operation is commutative), and indeed, there are abelian as well as non-abelian groups.

In order to prove that a certain statement φ is independent of a given (consistent) theory T, one could try to find two different models of T such that φ holds in one model and fails in the other. However, this task is quite difficult, in particular if one cannot prove that T has a model at all (as it happens for Set Theory).

Limits of First-Order Logic

We begin this section with a useful result, called COMPACTNESS THEOREM. On the one hand, it is just a consequence of the fact that formal proofs are finite (*i.e.*, finite sequences of formulae). On the other hand, the COMPACTNESS THEOREM is the main tool to prove that a certain sentence (or a set of sentences) is consistent with a given theory. In particular, the COMPACTNESS THEOREM is implicitly used in every set-theoretic consistency proof which is obtained by forcing (for details see Chapter 16).

THEOREM 3.7 (COMPACTNESS THEOREM). *Let T be an arbitrary set of $\mathscr{L}$-formulae. Then T is consistent if and only if every finite subset Φ of T is consistent.*

Proof. Obviously, if T is consistent, then every finite subset Φ of T must be consistent. On the other hand, if T is inconsistent, then there is a formula φ such that $\mathsf{T} \vdash \varphi \wedge \neg\varphi$. In other words, there is a proof of $\varphi \wedge \neg\varphi$ from T. Now, since every proof is finite, there are only finitely many formulae of T involved in this proof, and if Φ is this finite set of formulae, then $\Phi \vdash \varphi \wedge \neg\varphi$, which shows that Φ, a finite subset of T, is inconsistent. $\dashv$

A simple application of the COMPACTNESS THEOREM 3.7 shows that if PA is consistent, then there is more than one model of PA (*i.e.*, beside the intended model of natural numbers with domain $\mathbb{N}$, there are also so-called *non-standard* models of PA with larger domains):

Firstly we extend the language $\mathscr{L}_{\mathsf{PA}} = \{\mathtt{0}, \mathtt{s}, +, \cdot\}$ by adding a new constant symbol $\mathtt{n}$. Secondly we extend PA by adding the formulae

$$\underbrace{\mathtt{n} \neq \mathtt{0}}_{\varphi_0}, \quad \underbrace{\mathtt{n} \neq \mathtt{s}(\mathtt{0})}_{\varphi_1}, \quad \underbrace{\mathtt{n} \neq \mathtt{s}\big(\mathtt{s}(\mathtt{0})\big)}_{\varphi_2}, \quad \ldots,$$

and let Ψ be the set of these formulae. Now, if PA has a model $\mathbf{N}$ with domain say $\mathbb{N}$, and Φ is any finite subset of Ψ, then, by interpreting $\mathtt{n}$ in a suitable way, $\mathbf{N}$ is also a model of $\mathsf{PA} \cup \Phi$, which implies that $\mathsf{PA} \cup \Phi$ is consistent. Thus, by the COMPACTNESS THEOREM 3.7, $\mathsf{PA} \cup \Psi$ is also consistent and therefore has a model, say $\tilde{\mathbf{N}}$. Now, $\tilde{\mathbf{N}} \vDash \mathsf{PA} \cup \Psi$, but since $\mathtt{n}$ is different from every standard natural number of the form $\mathtt{s}(\mathtt{s}(\ldots \mathtt{s}(\mathtt{0}) \ldots))$, the domain of $\tilde{\mathbf{N}}$ must be essentially different from $\mathbb{N}$ (since it contains a kind of infinite number, whereas all standard natural numbers are finite).

This example shows that we cannot axiomatise Peano Arithmetic in First-Order Logic in such a way that all the models we get have essentially the same domain $\mathbb{N}$.

By PROPOSITION 3.5 we know that a set of first-order formulae T is consistent if and only if it has a model, *i.e.*, there is a model $\mathbf{M}$ such that $\mathbf{M} \vDash \mathsf{T}$. So, in order to prove for example that the axioms of Set Theory are consistent we only have to find a single model in which all these axioms hold. However, as a consequence of the following theorems—which we state again without proof—this turns out to be impossible (at least if one restricts oneself to methods formalisable in Set Theory).

THEOREM 3.8 (GÖDEL'S INCOMPLETENESS THEOREM). *Let T be a consistent set of first-order $\mathscr{L}$-formulae which is sufficiently strong to define the concept of natural numbers and to prove certain basic arithmetical facts (e.g., PA is such a theory, but also slightly weaker theories would suffice). Then there is always an $\mathscr{L}$-sentence φ which is independent of T, i.e., neither $\mathsf{T} \vdash \varphi$ nor $\mathsf{T} \vdash \neg\varphi$ (or in other words, there are models $\mathbf{M}_1$ and $\mathbf{M}_2$ of T such that $\mathbf{M}_1 \vDash \varphi$ and $\mathbf{M}_2 \vDash \neg\varphi$).*

In particular we find that there are number-theoretic statements which can neither be proved nor disproved in PA (*i.e.*, the theory PA is incomplete). Moreover, the following consequence of GÖDEL'S INCOMPLETENESS THEOREM 3.4 shows that not even the consistency of PA can be proved with number-theoretical methods.

THEOREM 3.9 (GÖDEL'S SECOND INCOMPLETENESS THEOREM). *Let T be a set of first-order $\mathscr{L}$-formulae. Then the statement* $\mathrm{Con}(\mathsf{T})$, *which says that $\mathsf{T} \nvdash \varphi \wedge \neg\varphi$ for some $\mathscr{L}$-formula φ, can be formulated as a number-theoretic sentence* $\mathrm{Con}^{\mathsf{T}}$. *Now, if T is consistent and is sufficiently strong to define the concept of natural numbers and to prove certain basic arithmetical facts, then $\mathsf{T} \nvdash \mathrm{Con}^{\mathsf{T}}$, i.e., T cannot prove its own consistency. In particular, $\mathsf{PA} \nvdash \mathrm{Con}^{\mathsf{PA}}$.*

On the one hand, GÖDEL'S INCOMPLETENESS THEOREM tells us that in any theory T which is sufficiently strong, there are always statements which are independent of T (*i.e.*, which can neither be proved nor disproved in T). On the other hand, statements which are independent of a given theory (*e.g.*, of Set Theory or of Peano Arithmetic) are often very interesting, since they say something unexpected, but in a language we can understand. From this point of view it is good to have Gödel's Incompleteness Theorem which guarantees the existence of such statements in theories like Set Theory or Peano Arithmetic.

In Part II we shall present a technique with which we can prove the independence of certain set-theoretical statements from the axioms of Set Theory, which are introduced and discussed below.

The Axioms of Zermelo–Fraenkel Set Theory

In 1905, Zermelo began to axiomatise Set Theory and in 1908 he published his first axiomatic system consisting of seven axioms. In 1922, Fraenkel and Skolem independently improved and extended Zermelo's original axiomatic system, and the final version was presented again by Zermelo in 1930. In this chapter we give the resulting axiomatic system called *Zermelo–Fraenkel Set Theory*, denoted ZF, which contains all axioms of Set Theory except the Axiom of Choice, which will be introduced and discussed in Chapter 5. Alongside the axioms of Set Theory we develop the theory of ordinals and give various notations which will be used throughout this book.

The language of Set Theory contains only one non-logical symbol, namely the binary **membership relation**, denoted by $\in$, and there exists just one type of object, namely sets. In other words, every object in the domain is a set and there are no other objects than sets. However, to make life easier, instead of $\in(a,b)$ we write $a \in b$ (or on rare occasions also $b \ni a$) and say that "a is an element of b", or that "a belongs to b". Later we will extend the language of Set Theory by defining some constants (like "$\emptyset$" and "ω"), relations (like "$\subseteq$"), and operations (like the power set operation "$\mathscr{P}$"), but in fact, all that can be formulated in Set Theory, can be written as a formula containing only the non-logical relation "$\in$" (but for obvious reasons, we will usually not do so).

0. The Axiom of Empty Set

$$\exists x \forall z (z \notin x).$$

This axiom not only postulates the existence of a set without any elements, *i.e.*, an empty set, it also shows that the set-theoretic universe is non-empty, because it contains at least an empty set (of course, the logical axioms L_{16} and L_{13} already incorporate this fact).

1. The Axiom of Extensionality

$$\forall x \forall y \big(\forall z (z \in x \leftrightarrow z \in y) \rightarrow x = y\big).$$