Chapter 1
Syntax: The Grammar of Symbols

The goal of this chapter is to develop the formal language of First-Order Logic from scratch. At the same time, we introduce some terminology of the so-called meta-language, which is the language we use when we speak *about* the formal language (*e.g.*, when we like to express that two strings of symbols are equal).

Alphabet

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

(a) **Variables** such as $x, y, v_0, v_1, \ldots$, 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). We use mainly lower case Latin letters (with or without subscripts) for variables.

(b) **logical operators** which are "$\neg$" (*not*), "$\wedge$" (*and*), "$\vee$" (*or*), and "$\rightarrow$" (*implies*).

(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).

More formally, to each function symbol F we adjoin a fixed FINITE string of place holders $\mathsf{x}\cdots\mathsf{x}$ and write $F\,\mathsf{x}\cdots\mathsf{x}$.

(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). More formally, to each relation symbol R we adjoin a fixed FINITE string of place holders $\mathsf{x}\cdots\mathsf{x}$ and write $R\,\mathsf{x}\cdots\mathsf{x}$.

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** (or **signature**) 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 "e" and "$\circ$", thus, $\mathscr{L} = \{\mathsf{e}, \circ\}$ is the language of Group Theory.

Terms & Formulae

With the symbols of our alphabet we can now start to compose words. In the language of First-Order Logic, these words are called called *terms*.

Terms. A string of symbols is a **term**, if it results from applying FINITELY many times the following rules:

(T0) Each variable is a term.
(T1) Each constant symbol is a term.
(T2) If $\tau_1, \ldots, \tau_n$ are any terms which we have already built and $F\,\mathsf{x}\cdots\mathsf{x}$ is an n-ary function symbol, then $F\tau_1\cdots\tau_n$ is a term (each place holder x is replaced with a term).

In order to define rule (T3) we had to use variables for terms, but since the variables of our alphabet stand just for objects of the domain and not for terms or other objects of the formal language, we had to introduce new symbols. For these new symbols, which do not belong to the alphabet of the formal language, we have chosen Greek letters. In fact, we shall mainly use Greek letters for variables which stand for objects of the formal language, also to emphasise the distinction between the formal language and the metalanguage However, we shall use the Latin letters F & R as variables for function and relation symbols respectively.

To make terms, relations, and other expressions in the formal language easier to read, it is convenient to introduce some more symbols, like brackets and commas, to our alphabet. For example we usually write $F(\tau_1, \ldots, \tau_n)$ rather than $F\tau_1\cdots\tau_n$.

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, or more precisely *formulae*, with these terms.

Formulae. A string of symbols is called a **formula**, if it results from applying F I N I T E L Y many times the following rules:

(F0) If τ_1 and τ_2 are terms, then $\tau_1 = \tau_2$ is a formula.
(F1) If $\tau_1, \ldots, \tau_n$ are any terms and $R\mathsf{x}\cdots\mathsf{x}$ is any non-logical n-ary relation symbol, then $R\tau_1 \cdots \tau_n$ is a formula.
(F2) If φ is any formula which we have already built, then $\neg\varphi$ is a formula.
(F3) If φ and ψ are formulae which we have already built, then $(\varphi \wedge \psi)$, $(\varphi \vee \psi)$, and $(\varphi \rightarrow \psi)$ 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*.)
(F4) If φ is a formula which we have already built, and ν is an arbitrary variable, then $\exists\nu\varphi$ and $\forall\nu\varphi$ are formulae.

Formulae of the form (F0) or (F1) 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 $R\mathsf{xx}$ 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. The 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 (the rules for this procedure are given later). For a formula φ, the set of variables occurring free in φ is denoted by $\text{free}(\varphi)$. A formula φ is a **sentence** (or a **closed formula**) 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.

In analogy to this definition we say that a term is a **closed term** if it contains no variables. Obviously, the only terms which are closed are the constant symbols and the function symbols followed by closed terms.

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

If φ is a formula, and τ a term, then $\varphi(x/\tau)$ is the formula we get after replacing all *free* instances of x by τ. A so-called **substitution** $\varphi(x/\tau)$ is **admissible** *iff* no free occurrence of ν in φ is in the range of a quantifier that binds any variable contained in τ (*i.e.*, for each variable ν appearing in τ, no place where ν occurs free in φ is in the range of "$\exists\nu$" or "$\forall\nu$"). For example, if $x \notin \text{free}(\varphi)$, then $\varphi(x/\tau)$ is admissible for any term τ. In this case, the formulae φ and $\varphi(x/\tau)$ are identical which we express by $\varphi \equiv \varphi(x/\tau)$. In general, we use the symbol "$\equiv$" in the metalanguage to denote equality of strings of symbols of the formal language. Furthermore, if φ is a formula and the substitution $\varphi(x/\tau)$ is admissible, then we write just $\varphi(\tau)$ instead of $\varphi(x/\tau)$. To express this we write $\varphi(\tau) :\equiv \varphi(x/\tau)$, where we use ":$\equiv$" in the metalanguage to define symbols (or strings of symbols) of the formal language.

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.

Axioms

Below we shall label certain formulae or types of formulae 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_0: $\varphi \vee \neg\varphi$,
L_1: $\varphi \rightarrow (\psi \rightarrow \varphi)$,
L_2: $(\psi \rightarrow (\varphi_1 \rightarrow \varphi_2)) \rightarrow ((\psi \rightarrow \varphi_1) \rightarrow (\psi \rightarrow \varphi_2))$,
L_3: $(\varphi \wedge \psi) \rightarrow \varphi$,
L_4: $(\varphi \wedge \psi) \rightarrow \psi$,
L_5: $\varphi \rightarrow (\psi \rightarrow (\psi \wedge \varphi))$,
L_6: $\varphi \rightarrow (\varphi \vee \psi)$,
L_7: $\psi \rightarrow (\varphi \vee \psi)$,
L_8: $(\varphi_1 \rightarrow \varphi_3) \rightarrow ((\varphi_2 \rightarrow \varphi_3) \rightarrow ((\varphi_1 \vee \varphi_2) \rightarrow \varphi_3))$,
L_9: $(\varphi \rightarrow \psi) \rightarrow ((\varphi \rightarrow \neg\psi) \rightarrow \neg\varphi)$,
L_{10}: $\neg\varphi \rightarrow (\varphi \rightarrow \psi)$.

If τ is a term, ν a variable, and the substitution $\varphi(\nu/\tau)$ is admissible, then:

L_{11}: $\forall\nu\varphi(\nu) \to \varphi(\tau)$,
L_{12}: $\varphi(\tau) \to \exists\nu\varphi(\nu)$.

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

L_{13}: $\forall\nu(\psi \to \varphi(\nu)) \to (\psi \to \forall\nu\varphi(\nu))$,
L_{14}: $\forall\nu(\varphi(\nu) \to \psi) \to (\exists\nu\varphi(\nu) \to \psi)$.

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* [1], 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 $\tau, \tau_1, \ldots, \tau_n, \tau_1', \ldots, \tau_n'$ are any terms, R an n-ary relation symbol (*e.g.*, the binary relation symbol "$=$"), and F an n-ary function symbol, then:

L_{15}: $\tau = \tau$,
L_{16}: $(\tau_1 = \tau_1' \wedge \cdots \wedge \tau_n = \tau_n') \to (R(\tau_1, \ldots, \tau_n) \to R(\tau_1', \ldots, \tau_n'))$,
L_{17}: $(\tau_1 = \tau_1' \wedge \cdots \wedge \tau_n = \tau_n') \to (F(\tau_1, \ldots, \tau_n) = F(\tau_1', \ldots, \tau_n'))$.

Finally, we define the logical operator "$\leftrightarrow$" and the binary relation symbol "$\neq$" by stipulating

$$\varphi \leftrightarrow \psi \quad :\Longleftrightarrow \quad (\varphi \to \psi) \wedge (\psi \to \varphi)$$

$$\tau \neq \tau' \quad :\Longleftrightarrow \quad \neg(\tau = \tau')$$

where we use "$:\Longleftrightarrow$" in the metalanguage to define relations between symbols (or strings of symbols) of the formal language (*i.e.*, "$\leftrightarrow$" & "$\neq$" are just abbreviations).

This completes the list of our logical axioms. In addition to these axioms, we are allowed to state arbitrarily many formulae. In logic, such a (possibly empty) set of formulae is also called a **theory**, or, when the signature $\mathscr{L}$ is specified, an **$\mathscr{L}$-theory**. Usually, a theory consists of arbitrarily many so-called **non-logical axioms** which are sentences (*i.e.*, closed formulae). Examples of theories (*i.e.*, of sets of non-logical axioms) which will be discussed in this book are the axioms of Set Theory (see Part **??**), the axioms of *Peano Arithmetic* PA (also known as *Number Theory*), and the axioms of *Group Theory* GT, which we discuss first.

GT: The language of Group Theory is $\mathscr{L}_{\mathsf{GT}} = \{\mathsf{e}, \circ\}$, where "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: $\forall x(\mathsf{e}\circ x = x)$ (*i.e.*, "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 "$+$" & "$\cdot$" are binary function symbols.

PA_0: $\neg \exists x (\mathtt{s}x = \mathtt{0})$
PA_1: $\forall x \forall y (\mathtt{s}x = \mathtt{s}y \rightarrow x = y)$,
PA_2: $\forall x (x + \mathtt{0} = x)$
PA_3: $\forall x \forall y (x + \mathtt{s}y = \mathtt{s}(x + y))$
PA_4: $\forall x (x \cdot \mathtt{0} = \mathtt{0})$
PA_5: $\forall x \forall y (x \cdot \mathtt{s}y = (x \cdot y) + x)$

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

PA_6: $\big(\varphi(\mathtt{0}) \wedge \forall x (\varphi(x) \rightarrow \varphi(\mathtt{s}(x)))\big) \rightarrow \forall x \varphi(x)$

Notice that PA_6 is an axiom schema, known as the **induction schema**, and not just a single axiom like PA_0–PA_5.

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 for example the expression $\mathtt{s0}$ with $\mathtt{1}$ and $\mathtt{ss0}$ with $\mathtt{2}$. More formally, we define

$$\mathtt{1} :\equiv \mathtt{s0} \qquad \text{and} \qquad \mathtt{2} :\equiv \mathtt{ss0}\,.$$

Obviously, all that can be expressed in the language $\mathscr{L}_{\mathsf{PA}} \cup \{\mathtt{1}, \mathtt{2}\}$ can also be expressed in $\mathscr{L}_{\mathsf{PA}}$.

Formal Proofs and Tautologies

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 until now, we do not have *inference rules* which allow us for example to infer a certain sentence from a given set of axioms.

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

MODUS PONENS (MP): $\dfrac{\varphi \rightarrow \psi,\ \varphi}{\psi}$ and GENERALISATION ($\forall$): $\dfrac{\varphi}{\forall \nu \varphi}$.

In the former case we say that ψ is obtained from $\varphi \rightarrow \psi$ and φ by MODUS PONENS, abbreviated (MP), and in the latter case we say that $\forall \nu \varphi$ (where ν can be any variable) is obtained from φ by GENERALISATION, abbreviated ($\forall$).

Using these two inference rules, we are now able to define the notion of **formal proof**: Let $\mathscr{L}$ be a signature (*i.e.*, a possibly empty set of non-logical symbols) and let T be an $\mathscr{L}$-theory (*i.e.*, a possibly empty set of $\mathscr{L}$-formulae). An $\mathscr{L}$-formula ψ is **provable** from T (or provable in T), denoted $\mathsf{T} \vdash \psi$, if there is a F I N I T E sequence $\varphi_0, \ldots, \varphi_n$ of $\mathscr{L}$-formulae such that $\varphi_n \equiv \psi$ (*i.e.*, the formulae φ_n and ψ are identical), and for all i with $i \leqslant n$ we have:

- φ_i is a logical axiom, or
- $\varphi_i \in \mathsf{T}$, or
- there are $j, k < i$ such that $\varphi_j \equiv \varphi_k \to \varphi_i$, or
- there is a $j < i$ such that $\varphi_i \equiv \forall \nu\, \varphi_j$ for some variable ν.

If a formula ψ is not provable from 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. Nevertheless, we shall give two examples:

Example 1.1. For every formula φ we have:

$$\vdash \varphi \to \varphi$$

A formal proof of $\varphi \to \varphi$ is given by

φ_0:	$(\varphi \to ((\varphi \to \varphi) \to \varphi)) \to ((\varphi \to (\varphi \to \varphi)) \to (\varphi \to \varphi))$	instance of L_2
φ_1:	$\varphi \to ((\varphi \to \varphi) \to \varphi)$	instance of L_1
φ_2:	$(\varphi \to (\varphi \to \varphi)) \to (\varphi \to \varphi)$	from φ_0 and φ_1 by (MP)
φ_3:	$\varphi \to (\varphi \to \varphi)$	instance of L_1
φ_4:	$\varphi \to \varphi$	from φ_2 and φ_3 by (MP)

Example 1.2. $\mathsf{PA} \vdash \mathtt{s0} + \mathtt{s0} = \mathtt{ss0}$

φ_0:	$\forall x \forall y(x + \mathtt{s}y = \mathtt{s}(x + y))$	instance of PA_3
φ_1:	$\forall x \forall y(x + \mathtt{s}y = \mathtt{s}(x + y)) \rightarrow \forall y(\mathtt{s0} + \mathtt{s}y = \mathtt{s}(\mathtt{s0} + y))$	instance of L_{11}
φ_2:	$\forall y(\mathtt{s0} + \mathtt{s}y = \mathtt{s}(\mathtt{s0} + y))$	from φ_1 and φ_0 by (MP)
φ_3:	$\forall y(\mathtt{s0} + \mathtt{s}y = \mathtt{s}(\mathtt{s0} + y)) \rightarrow \mathtt{s0} + \mathtt{s0} = \mathtt{s}(\mathtt{s0} + \mathtt{0})$	instance of L_{11}
φ_4:	$\mathtt{s0} + \mathtt{s0} = \mathtt{s}(\mathtt{s0} + \mathtt{0})$	from φ_3 and φ_2 by (MP)
φ_5:	$\forall x(x + \mathtt{0} = x)$	instance of PA_2
φ_6:	$\forall x(x + \mathtt{0} = x) \rightarrow \mathtt{s0} + \mathtt{0} = \mathtt{s0}$	instance of L_{11}
φ_7:	$\mathtt{s0} + \mathtt{0} = \mathtt{s0}$	from φ_6 and φ_5 by (MP)
φ_8:	$\mathtt{s0} + \mathtt{0} = \mathtt{s0} \rightarrow \mathtt{s}(\mathtt{s0} + \mathtt{0}) = \mathtt{ss0}$	instance of L_{17}
φ_9:	$\mathtt{s}(\mathtt{s0} + \mathtt{0}) = \mathtt{ss0}$	from φ_8 and φ_7 by (MP)
φ_{10}:	$\mathtt{s0} + \mathtt{s0} = \mathtt{s0} + \mathtt{s0}$	instance of L_{15}
φ_{11}:	$\varphi_{10} \rightarrow (\varphi_9 \rightarrow (\varphi_{10} \wedge \varphi_9))$	instance of L_5
φ_{12}:	$\varphi_9 \rightarrow (\varphi_{10} \wedge \varphi_9)$	from φ_{11} and φ_{10} by (MP)
φ_{13}:	$\varphi_{10} \wedge \varphi_9$	fromφ_{12} and φ_9 by (MP)
φ_{14}:	$(\varphi_{10} \wedge \varphi_9) \rightarrow (\mathtt{s0} + \mathtt{s0} = \mathtt{s}(\mathtt{s0} + \mathtt{0}) \rightarrow \mathtt{s0} + \mathtt{s0} = \mathtt{ss0})$	instance of L_{16}
φ_{15}:	$\mathtt{s0} + \mathtt{s0} = \mathtt{s}(\mathtt{s0} + \mathtt{0}) \rightarrow \mathtt{s0} + \mathtt{s0} = \mathtt{ss0}$	from φ_{14} and φ_{13} by (MP)
φ_{16}:	$\mathtt{s0} + \mathtt{s0} = \mathtt{ss0}$	from φ_{15} and φ_4 by (MP)

We say that two formulae φ and ψ are **logically equivalent** (or just **equivalent**), denoted $\varphi \Leftrightarrow \psi$, if $\vdash \varphi \leftrightarrow \psi$. More formally:

$$\varphi \Leftrightarrow \psi \quad :\Longleftrightarrow \quad \vdash \varphi \leftrightarrow \psi$$

In other words, if $\varphi \Leftrightarrow \psi$, 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 & ψ are equivalent if and only if $\varphi \leftrightarrow \psi$ is a tautology.

Example 1.3. For every formula φ we have:

$$\varphi \Leftrightarrow \varphi$$

This follows from Example 1.1 and the method of proof $(\rightarrow)$ which will be introduced in the next section.

Example 1.4. Commutativity and associativity of $\wedge$ and $\vee$ are tautological, i.e. $\varphi \wedge \psi \Leftrightarrow \psi \wedge \varphi$ and $\varphi \wedge (\psi \wedge \chi) \Leftrightarrow (\varphi \wedge \psi) \wedge \chi$; and similarly for $\vee$. Again, we omit the proof since it will be trivial once we have proved the DEDUCTION THEOREM (Theorem 1.1) and Proposition 1.2. This legitimizes the notations $\varphi_0 \wedge \cdots \wedge \varphi_n$ resp. $\varphi_0 \vee \cdots \varphi_n$ for $\varphi_0 \wedge (\varphi_1 \wedge (\ldots \wedge \varphi_n) \ldots)$ resp. $\varphi_0 \vee (\varphi_1 \vee (\ldots \vee \varphi_n) \ldots)$.

In Appendix A there is a list of tautologies which will be frequently used in formal proofs.

The Art of Proof

In Example 1.2 we gave a proof of $\mathtt{s0} + \mathtt{s0} = \mathtt{ss0}$ in 17 (!) proof steps. At that point you may have probably asked yourself that if it takes that much effort to prove such a simple statement, how can one ever prove any non-trivial mathematical result using formal proofs. This objection is of course justified; however we will show in this section how one can simplify formal proofs using some methods of proof such as proofs by cases or by contradiction. It is crucial to note that the next results are not theorems of a formal theory but theorems about formal proofs, i.e. they show how - under certain conditions - a formal proof can be transformed into another.

One of the most useful methods is the so-called DEDUCTION THEOREM:

THEOREM 1.1 (DEDUCTION THEOREM). *If* T *is a theory and* $\mathsf{T} \cup \{\psi\} \vdash \varphi$, *where in the proof of* φ *from* $\mathsf{T} \cup \{\psi\}$ *the rule of* GENERALISATION *is not applied to the free variables of* ψ, *then* $\mathsf{T} \vdash \psi \to \varphi$*; and vice versa, if* $\mathsf{T} \vdash \psi \to \varphi$, *then* $\mathsf{T} \cup \{\psi\} \vdash \varphi$*:*

$$\mathsf{T} \cup \{\psi\} \vdash \varphi \iff \mathsf{T} \vdash \psi \to \varphi \tag{DT}$$

Proof. It is clear that $\mathsf{T} \vdash \psi \to \varphi$ implies $\mathsf{T} \cup \{\psi\} \vdash \varphi$. Conversely, suppose that $\mathsf{T} \cup \{\psi\} \vdash \varphi$ holds and let the sequence $\varphi_0, \ldots, \varphi_n$ with $\varphi_n \equiv \varphi$ be a formal proof for φ from $\mathsf{T} \cup \{\psi\}$. For each $i \leqslant n$ we will replace the formula φ_i by a sequence of formulae which ends with $\psi \to \varphi_i$. Let $i \leqslant n$ and assume $\mathsf{T} \vdash \psi \to \varphi_j$ for every $j < i$.

- If φ_i is a logical axiom or $\varphi_i \in \mathsf{T}$, we have

$\varphi_{i,0}$:	φ_i	$\varphi_i \in \mathsf{T}$ or φ_i is a logical axiom
$\varphi_{i,1}$:	$\varphi_i \to (\psi \to \varphi_i)$	instance of L_1
$\varphi_{i,2}$:	$\psi \to \varphi_i$	from $\varphi_{i,1}$ and $\varphi_{i,0}$ by (MP)

- The case $\varphi_i \equiv \psi$ follows directly from Example 1.1.
- If φ_i is obtained by MODUS PONENS from φ_j and $\varphi_k \equiv (\varphi_j \to \varphi_i)$ for some $j < k < i$, we have

$\varphi_{i,0}$:	$\psi \to \varphi_j$	since $j < i$
$\varphi_{i,1}$:	$\psi \to (\varphi_j \to \varphi_i)$	since $k < i$
$\varphi_{i,2}$:	$\varphi_{i,1} \to ((\psi \to \varphi_j) \to (\psi \to \varphi_i))$	instance of L_2
$\varphi_{i,3}$:	$(\psi \to \varphi_j) \to (\psi \to \varphi_i)$	from $\varphi_{i,2}$ and $\varphi_{i,1}$ by (MP)
$\varphi_{i,4}$:	$\psi \to \varphi_i$	from $\varphi_{i,3}$ and $\varphi_{i,0}$ by (MP)

- If $\varphi_i \equiv \forall x \varphi_j$ with $j < i$ and $x \notin \mathrm{free}(\psi)$, the claim follows from

$\varphi_{i,0}$:	$\psi \to \varphi_j$	since $j < i$
$\varphi_{i,1}$:	$\forall x(\psi \to \varphi_j)$	from $\varphi_{i,0}$ by $(\forall)$
$\varphi_{i,2}$:	$\forall x(\psi \to \varphi_j) \to (\psi \to \varphi_i)$	instance of L_{13}
$\varphi_{i,3}$:	$\psi \to \varphi_i$	from $\varphi_{i,2}$ and $\varphi_{i,1}$ by (MP)

⊣

As an application of the DEDUCTION THEOREM, we show that the equality relation is symmetric, which is (??). We first work with the empty theory and show that $\{x = y\} \vdash y = x$:

φ_0:	$(x = y \wedge x = x) \to (x = x \to y = x)$	instance of L_{17}
φ_1:	$x = x$	instance of L_{16}
φ_2:	$x = y$	$x = y$ belongs to $\{x = y\}$
φ_3:	$x = x \to (x = y \to (x = y \wedge x = x))$	instance of L_5
φ_4:	$x = y \to (x = y \wedge x = x)$	from φ_3 and φ_1 by (MP)
φ_5:	$x = y \wedge x = x$	from φ_4 and φ_2 by (MP)
φ_6:	$x = x \to y = x$	from φ_0 and φ_5 by (MP)
φ_7:	$y = x$	from φ_6 and φ_1 by (MP)

Thus, we have $\{x = y\} \vdash y = x$, and by the Deduction Theorem 1.1 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 (see EXERCISE 1).

PROPOSITION 1.2. *Let T be an $\mathscr{L}$-theory, and $\varphi \,\&\, \psi$ any two $\mathscr{L}$-formulae. Then we have:*

$$\mathsf{T} \vdash \varphi \quad \textit{and} \quad \mathsf{T} \vdash \psi \quad \Longleftrightarrow \quad \mathsf{T} \vdash \varphi \wedge \psi \tag{$\wedge$}$$

Proof. First we assume $\mathsf{T} \vdash \varphi$ and $\mathsf{T} \vdash \psi$, and show $\mathsf{T} \vdash \varphi \wedge \psi$:

φ_0:	$\psi \to \big(\varphi \to (\varphi \wedge \psi)\big)$	instance of L_5
φ_1:	ψ	provable from T by assumption
φ_2:	$\varphi \to (\varphi \wedge \psi)$	from φ_0 and φ_1 by (MP)
φ_3:	φ	provable from T by assumption
φ_4:	$\varphi \wedge \psi$	from φ_2 and φ_3 by (MP)

Now we assume $\mathsf{T} \vdash \varphi \wedge \psi$, and show $\mathsf{T} \vdash \varphi$ ($\mathsf{T} \vdash \psi$ is similar):

φ_0:	$(\varphi \wedge \psi) \to \varphi$	instance of L_3
φ_1:	$\varphi \wedge \psi$	provable from T by assumption
φ_2:	φ	from φ_0 and φ_1 by (MP)

$\dashv$

As an immediate consequence of the definition of "$\leftrightarrow$" and PROPOSITION 1.2 we get:

$$\mathsf{T} \vdash \varphi \to \psi \quad \text{and} \quad \mathsf{T} \vdash \psi \to \varphi \quad \Longleftrightarrow \quad \mathsf{T} \vdash \varphi \leftrightarrow \psi \tag{$\leftrightarrow$}$$

PROPOSITION 1.3 (Ex falso quodlibet). *Let T be an $\mathscr{L}$-theory and φ an arbitrary $\mathscr{L}$-formula. Then for every $\mathscr{L}$-formula ψ we have:*

$$\mathsf{T} \vdash \varphi \wedge \neg\varphi \quad \Longrightarrow \quad \mathsf{T} \vdash \psi \tag{$\boxtimes$}$$

Proof. Let ψ be any $\mathscr{L}$-formula and assume that $\mathsf{T} \vdash \varphi \wedge \neg\varphi$ for some $\mathscr{L}$-formula φ. By $(\wedge)$ we clearly have $\mathsf{T} \vdash \varphi$ and $\mathsf{T} \vdash \neg\varphi$. Now the instance $\neg\varphi \to (\varphi \to \psi)$ of the logical axiom L_{10} and two applications of MODUS PONENS imply $\mathsf{T} \vdash \psi$. ⊣

Notice that PROPOSITION 1.3 implies that if we can derive a contradiction from T, we can derive *every* formula we like, even the impossible, which shall be denoted by

$$\mathsf{T} \vdash \boxtimes.$$

PROPOSITION 1.4 (Proof by Cases). *Let T be an $\mathscr{L}$-theory and φ, ψ, and α some $\mathscr{L}$-formulae. Then the following four statements hold:*

$$\mathsf{T} \vdash \varphi \vee \psi \ \text{ and } \ \mathsf{T} \cup \{\varphi\} \vdash \alpha \ \text{ and } \ \mathsf{T} \cup \{\psi\} \vdash \alpha \quad \Longrightarrow \quad \mathsf{T} \vdash \alpha \tag{$\vee 0$}$$

$$\mathsf{T} \cup \{\varphi\} \vdash \alpha \ \text{ and } \ \mathsf{T} \cup \{\neg\varphi\} \vdash \alpha \quad \Longrightarrow \quad \mathsf{T} \vdash \alpha \tag{$\vee 1$}$$

where $(\forall)$ is not applied to any of the free variables of φ or ψ in the proof of α from $\mathsf{T} \cup \{\varphi\}$, $\mathsf{T} \cup \{\psi\}$ or $\mathsf{T} \cup \{\neg\varphi\}$ respectively. Furthermore, we have

$$\mathsf{T} \vdash \varphi \vee \psi \quad \Longrightarrow \quad \mathsf{T} \cup \{\neg\varphi\} \vdash \psi \tag{$\vee 2$}$$

$$\mathsf{T} \vdash \varphi \vee \psi \ \text{ and } \ \mathsf{T} \cup \{\varphi\} \vdash \boxtimes \quad \Longrightarrow \quad \mathsf{T} \vdash \psi \tag{$\vee 3$}$$

where $(\forall)$ is not applied to any of the free variables of φ in the proof of $\boxtimes$ from $\mathsf{T} \cup \{\varphi\}$.

Proof. For $(\vee 0)$ we assume $\mathsf{T} \vdash \varphi \vee \psi$, $\mathsf{T} \cup \{\varphi\} \vdash \alpha$ and $\mathsf{T} \cup \{\psi\} \vdash \alpha$.

φ_0:	$\varphi \to \alpha$	from $\mathsf{T} \cup \{\varphi\} \vdash \alpha$ by (DT)
φ_1:	$\psi \to \alpha$	from $\mathsf{T} \cup \{\psi\} \vdash \alpha$ by (DT)
φ_2:	$(\varphi \to \psi) \to ((\psi \to \psi) \to ((\varphi \vee \psi) \to \psi))$	instance of L_8
φ_3:	$(\psi \to \alpha) \to ((\varphi \vee \psi) \to \alpha)$	from φ_2 and φ_0 by (MP)
φ_4:	$(\varphi \vee \psi) \to \alpha$	from φ_3 and φ_1 by (MP)
φ_4:	$\varphi \vee \psi$	by assumption
φ_5:	α	from φ_4 and φ_5 by (MP)

$(\vee 1)$ is a special case of $(\vee 0)$, since $\mathsf{T} \vdash \varphi \vee \neg\varphi$ by L_0. For $(\vee 2)$ suppose $\mathsf{T} \vdash \varphi \vee \psi$ and put $\mathsf{T}' = \mathsf{T} \cup \{\neg\varphi\}$. By $(\vee 0)$ it suffices to prove $\mathsf{T}' \cup \{\varphi\} \vdash \psi$ and $\mathsf{T}' \cup \{\psi\} \vdash \psi$. The first statement follows from $(\wedge)$ and Proposition 1.3 and the second one is trivial. For $(\vee 3)$ observe that $(\vee 1)$ reduces the claim $\mathsf{T} \cup \{\varphi\} \vdash \psi$ and $\mathsf{T} \cup \{\neg\varphi\} \vdash \psi$ which are direct consequences of $(\boxtimes)$ and $(\vee 2)$ respectively. ⊣

COROLLARY 1.5 (Generalised Proof by Cases). *Let T be an $\mathscr{L}$-theory and $\psi_0, \ldots, \psi_n, \varphi$ some $\mathscr{L}$-formulae. Then we have:*

$$\mathsf{T} \vdash \psi_0 \vee \cdots \vee \psi_n \ \text{ and } \ \mathsf{T} \cup \{\psi_i\} \vdash \varphi \text{ for all } i \leqslant n \quad \Longrightarrow \quad \mathsf{T} \vdash \varphi,$$

where $(\forall)$ *is not applied to any of the free variables of* ψ_i *in the proof of* φ *from* $\mathsf{T} \cup \{\psi_i\}$.

Since Corollary 1.5 is just a generalization of $(\vee 1)$, we will also denote all instance of this form by $(\vee 1)$.

Proof of Corollary 1.5. We proceed by induction on $n \geqslant 1$. For $n = 1$ the statement is exactly $(\vee 0)$. Now assume that $\mathsf{T} \vdash \psi_0 \vee \ldots \vee \psi_n \vee \psi_{n+1}$ and $\mathsf{T} \cup \{\psi_i\} \vdash \varphi$ for all $i \leqslant n+1$. Let $\mathsf{T}' :\equiv \mathsf{T} \cup \{\psi_0 \vee \ldots \vee \psi_n\}$ and observe that $\mathsf{T}' \vdash \psi_0 \vee \ldots \vee \psi_n$ and $\mathsf{T}' \cup \{\psi_i\} \vdash \varphi$, so by induction hypothesis $\mathsf{T}' \vdash \varphi$. By the DEDUCTION THEOREM this implies $\mathsf{T} \vdash \psi_0 \vee \ldots \vee \psi_n \to \varphi$. Moreover, by another application of (DT) we also have $\mathsf{T} \vdash \psi_{n+1} \to \varphi$. Using L_8 and twice (DT) we obtain $\mathsf{T} \vdash \psi_0 \vee \ldots \vee \psi_n \vee \psi_{n+1} \to \varphi$, hence (DT) yields the claim. ⊣

COROLLARY 1.6 (Proof by contradiction). *Let* T *be a theory, and* φ *be an arbitrary formula. Then the following statements hold:*

$$\begin{aligned} \mathsf{T} \cup \{\neg\varphi\} \vdash \square \quad &\Longrightarrow \quad \mathsf{T} \vdash \varphi, \textit{ respectively} \\ \mathsf{T} \cup \{\varphi\} \vdash \square \quad &\Longrightarrow \quad \mathsf{T} \vdash \neg\varphi. \end{aligned}$$

Proof. We consider only the first statement, since both proofs are similar. By $(\vee 1)$ it is enough to check $\mathsf{T} \cup \{\varphi\} \vdash \varphi$ and $\mathsf{T} \cup \{\neg\varphi\} \vdash \varphi$. The first condition is clearly satisfied and the second one follows directly from $(\wedge)$ and $(\square)$. ⊣

PROPOSITION 1.7 (Contrapositon). *Let* T *be an* $\mathscr{L}$*-theory and* φ *&* ψ *two arbitrary* $\mathscr{L}$*-formulae. Then we have:*

$$\mathsf{T} \cup \{\neg\psi\} \vdash \neg\varphi \quad \Longrightarrow \quad \mathsf{T} \cup \{\varphi\} \vdash \psi \tag{CP}$$

Proof. By $(\vee 2)$ it suffices to show $\mathsf{T} \cup \{\varphi, \psi\} \vdash \psi$ and $\mathsf{T} \cup \{\varphi, \neg\psi\} \vdash \psi$. The first statement is obvious and for the second one note that $\mathsf{T} \cup \{\varphi, \neg\psi\} \vdash \varphi \wedge \neg\varphi$ by $(\wedge)$ and thus by Proposition 1.3 $\mathsf{T} \cup \{\varphi, \neg\psi\} \vdash \psi$. ⊣

PROPOSITION 1.8 ($\exists$−Introduction). *Let* T *be a set of formulae,* $\varphi(x)$ *a formula with* $x \in \operatorname{free}(\varphi)$ *and* ψ *an arbitary formula. Then:*

$$\mathsf{T} \cup \{\varphi(x)\} \vdash \psi \quad \Longrightarrow \quad \mathsf{T} \cup \{\exists x \varphi(x)\} \vdash \psi. \tag{$\exists$}$$

Proof. Using the DEDUCTION THEOREM we obtain $\mathsf{T} \vdash \varphi(x) \to \psi$. Then the following formal proof implies the claim:

$$\begin{array}{lll} \varphi_0: & \varphi(x) \to \psi & \text{since } \mathsf{T} \vdash \varphi(x) \to \psi \\ \varphi_1: & \forall x(\varphi(x) \to \psi & \text{from } \varphi_0 \text{ by } (\forall) \\ \varphi_2: & \forall x(\varphi(x) \to \psi) \to (\exists x \varphi(x) \to \psi) & \text{instance of } \mathsf{L}_{14} \\ \varphi_3: & \exists x \varphi(x) \to \psi & \text{from } \varphi_2 \text{ and } \varphi_1 \text{ by (MP)} \end{array}$$

⊣

THEOREM 1.9 (DEMORGAN'S LAWS). *If $\varphi_0, \dots, \varphi_n$ are formulae, then:*

(a) $\neg(\varphi_0 \wedge \dots \wedge \varphi_n) \Leftrightarrow (\neg\varphi_1 \vee \dots \vee \neg\varphi_n)$
(b) $\neg(\varphi_0 \vee \dots \vee \varphi_n) \Leftrightarrow (\neg\varphi_1 \wedge \dots \wedge \neg\varphi_n)$
(c) $\varphi_0 \to \big(\varphi_1 \to (\dots \to \varphi_n)\dots\big) \Leftrightarrow \neg(\varphi_0 \wedge \dots \wedge \varphi_n)$

Proof. ⊣

THEOREM 1.10 (GENERALISED DEDUCTION THEOREM). *If T is any theory and $\mathsf{T} \cup \{\psi_1, \dots, \psi_n\} \vdash \varphi$, where in the proof of φ from $\mathsf{T} \cup \{\psi_1, \dots, \psi_n\}$ the rule of* GENERALISATION *is not applied to any of the free variables of $\psi_1, \dots, \psi_n$, then $\mathsf{T} \vdash (\psi_1 \wedge \dots \wedge \psi_n) \to \varphi$; and vice versa:*

$$\mathsf{T} \cup \{\psi_1, \dots, \psi_n\} \vdash \varphi \iff \mathsf{T} \vdash (\psi_1 \wedge \dots \wedge \psi_n) \to \varphi \qquad \text{(GDT)}$$

Proof. Follows immediately from the DEDUCTION THEOREM and from part (c) of DEMORGAN'S LAWS. ⊣

THEOREM 1.11 (3-SYMBOLS). *For every each φ there is an equivalent formula ψ which contains only "$\neg$" and "$\wedge$" as logical operators and "$\exists$" as quantifier.*

Proof. ⊣

Definition of **Prenex Normal Form**, abbreviated PNF.

THEOREM 1.12 (PRENEX NORMAL FORM THEOREM). *For every formula φ there is an equivalent formula ψ which is in* PNF.

Proof. ⊣

THEOREM 1.13 (VARIABLE SUBSTITUTION THEOREM). *For every formula φ there is an equivalent formula ψ which contains just variables among $v_0, v_1, \dots$*

Proof. ⊣

Consistency & Compactness

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 1.14. *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* $\mathsf{T} \nvdash \neg\varphi$*.*

(c) *If* $\neg\,\mathrm{Con}(\mathsf{T} + \varphi)$*, for some* $\mathscr{L}$*-formula* φ*, then* $\mathsf{T} \vdash \neg\varphi$*.*

(d) *If* $\mathsf{T} \vdash \neg\varphi$*, for some* $\mathscr{L}$*-formula* φ*, then* $\neg\,\mathrm{Con}(\mathsf{T} + \varphi)$*.*

Proof. (a) This is just PROPOSITION 1.3.

(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

(c) Assume that for some $\mathscr{L}$-formula φ we have $\neg\,\mathrm{Con}(\mathsf{T} + \varphi)$. By (b) we get $\mathsf{T} + \varphi \vdash \psi$ for every $\mathscr{L}$-formula ψ. In particular we get $\mathsf{T} + \varphi \vdash \neg\varphi$ and by the DEDUCTION THEOREM we get $\mathsf{T} \vdash \varphi \to \neg\varphi$:

$\mathsf{T} \vdash \varphi \to \neg\varphi$	consequence of assumption
$\mathsf{T} \vdash \varphi \to \varphi$	TAUTOLOGY (A.1)
$\mathsf{T} \vdash (\varphi \to \varphi) \to ((\varphi \to \neg\varphi) \to \neg\varphi)$	L_9
$\mathsf{T} \vdash (\varphi \to \neg\varphi) \to \neg\varphi$	by MODUS PONENS
$\mathsf{T} \vdash \neg\varphi$	by MODUS PONENS

(d) Assume that for some $\mathscr{L}$-formula φ we have $\mathsf{T} \vdash \neg\varphi$. By extending T, we also have $\mathsf{T} + \varphi \vdash \neg\varphi$:

$\mathsf{T} + \varphi \vdash \neg\varphi$	consequence of assumption
$\mathsf{T} + \varphi \vdash \varphi$	φ belongs to $\mathsf{T} + \varphi$
$\mathsf{T} + \varphi \vdash \varphi \wedge \neg\varphi$	TAUTOLOGY (B)

Hence, $\mathsf{T} + \varphi$ is inconsistent, *i.e.*, $\neg\,\mathrm{Con}(\mathsf{T} + \varphi)$. $\dashv$

If we design a theory T (*e.g.*, a set of axioms), we have to make sure that T is consistent. However, as we shall see later, in many cases this task is impossible.

We conclude this chapter with the COMPACTNESS THEOREM, which is a powerful tool in order to construct non-standard models of Peano Arithmetic or of Set Theory. On the one hand, it is just a consequence of the fact that formal proofs are F I N I T E sequences of formulae. On the other hand, the COMPACTNESS THEOREM is the main tool to prove that a given set of sentences is consistent with some given theory.

THEOREM 1.15 (COMPACTNESS THEOREM). *Let* T *be an arbitrary set of formulae. Then* T *is consistent if and only if every finite subset* T' *of* T *is consistent.*

Proof. Obviously, if T is consistent, then every finite subset T' 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 T' is this finite set of formulae, then $\mathsf{T}' \vdash \varphi \wedge \neg\varphi$, which shows that T', a finite subset of T, is inconsistent. $\dashv$

Semi-formal Proofs

Previously we have shown that formal proofs can be simplified by applying methods of proof such as case distinctions, proofs by contradiction or contraposition. However, to make proofs even more natural, it is useful to use natural language in order to describe a proof step as in an "informal" mathematical proof.

Example 1.1 We want to prove the tautology $\vdash \varphi \to \neg\neg\varphi$. Instead of writing out the whole formal proof which is quite tedious, we can apply our methods of proof introduced above.

The first modification we make is to use (DT) to obtain the new goal

$$\{\varphi\} \vdash \neg\neg\varphi.$$

This seems like a target which can easily be shown using a proof by contradiction: We apply $\lightning$ and get the goal

$$\{\varphi, \neg\varphi\} \vdash \lightning$$

which by $(\wedge)$ is again a consequence of the tautological goals

$$\{\varphi, \neg\varphi\} \vdash \varphi \quad \text{and} \quad \{\varphi, \neg\varphi\} \vdash \neg\varphi.$$

To sum up, this procedure can actually be transformed back into a formal proof, so it suffices as a proof of $\vdash \varphi \to \neg\neg\varphi$. Now this is still not completely satisfactory, since we would like to write the proof in natural language. A possible translation could thus be

Proof. We want to prove that φ implies $\neg\neg\varphi$. Assume φ. Suppose for a contradiction that $\neg\varphi$. But then we have φ and $\neg\varphi$. Contradiction. $\dashv$

We will now show in a systematic way how formal proofs can - in principal - be replaced by semi-formal proofs, which make use of a **controlled natural language**, *i.e.*, a limited vocabulary consisting of natural language phrases such as "assume that" which are often used in mathematical proof texts. This language is controlled in the sense that its allowed vocabulary is only a subset of the entire English vocabulary and that every word resp. phrase has a unique precisely defined interpretation. However, for the sake of a nice proof style, we will not always stick to this limited vocabulary. Moreover, this section should be considered as a hint of how formal proofs can be formulated using a controlled natural language as well as a justification for working with natural language proofs rather than formal ones.

Every statement we would like to prove formally is of the form $\Gamma \vdash \varphi$, where Γ is a set of formulae, denoted the set of **premises**, and φ is the formula to be verified, named **target**. The whole statement $\Gamma \vdash \varphi$ (which we will describe as a **goal**) states that there is a formal proof of φ from the logical axioms L_0-L_{17} and the premises as non-logical axioms using the inference rules (MP) and $(\forall)$. Now instead of listing a formal proof, we can step by step reduce our current goal to a simpler one using the methods of proof from before; the idea is to follow this procedure until the target is tautological as in the case of Example 1.1.

Methods of proof can be considered as operations on the premises and the targets. the proof by contraposition for example adds the negation of the target to the premises and replaces the original target by the negation of the premise from which it shall be derived:

If we want to show

$$\Gamma \cup \{\psi\} \vdash \varphi,$$

we can prove instead

$$\Gamma \cup \{\neg\varphi\} \vdash \neg\psi.$$

A slightly different example is the proof of a conjunction

$$\Gamma \vdash \varphi \wedge \psi,$$

which is usually split into the two goals given by

$$\Gamma \vdash \varphi \text{ and } \Gamma \vdash \psi.$$

Thus we have to revise our first attempt and interpret methods of proof as operations on FINITE lists of goals consisting of premises and targets.

In the following we will list some of the most common methods and assign them their (respectively one possible) natural language meaning:

Operations on targets (Backward reasoning)

- Targets are often of the universal conditional form $\forall\boldsymbol{\nu}(\varphi(\boldsymbol{\nu}) \rightarrow \psi(\boldsymbol{\nu}))$, where $\boldsymbol{\nu}$ is a FINITE sequence of variables. In particular, this pattern includes the purely universal formulae $\forall\boldsymbol{\nu}\psi(\boldsymbol{\nu})$ by taking φ to be $\top$ as well as simple conditionals of the form $\varphi \rightarrow \psi$. Now the usual procedure is to reduce $\Gamma \vdash \forall\boldsymbol{\nu}(\varphi(\boldsymbol{\nu}) \rightarrow \psi(\boldsymbol{\nu}))$ to $\Gamma \cup \{\varphi(\boldsymbol{\nu})\} \vdash \psi(\boldsymbol{\nu})$ using $(\forall)$ various times and (DT). This can be rephrased as

  Assume $\varphi(\boldsymbol{\nu})$. Then ... This shows $\psi(\boldsymbol{\nu})$.

- As already mentioned above, if the target is a conjunction $\varphi \wedge \psi$, one can show them separately using $(\wedge)$. This step is usually executed without mentioning it explicitly.
- If the target is a negation $\neg\varphi$, one often uses a proof by contradiction or by contraposition: In the first case we transform $\Gamma \vdash \neg\varphi$ to $\Gamma \cup \{\varphi\} \vdash \square$ and use the natural language notation

  Suppose for a contradiction that φ. Then ... Contradiction.

  In the latter case, we want to go from $\Gamma \cup \{\neg\psi\} \vdash \neg\varphi$ to $\Gamma \cup \{\varphi\} \vdash \psi$ resp. in its positive version from $\Gamma \cup \{\psi\} \vdash \neg\varphi$ to $\Gamma \cup \{\varphi\} \vdash \neg\psi$. In both cases we can mark this with the keyword *contraposition*, e.g. as

  We proceed by *contraposition... This shows $\neg\varphi$.*

- If the target is an existential formula $\exists x\varphi$, then by L_{12} and (MP) it suffices to find a witness τ such that $\Gamma \vdash \varphi(\tau)$.
- In order to prove a disjunction $\varphi \vee \psi$, by L_5 and L_6 and (MP) it is enough to prove either $\Gamma \vdash \varphi$ or $\Gamma \vdash \psi$.

Observe that the last two operations are usually performed in the very end after many operations on the premises such as adding further premises and splitting goals into subgoals.

Operations on premises (Forwards reasoning)

- By $(\wedge)$, conjunctive premises $\varphi \wedge \psi$ can be split into two premises φ, ψ; i.e. $\Gamma \cup \{\varphi \wedge \psi\} \vdash \chi$ can be reduced to $\Gamma \cup \{\varphi, \psi\} \vdash \chi$.
- Disjunctive premises are used
- Intermediate proof steps: Often we want to prove first some intermediate statement which shall then be applied to resolve the target. Formally this means that we want to show $\Gamma \vdash \varphi$ by showing first $\Gamma \vdash \psi$ and then we add ψ to the list of premises and check $\Gamma \cup \{\psi\} \vdash \varphi$. Clearly, if we have $\Gamma \vdash \psi$ and $\Gamma \cup \{\psi\} \vdash \varphi$, using (DT) and (MP) we obtain that $\Gamma \vdash \varphi$. In a semi-formal proof this can be described by

  We show *first ψ... This proves ψ.*

  Note that it is important to mark when the proof of the intermediate statement ψ ends, because from this point on, ψ can be used as a new premise.

Observe that in any case, once a goal $\Gamma \vdash \varphi$ is reduced to a tautology, it can be removed from the list of goals. This should be marked by a phrase like

This shows/proves φ

so that it is clear that we go on to the next goal. The proof is complete as soon as no unresolved goals remain.

What is the use of such a formalized natural proof language? First of all, it increases readibility for the audience. Secondly, by giving some of the common natural language phrases appearing in proof texts a precise formal definition, we show how - in principal - one could write formal proofs with a controlled natural language input. This input could then be parsed into a formal proof and then be verified by a proof checking system.

We would like to emphasize that this section should only be considered a motivation rather than a precise description of how formal proofs can be translated into semi-formal ones and vice versa. Nevertheless, it suffices to understand that and how this can theoretically be achieved. Therefore, in subsubsequent chapters, especially in Chapters **??** and **??**, we will often present semi-formal proofs.

Exercises

0. Something with terms.
1. The equality relation is transitive.

Appendix A
Tautologies

In this section we give a list of some of the most important tautologies. Many of them have been used explicitly and implicitly in several formal proofs.

(A.1) $\vdash \varphi \to \varphi$
(A.0) $\vdash \varphi \leftrightarrow \varphi$

(B) $\{\psi, \varphi\} \vdash \varphi \wedge \psi$

(C) $\vdash (\psi \to \varphi) \to (\psi \to \forall \nu \varphi)$ [for $\nu \notin \text{free}(\psi)$]

(D.1) $\{\varphi_0 \to \varphi_1, \varphi_1 \to \varphi_2\} \vdash \varphi_0 \to \varphi_2$
(D.2) $\{\varphi_0 \to \psi, \varphi_1 \to \psi\} \vdash (\varphi_0 \vee \varphi_1) \to \psi$
(D.3) $\{\psi \to \varphi_0, \psi \to \varphi_1\} \vdash \psi \to (\varphi_0 \wedge \varphi_1)$

(E) $\vdash \varphi \to \big(\psi \to (\varphi \wedge \psi)\big)$

(F.1) $\vdash \varphi \to \neg\neg\varphi$
(F.2) $\vdash \neg\neg\varphi \to \varphi$
(F.0) $\vdash \varphi \leftrightarrow \neg\neg\varphi$

(G.1) $\vdash (\varphi \to \psi) \to (\neg\psi \to \neg\varphi)$
(G.2) $\vdash (\neg\psi \to \neg\varphi) \to (\varphi \to \psi)$
(G.0) $\vdash (\varphi \to \psi) \leftrightarrow (\neg\psi \to \neg\varphi)$

(H.0) $\{\varphi \leftrightarrow \psi\} \vdash \neg\varphi \leftrightarrow \neg\psi$
(H.1) $\{\varphi \leftrightarrow \varphi', \psi \leftrightarrow \psi'\} \vdash (\varphi \to \psi) \leftrightarrow (\varphi' \to \psi')$
(H.2) $\{\varphi \leftrightarrow \varphi', \psi \leftrightarrow \psi'\} \vdash (\varphi \vee \psi) \leftrightarrow (\varphi' \vee \psi')$
(H.3) $\{\varphi \leftrightarrow \varphi', \psi \leftrightarrow \psi'\} \vdash (\varphi \wedge \psi) \leftrightarrow (\varphi' \wedge \psi')$

(I.1) $\vdash (\varphi_1 \wedge \varphi_2) \leftrightarrow (\varphi_2 \wedge \varphi_1)$
(I.2) $\vdash (\varphi_1 \wedge \varphi_2) \wedge \varphi_3 \leftrightarrow \varphi_1 \wedge (\varphi_2 \wedge \varphi_3)$

(J.1) $\vdash (\varphi_1 \vee \varphi_2) \leftrightarrow (\varphi_2 \vee \varphi_1)$
(J.2) $\vdash (\varphi_1 \vee \varphi_2) \vee \varphi_3 \leftrightarrow \varphi_1 \vee (\varphi_2 \vee \varphi_3)$

(K.1) $\vdash (\neg\varphi \vee \psi) \rightarrow (\varphi \rightarrow \psi)$
(K.2) $\vdash (\varphi \rightarrow \psi) \rightarrow (\neg\varphi \vee \psi)$
(K.0) $\vdash (\varphi \rightarrow \psi) \leftrightarrow (\neg\varphi \vee \psi)$

(L.1) $\vdash (\neg\varphi \vee \neg\psi) \rightarrow \neg(\varphi \wedge \psi)$
(L.2) $\vdash \neg(\varphi \wedge \psi) \rightarrow (\neg\varphi \vee \neg\psi)$
(L.0) $\vdash \neg(\varphi \wedge \psi) \leftrightarrow (\neg\varphi \vee \neg\psi)$

(M.1) $\vdash \big(\varphi_1 \rightarrow (\varphi_2 \rightarrow \varphi_3)\big) \leftrightarrow \big((\varphi_1 \wedge \varphi_2) \rightarrow \varphi_3\big)$
(M.2) $\vdash \neg(\varphi \vee \psi) \leftrightarrow (\neg\varphi \wedge \neg\psi)$

(N.1) $\vdash (\varphi_1 \wedge \varphi_2) \vee \varphi_3 \rightarrow (\varphi_1 \vee \varphi_3) \wedge (\varphi_2 \vee \varphi_3)$
(N.2) $\vdash (\varphi_1 \vee \varphi_3) \wedge (\varphi_2 \vee \varphi_3) \rightarrow (\varphi_1 \wedge \varphi_2) \vee \varphi_3$
(N.0) $\vdash (\varphi_1 \wedge \varphi_2) \vee \varphi_3 \leftrightarrow (\varphi_1 \vee \varphi_3) \wedge (\varphi_2 \vee \varphi_3)$

(O) $\vdash (\varphi_1 \vee \varphi_2) \wedge \varphi_3 \leftrightarrow (\varphi_1 \wedge \varphi_3) \vee (\varphi_2 \wedge \varphi_3)$

(P.1) $\vdash x = y \leftrightarrow y = x$
(P.2) $\vdash (x = y \wedge y = z) \rightarrow x = z$

(Q.1) $\vdash \varphi(\nu) \leftrightarrow \varphi(\nu')$ [if ν' does not appear in $\varphi(\nu)$]
(Q.2) $\vdash \exists\nu\varphi(\nu) \leftrightarrow \exists\nu'\varphi(\nu')$ [if ν' does not appear in $\varphi(x)$]
(Q.3) $\vdash \forall\nu\varphi(\nu) \leftrightarrow \forall y\varphi(\nu')$ [if ν' does not appear in $\varphi(\nu)$]

(R.1) $\{\varphi \leftrightarrow \psi\} \vdash \forall\nu\varphi \leftrightarrow \forall\nu\psi$
(R.2) $\{\varphi \leftrightarrow \psi\} \vdash \exists\nu\varphi \leftrightarrow \exists\nu\psi$

(S.1) $\vdash \neg\exists\nu\varphi \rightarrow \forall\nu\neg\varphi$
(S.2) $\vdash \neg\forall\nu\neg\varphi \rightarrow \exists\nu\varphi$
(S.3) $\vdash \exists\nu\varphi \rightarrow \neg\forall\nu\neg\varphi$
(S.0) $\vdash \exists\nu\varphi \leftrightarrow \neg\forall\nu\neg\varphi$

(T) $\vdash \forall\nu\varphi \leftrightarrow \neg\exists\nu\neg\varphi$

(U.1) $\vdash \exists x \exists y \varphi \leftrightarrow \exists y \exists x \varphi$
(U.2) $\vdash \exists x \exists x \varphi \leftrightarrow \exists x \varphi$
(U.3) $\vdash \forall x \exists x \varphi \leftrightarrow \exists x \varphi$
(U.4) $\vdash \exists x \forall x \varphi \leftrightarrow \forall x \varphi$

(V.1) $\vdash \big(\exists x \varphi \wedge \exists y \psi\big) \leftrightarrow \big(\exists x \exists y (\varphi \wedge \psi)\big)$ [for $x \notin \mathrm{free}(\psi)$, $y \notin \mathrm{free}(\varphi)$]
(V.2) $\vdash \big(\forall x \varphi \wedge \forall y \psi\big) \leftrightarrow \big(\forall x \forall y (\varphi \wedge \psi)\big)$ [for $x \notin \mathrm{free}(\psi)$, $y \notin \mathrm{free}(\varphi)$]
(V.3) $\vdash \big(\exists x \varphi \wedge \forall y \psi\big) \leftrightarrow \big(\exists x \forall y (\varphi \wedge \psi)\big)$ [for $x \notin \mathrm{free}(\psi)$, $y \notin \mathrm{free}(\varphi)$]
(V.4) $\vdash \big(\exists x \varphi \wedge \psi\big) \leftrightarrow \big(\exists x (\varphi \wedge \psi)\big)$ [for $x \notin \mathrm{free}(\psi)$]
(V.5) $\vdash \big(\forall x \varphi \wedge \psi\big) \leftrightarrow \big(\forall x (\varphi \wedge \psi)\big)$ [for $x \notin \mathrm{free}(\psi)$]

(W.1) $\vdash \big(\exists x \varphi \vee \exists y \psi\big) \leftrightarrow \big(\exists x \exists y (\varphi \vee \psi)\big)$ [for $x \notin \mathrm{free}(\psi)$, $y \notin \mathrm{free}(\varphi)$]
(W.2) $\vdash \big(\forall x \varphi \vee \forall y \psi\big) \leftrightarrow \big(\forall x \forall y (\varphi \vee \psi)\big)$ [for $x \notin \mathrm{free}(\psi)$, $y \notin \mathrm{free}(\varphi)$]
(W.3) $\vdash \big(\exists x \varphi \vee \forall y \psi\big) \leftrightarrow \big(\exists x \forall y (\varphi \vee \psi)\big)$ [for $x \notin \mathrm{free}(\psi)$, $y \notin \mathrm{free}(\varphi)$]
(W.4) $\vdash \big(\exists x \varphi \vee \psi\big) \leftrightarrow \big(\exists x (\varphi \vee \psi)\big)$ [for $x \notin \mathrm{free}(\psi)$]
(W.5) $\vdash \big(\forall x \varphi \vee \psi\big) \leftrightarrow \big(\forall x (\varphi \vee \psi)\big)$ [for $x \notin \mathrm{free}(\psi)$]

References

1. ARISTOTLE, ***Topics***, Athens, *published by Andronikos of Rhodos around* 40 B.C.
2. JOHN L. BELL AND ALAN B. SLOMSON, ***Models and Ultraproducts: An Introduction***, North-Holland, Amsterdam, 1969.
3. VILNIS DETLOVS, ***Matemātiskā logika***, Izdevniecība "Zvaigzne", Riga, 1974.
4. HEINZ-DIETER EBBINGHAUS, JÖRG FLUM, AND WOLFGANG THOMAS, ***Mathematical Logic***, 2nd ed., *English translation of* [5], [Undergraduate Texts in Mathematics], Springer-Verlag, New York, 1994.
5. ______, ***Einführung in die mathematische Logik***, 4th ed., Spektrum Akademischer Verlag, Heidelberg · Berlin, 1996.
6. EUCLID, ***The Thirteen Books of the Elements***, Volume I: Books I & II [translated with introduction and commentary by Sir Thomas L. Heath], Dover, 1956.
7. KURT GÖDEL, *Über die Vollständigkeit des Logikkalküls*, Dissertation (1929), University of Vienna (Austria), (reprinted and translated into English in [11].
8. ______, *Die Vollständigkeit der Axiome des logischen Funktionenkalküls*, ***Monatshefte für Mathematik und Physik***, vol. 37 (1930), 349–360 (see [17, 11] for a translation into English).
9. ______, *Einige metamathematische Resultate über Entscheidungsdefinitheit und Widerspruchsfreiheit*, ***Anzeiger der Akademie der Wissenschaften in Wien, mathematisch-naturwissenschaftliche Klasse***, vol. 67 (1930), 214–215 (see [17, 11] for a translation into English).
10. ______, *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme*, ***Monatshefte für Mathematik und Physik***, vol. 38 (1931), 173–198 (see [17, 11] for a translation into English).
11. ______, ***Collected Works, Volume I: Publications 1929–1936***, edited by S. Feferman (Editor-in-chief), J. W. Dawson, Jr., S. C. Kleene, G. H. Moore, R. M. Solovay, J. van Heijenoort, Oxford University Press, New York, 1986.
12. LORENZ HALBEISEN, ***Combinatorial Set Theory***, Springer Monographs in Mathematics, Springer, London, 2012, with a gentle introduction to forcing.

13. LEON HENKIN, *The completeness of the first-order functional calculus*, ***The Journal of Symbolic Logic***, vol. 14 (1949), 159–166.
14. ______, *The discovery of my completeness proofs*, ***The Bulletin of Symbolic Logic***, vol. 2 (1996), 127–158.
15. HANS HERMES, ***Einführung in die mathematische Logik. Klassische Prädikatenlogik***, Mathematische Leitfäden, B. G. Teubner Verlagsgesellschaft, Stuttgart, 1963.
16. JOSEPH R. SHOENFIELD, ***Mathematical Logic***, Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1967.
17. JEAN VAN HEIJENOORT, ***From Frege to Gödel. A Source Book in Mathematical Logic, 1879–1931***, [Source Books in the History of Science], Harvard University Press, Cambridge, Massachusetts, 1967.
18. ERNST ZERMELO, ***Collected Works / Gesammelte Werke***, *Volume I: Set Theory, Miscellanea / Band I: Mengenlehre, Varia*, [Schriften der Mathematisch-naturwissenschaftlichen Klasse der Heidelberger Akademie der Wissenschaften, Nr. 21 (2010)], edited by Heinz-Dieter Ebbinghaus, Craig G. Fraser, and Akihiro Kanamori, Springer-Verlag, Berlin · Heidelberg, 2010.