Mathematical Logic, an Introduction

by Peter Koepke

Bonn, Summer 2018

Wann sollte die Mathematik je zu einem Anfang gelangen, wenn sie warten wollte, bis die Philosophie über unsere Grundbegriffe zur Klarheit und Einmüthigkeit gekommen ist? Unsere einzige Rettung ist der formalistische Standpunkt, undefinirte Begriffe (wie Zahl, Punkt, Ding, Menge) an die Spitze zu stellen, um deren actuelle oder psychologische oder anschauliche Bedeutung wir uns nicht kümmern, und ebenso unbewiesene Sätze (Axiome), deren actuelle Richtigkeit uns nichts angeht. Aus diesen primitiven Begriffen und Urtheilen gewinnen wir durch Definition und Deduction andere, und nur diese Ableitung ist unser Werk und Ziel. (Felix Hausdorff, 12. Januar 1918)

1 Introduction

Mathematics models real world phenomena like space, time, number, probability, games, etc. It proceeds from initial assumptions to conclusions solely by the application of rigorous arguments. Its results are "universal" and "logically valid", in that they do not depend on external or implicit conditions which may change with time, nature or society.

It is remarkable that mathematics is also able to *model itself*: mathematical logic defines exactly what mathematical statements and rigorous arguments are. The mathematical enquiry into the mathematical method leads to deep insights into mathematics, applications to classical field of mathematics, and to new mathematical theories. The study of mathematical language has also influenced the theory of formal and natural languages in computer science, linguistics and philosophy.

(Pure) mathematics is a formal science. The formal character of mathematical statements and arguments is the basis for the self-modelling of mathematics in mathematical logic. We sketch some aspects of mathematical logic in the following subsections.

1.1 A simple proof

We want to indicate that rigorous mathematical proofs can be generated by applying simple text manipulations to mathematical statements. Let us consider a fragment of the elementary theory of functions which expresses that the composition of two surjective maps is surjective as well:

> Let f and g be *surjective*, i.e., for all y there is x such that $y = f(x)$, and for all y there is x such that $y = g(x)$.
> *Theorem.* $g \circ f$ is surjective, i.e., for all y there is x such that $y = g(f(x))$.
> *Proof*. Consider any y. Choose z such that $y = g(z)$. Choose x such that $z = f(x)$. Then $y = g(f(x))$. Thus there is x such that $y = g(f(x))$. Thus for all y there is x such that $y = g(f(x))$.
> *Qed.*

These statements and arguments are expressed in an austere and systematic language, which can be further normalized. Logical symbols like $\forall$ and $\exists$ abbreviate language phrases like “for all” or “there exists”:

Let $\forall y \exists x\, y = f(x)$.
Let $\forall y \exists x\, y = g(x)$.
Theorem. $\forall y \exists x\, y = g(f(x))$.
Proof. Consider y.
$\exists x\, y = g(x)$.
Take z such that $y = g(z)$.
$\exists x\, z = f(x)$.
Take x such that $z = f(x)$.
$y = g(f(x))$.
Thus $\exists x\, y = g(f(x))$.
Thus $\forall y \exists x\, y = g(f(x))$.
Qed.

These lines can be considered as formal sequences of symbols. Certain sequences of symbols are acceptable as mathematical formulas, others like „let“, „take“ or „thus“ serve to structure the formal text. There are rules for the formation of formulas which are acceptable in a proof. These rules have a purely formal character and they can be applied irrespectively of some intuitive “meaning” of the symbols and formulas.

1.2 Formal proofs

In the example, $\exists x\, y = g(f(x))$ is inferred from $y = g(f(x))$. The rule of *existential quantification*: “put $\exists x$ in front of a formula” can usually be applied. It has the character of a left-multiplication by $\exists x$.

$$\exists x\,, \varphi \mapsto \exists x\, \varphi.$$

Logical rules satisfy certain laws which are similar to algebraic laws like associativity. Another interesting operation is *substitution*: From $y = g(z)$ and $z = f(x)$ infer $y = g(f(x))$ by a “find-and-replace”-substitution of z by $f(x)$.

Given a sufficient collection of rules, the above sequence of formulas, involving “keywords” like “let” and “thus” is a *deduction* or *derivation* in which every line is generated from earlier ones by syntactical rules. Mathematical results may be provable simply by the application of formal rules. In analogy with the formal rules of the infinitesimal “calculus” one calls a system of rules a *calculus*.

1.3 Syntax and semantics

Obviously we do not just want to describe a formal derivation as a kind of domino but we want to *interpret* the occuring symbols as mathematical objects. Thus we let variables x, $y, \ldots$ range over some domain like the real numbers $\mathbb{R}$ and let f and g stand for functions F, G: $\mathbb{R} \to \mathbb{R}$. Observe that the symbol or “name” f is not identical to the function F, and indeed f might also be interpretated as another function F'. To emphasize the distinction between names and objects, we classify symbols, formulas and derivations as *syntax* whereas the interpretations of symbols belong to the realm of *semantics*.

By interpreting x, y, ... and f, g, ... in a structure like ($\mathbb{R}$, F, G) we can define straightforwardly whether a formula like $\exists x\, g(f(x))$ is *satisfied* in the structure. A formula is *logically valid* if it is satisfied under *all* interpretations. The fundamental theorem of mathematical logic and the central result of this course is Gödel's completeness theorem:

Theorem. *There is a calculus with finitely many rules such that a formula is derivable in the calculus iff it is logically valid.*

1.4 Object theory and meta theory

We shall use the common, *informal* mathematical language to express properties of a *formal* mathematical language. The formal language forms the *object theory* of our studies, the informal mathematical language is the “higher” or *meta theory* of mathematical logic. There will be strong parallels between object and meta theory which say that the modelling is faithful.

1.5 Set theory

In modern mathematics notions can usually be reduced to set theory: non-negative integers correspond to cardinalities of finite sets, integers can be obtained via pairs of non-negative integers, rational numbers via pairs of integers, and real numbers via subsets of the rationals, etc. Geometric notions can be defined from real numbers using analytic geometry: a point is a pair of real numbers, a line is a set of points, etc. It is remarkable that the basic set theoretical axioms can be formulated in the logical language indicated above. So mathematics may be understood abstractly as

$$\text{Mathematics} = \text{(first-order) logic} + \text{set theory}.$$

Note that we only propose this as a reasonable abstract viewpoint corresponding to the logical analysis of mathematics. This perspective leaves out many important aspects like the applicability, intuitiveness and beauty of mathematics.

1.6 Set theory as meta theory

Our meta theory will be informal using the common notions of set, function, relation, natural, rational, real numbers, finite, infinite etc. We shall work informally but with a view towards possible formalization. Since we shall be interested in a weak theory able to carry out logical syntax, we shall attempt to

not use the existence of infinite sets in *syntactical considerations*.

On the other hand the standard semantics requires infinite structures like the sets $\mathbb{N}$ or $\mathbb{R}$ of numbers and we shall

use the existence of infinite sets in *semantical considerations*.

There will be exceptions to these rules where syntax and semantics merge like in the construction of structures out of infinite *sets* of terms.

In set theory, one usually distinguishes between *sets* and *classes*. A class is a collection $\{x|\ \varphi(x)\}$ of objects x which satisfy some property φ. The axioms of set theory determine which of these classes are sets, i.e., can be taken as objects themselves.

1.7 Circularity

We shall use *sets* as symbols which can then be used to formulate the axioms of *set* theory. We shall *prove* theorems about *proofs*. This kind of circularity seems to be unavoidable in comprehensive foundational science: linguistics has to *talk* about *language*, *brain research* has to be carried out by brains. Circularity can lead to paradoxes like the liar’s paradox: “I am a liar”, or “this sentence is false”. Circularity poses many problems and seems to undermine the value of foundational theories. We suggest that the reader takes a *naive* standpoint in these matters: there are sets and proofs which are just as obvious as natural numbers. Then theories are formed which abstractly describe the naive objects.

A closer analysis of circularity in logic leads to the famous *incompleteness theorems* of Gödel:

Theorem. *Formal theories which are strong enough to "formalize themselves" are not complete, i.e., there are statements such that neither it nor its negation can be proved in that theory. Moreover such theories cannot prove their own consistency.*

These results, besides their initial mathematical meaning, had a tremendous impact on the theory of knowledge outside mathematics, e.g., in philosophy, psychology, linguistics.

2 The Syntax of first-order logic: Symbols, terms, and formulas

> *The art of free society consists first in the maintenance of the symbolic code.*
> A. N. Whitehead

Formal mathematical statements will be finite sequences of symbols, just like ordinary sentences are sequences of alphabetic letters. These sequences can be studied mathematically. We shall treat sequences as mathematical objects, similar to numbers or vectors.

The study of the formal properties of symbols, words, sentence,... is called *syntax*. Syntax will later be related to the "meaning" of symbolic material, its *semantics*. The interplay between syntax and semantics is at the core of logic. A strong logic is able to present interesting semantic properties, i.e., properties of interesting mathematical structure, already in its syntax.

We build the formal language with formulas like $\forall y \exists x\, y = g(f(x))$ recursively from atomic building blocks.

2.1 Symbols

> *Man muß jederzeit an Stelle von 'Punkte, Geraden, Ebenen', 'Tische, Stühle, Bierseidel' sagen können".*
> Quote ascribed to David Hilbert

A symbol has some basic information about its role within larger contexts like words and sentences. E.g., the symbol $\leqslant$ is usually used to stand for a *binary relation*. So we let symbols include information on its function, like denoting a "relation", together with further details, like "binary".

Definition 1. *The* basic symbols *of first-order logic are*

a) $\equiv$ *for* equality*,*

b) $\neg, \rightarrow, \bot$ *for the logical operations of* negation*,* implication *and the truth value* false*,*

c) $\forall$ *for* universal quantification*,*

d) (*and*) *for auxiliary* bracketing*.*

e) variables v_n *for* $n \in \mathbb{N}$*.*

Let $\mathrm{Var} = \{v_n | n \in \mathbb{N}\}$ *be the class of variables and let* S_0 *be the class of basic symbols.*

There is a sufficiently rich class of relation symbols*. Every relation symbol R possesses an arity which is a natural number. 1-ary relation symbols are called* unary*, 2-ary relation symbols are called* binary*. A 0-ary relation symbol is also called a* propositional constant (symbol)*.*

Moreover there is a sufficiently rich class of function symbols*. Every function symbol f possesses an arity which is a natural number. A 0-ary function symbol is also called a* constant (symbol)l*.*

We assume that the basic symbols, the relation symbols, and the function symbols are all pairwise distinct.

A symbol class *or a* language *is a class of relation symbols and function symbols.*

An n-ary relation symbol is intended to denote an n-ary relation; an n-ary function symbol is intended to denote an n-ary function in some structure. A symbol class is also called a *type* because it describes the type of structures which will later interpret the symbols. We shall denote variables by letters like x, y, z, ..., relation symbols by P, Q, R,..., functions symbols by $f, g, h, \ldots$ and constant symbols by $c, c_0, c_1, \ldots$ We shall also use other typographical symbols in line with standard mathematical practice. A symbol like $<$, e.g., usually denotes a binary relation, and we could assume for definiteness that there is some fixed formalization of $<$ like $< = (1, 999, 2)$, where 1 indicates a relation (symbol), 999 is the "name" of the symbol, and 2 is its arity. Instead of the arbitrary 999 one could also take the number of $<$ in some typographical coding system like unicode; there $<$ is coded by the decimal number 60 and we could set $< = (1, 60, 2)$.

Example 2. The *language of group theory* is the language

$$S_{\mathrm{Gr}} = \{\circ, e\},$$

where $\circ$ is a binary function symbol and e is a constant (symbol). Again one could be definite about the coding of symbols and set $S_{\mathrm{Gr}} = \{(2, 9900, 2), (2, 101, 0)\}$, following unicode, but we shall not care about such detail. As usual in algebra, one also uses an *extended language of group theory*

$$S_{\mathrm{Gr}'} = \{\circ, {}^{-1}, e\}$$

to describe groups, where ${}^{-1}$ is a *unary* function symbol (for forming inverses).

2.2 Words

Words:
A letter and a letter on a string
Will hold forever humanity spell-bound
The Real Group

Definition 3. *Let S be a language. A* word *over S is a finite sequence*

$$w = s_0 s_1 \ldots s_{n-1}$$

where each s_i is an element of $S_0 \cup S$. The number n is called the length *of w:* $\mathrm{length}(w) = n$ *. The empty sequence $\emptyset$ is also called the* empty word*. Let S^* be the class of all words over S.*

Definition 4. *If $w = s_0 s_1 \ldots s_{m-1}$ and $w' = s_0' s_1' \ldots s_{n-1}'$ are words over S then*

$$w^\frown w' = s_0 s_1 \ldots s_{m-1} s_0' s_1' \ldots s_{n-1}'$$

is the concatenation *of w and w'. We also write $w w'$ instead of $w^\frown w'$.*

Exercise 1. The operation of concatenation satisfies some canonical laws:

a) $^\frown$ is associative: $(ww')w''=w(w'w'')$.

b) $\emptyset$ is a neutral element for $^\frown$: $\emptyset w=w\emptyset=w$.

c) $^\frown$ satisfies cancelation: if $uw=u'w$ then $u=u'$; if $wu=wu'$ then $u=u'$.

2.3 Terms

Fix a language S.

Definition 5. *The class T^S of all S-*terms *is the smallest subclass of S^* such that*

a) $x\in T^S$ for all variables x;

b) $ft_0...t_{n-1}\in T^S$ for all $n\in\mathbb{N}$, all n-ary function symbols $f\in S$, and all t_0, ..., $t_{n-1}\in T^S$.

Terms are written in *Polish* notation, meaning that function symbols come first and that no brackets are needed. Indeed, terms in T^S have *unique readings* according to the following

Lemma 6. *For every term $t\in T^S$ exactly one of the following holds:*

a) t is a variable;

b) there is a uniquely defined function symbol $f\in S$ and a uniquely defined sequence $t_0,...,t_{n-1}\in T^S$ of terms, where f is n-ary, such that $t=ft_0...t_{n-1}$.

Proof. Exercise. □

Remark 7. Unique readability is essential for working with terms. Therefore if this Lemma would not hold one would have to alter the definition of terms.

Example 8. For the language $S_{\mathrm{Gr}}=\{\circ,e\}$ of group theory, terms in $T^{S_{\mathrm{Gr}}}$ look like

$$e,v_0,v_1,...,\circ ee,\circ ev_m,\circ v_m e,\circ ee,\circ e\circ ee,...,\circ v_i\circ v_j v_k,\circ\circ v_i v_j v_k,....$$

In standard notation we would write $\circ v_i\ \circ v_j\ v_k$ as $(v_i\circ(v_j\circ v_k))$ and $\circ\circ v_i\ v_j\ v_k$ as $=((v_i\circ v_j)\circ v_k)$. Later, if the operation $\circ$ should be seen to be associative, one might "leave out" some brackets.

Exercise 2. Show that every term $t\in T^{S_{\mathrm{Gr}}}$ has odd length $2n+1$ where n is the number of $\circ$-symbols in t.

2.4 Formulas

Definition 9. *The class L^S of all S-*formulas *is the smallest subclass of S^* such that*

a) $\bot\in L^S$ (the false formula);

b) $t_0\equiv t_1\in L^S$ for all S-terms $t_0,t_1\in T^S$ (equality);

c) $Rt_0...t_{n-1}\in L^S$ for all n-ary relation symbols $R\in S$ and all S-terms $t_0,...,t_{n-1}\in T^S$ (relational formula);

d) $\neg\varphi\in L^S$ for all $\varphi\in L^S$ (negation);

e) $(\varphi\rightarrow\psi)\in L^S$ for all $\varphi,\psi\in L^S$ (implication);

f) $\forall x\varphi\in L^S$ for all $\varphi\in L^S$ and all variables x (universalisation).

L^S is also called the first-order language *for the symbol class S. Formulas produced by conditions a) - c) only are called* atomic formulas *since they constitute the initial steps of the formula calculus.*

We restrict L^S to just the logical connectives $\neg$ and $\rightarrow$, and the quantifier $\forall$. The next definition introduces other connectives and quantifiers as convenient abbreviations for formulas in L^S. For theoretical considerations it is however advantageous to work with a "small" language.

Definition 10. *For S-formulas φ and ψ and a variable x write*

- $\top$ *("true") instead of* $\neg\bot$ *;*
- $(\varphi \vee \psi)$ *("φ or ψ") instead of* $(\neg\varphi \rightarrow \psi)$ *is the* disjunction *of* φ, ψ *;*
- $(\varphi \wedge \psi)$ *("φ and ψ") instead of* $\neg(\varphi \rightarrow \neg\psi)$ *is the* conjunction *of* φ, ψ *;*
- $(\varphi \leftrightarrow \psi)$ *("φ iff ψ") instead of* $((\varphi \rightarrow \psi) \wedge (\psi \rightarrow \varphi))$ *is the* equivalence *of* φ, ψ *;*
- $\exists x \varphi$ *("for all x holds φ") instead of* $\neg\forall x \neg\varphi$ *is an* existential quantification.

For the sake of simplicity one often omits redundant brackets, in particular outer brackets. So we usually write $\varphi \vee \psi$ instead of $(\varphi \vee \psi)$.

Exercise 3. Formulate and prove the unique readability of formulas in L^S.

Exercise 4. Formulate the standard axioms of group theory in $L^{S_{\mathrm{Gr}}}$.

3 Implementations of first-order syntax

We have defined the syntactic notions in informal mathematical language. To be more formal, one could formalize those notions in some foundational mathematical theory. We shall consider formalizations in set theory and in some programming language.

3.1 Formalization in set theory without infinity

Set theory the widely accepted foundation of mathematics. Hence the syntactical notions introduced so far should be formalizable in set theory. Since the notions of symbol and formula are finitary, they should not require the axiom of infinity, which is equivalent to the existence of infinite sets. So we shall work in standard set theory *without* infinity. We shall later present set theory axiomatically. For the moment we work in the usual "naive" set theory, without assuming the existence of infinite sets.

We introduce some basic notions of set theory. The term

$$\{x \mid A(x)\}$$

denotes the *class* of all sets x which satisfy the property A. In general we can not and do not require $\{x \mid A(x)\}$ to be a set (which can then be used as an element of another class). Some (finite) classes will always be sets. $V = \{x \mid x = x\}$ is the class of all sets or the set theoretical *universe* and $\emptyset = \{x \mid x \neq x\}$ is the *empty set*. We can form pairs of elements as the set

$$\{x, y\} = \{z \mid z = x \text{ or } z = y\}.$$

Ordered pairs and triples can be formalized as

$$\begin{aligned} (x, y) &= \{\{x\}, \{x, y\}\}, \text{ with } \{x\} = \{x, x\} \\ (x, y, z) &= ((x, y), z) \end{aligned}$$

The natural numbers can be defined, without assuming infinite sets, as a class $\mathbb{N}$ which contains the recursively defined numbers

$$\begin{aligned} 0 &= \emptyset \\ 1 &= \{0\} \\ 2 &= \{0,1\} \\ 3 &= \{0,1,2\} \\ \ldots &= \ldots \end{aligned}$$

The *cartesian product* of A and B is

$$A \times B = \{(x,y) | x \in A \text{ and } y \in B\}.$$

The *cartesian powers* of A are defined recursively as

$$\begin{aligned} A^0 &= \{0\} \\ A^1 &= A \\ A^{n+2} &= A^{n+1} \times A \end{aligned}$$

An n-ary relation on A is a subclass $R \subseteq A^n$. An n-ary function on A is a function f: $A^n \rightarrow A$.

We can now begin to embed syntax into set theory by defining symbols to be certain fixed sets.

Definition 11. *Set*

a) $\equiv = 0$,

b) $\neg = 1$,

c) $\rightarrow = 2$,

d) $\bot = 3$,

e) $\forall = 4$,

f) $(= 5$,

g) $) = 6$,

h) $v_n = (0, n, 0)$,

i) an n-ary relation symbol is a set R of the form $R = (1, x, n)$, where x is some set,

j) an n-ary function symbol is a set f of the form $f = (2, x, n)$, where x is some set.

This provides us with sufficiently many pairwise distinct relation symbols and function symbols.

Words are sequences of symbols, and these can be formalized set-theoretically as functions from natural numbers to symbols.

Definition 12. *Let S be a language. A word over S is a function*

$$w : n \rightarrow S_0 \cup S$$

for some number $n \in \mathbb{N}$ which is the length of w. Note that $n = \{0, \ldots, n-1\}$ so that

$$w : \{0, \ldots, n-1\} \rightarrow S_0 \cup S.$$

We denote w also by $w_0 \ldots w_{n-1}$. For finite sequences $w = w_0 \ldots w_{m-1}$ and $w' = w'_0 \ldots w'_{n-1}$ define the concatenation $w^\frown w' = w_0 \ldots w_{m-1} w'_0 \ldots w'_{n-1}$ of w and w' by

$$w^\frown w' : m + n \rightarrow S_0 \cup S$$

and

$$w\hat{\ }w'(i) = \begin{cases} w(i), \ \textit{if } i < m; \\ w'(i-m), \ \textit{if } i \geqslant m. \end{cases}$$

These formalizations will also allow the further development to be carried out in set theory.

3.2 Implementations in programming languages

Finite sequences of symbols can be handled efficiently by computers and programming languages. Indeed one can argue that finite sequences of symbols are the prime data type for computers. We quote some code from the implementation of first-order logic in OCaml by John Harrison, where the syntactic categories are defined as inductive data types.

```
type term = Var of string
          | Fn of string * term list;;

(* ------------------------------------- *)
(* Example.
     *)
(* ------------------------------------- *)

START_INTERACTIVE;;
Fn("sqrt",[Fn("-",[Fn("1",[]);
                   Fn("cos",[Fn("power",[Fn("+",[Var "x"; Var "y"]);
                                        Fn("2",[])])])])]);;
END_INTERACTIVE;;
```

Any string can become a variable by prefixing it with the “constructor” `Var`: `Var "x"`. Any string can become a function symbol by using the constructor `Fn`: `Fn("+",[Var "x"; Var "y"])` turns the string `"+"` into a function symbol; its argument are contained in the 2-element list `[Var "x"; Var "y"]`; this makes `"+"` a binary symbol.

Atomic relational formulas are implemented as

```
type fol = R of string * term list;;
```

So any string can also become a relation symbol. Note that there is no distinguished equality symbol in Harrison’s approach; one could use the string `"="` for it.

General first-order formulas with atomic formulas of type `'a` are defined as

```
type ('a)formula = False
                 | True
                 | Atom of 'a
                 | Not of ('a)formula
                 | And of ('a)formula * ('a)formula
                 | Or of ('a)formula * ('a)formula
                 | Imp of ('a)formula * ('a)formula
                 | Iff of ('a)formula * ('a)formula
                 | Forall of string * ('a)formula
                 | Exists of string * ('a)formula;;
```

Then `fol formula` is the type of first order formulas.

Harrison also introduces parsing and printing routines that improve the readability of input and output formulas, like in

```
<<(forall x y.  exists z.  forall w.  P(x) /\ Q(y) ==>R(z) /\ U(w))
   ==> (exists x y.  P(x) /\ Q(y)) ==> (exists z.  R(z))>>;;
```

Note that in implementation of first-order logic in a computer language amounts to the definition of a formal language within an other formal language. A computer language can have some formal semantics within some abstract mathematical domain, or we can let it have a concrete semantics in terms of steering a concrete electronic device like a PC.

4 Semantics

We shall *interpret* formulas like $\forall y \exists x\, y = g(f(x))$ in adequate *structures*. The interaction between language and structures is usually called *semantics*. Technically it will consist in mapping all syntactic material to semantic material centered around structures. We shall obtain a mapping schema like:

$\forall$	domain A of a structure $\mathfrak{A}$
variable	element of A
function symbol	function on A
relation symbol	relation on A
term	element of A
formula	truth value
...	...

Fix a symbol class S.

Definition 13. *An S-*structure *is a function $\mathfrak{A}$ defined on $\{\forall\} \cup S$ such that*

a) $\mathfrak{A}(\forall)$ is a nonempty set; $\mathfrak{A}(\forall)$ is called the underlying set *or the* domain *of $\mathfrak{A}$ and is often denoted by A or $|\mathfrak{A}|$;*

b) for every n-ary relation symbol $R \in S$, $\mathfrak{A}(R)$ is an n-ary relation *on A, i.e., $\mathfrak{A}(R) \subseteq A^n$;*

c) for every n-ary function symbol $f \in S$, $\mathfrak{A}(f)$ is an n-ary function *on A, i.e., $\mathfrak{A}(f)\colon A^n \to A$.*

Again we use customary and convenient notations for the *components* of the structure $\mathfrak{A}$, i.e., the values of $\mathfrak{A}$. One often writes $R^{\mathfrak{A}}$, $f^{\mathfrak{A}}$, or $c^{\mathfrak{A}}$ instead of $\mathfrak{A}(r)$, $\mathfrak{A}(f)$, or $\mathfrak{A}(c)$ resp. In simple cases, one may simply list the components of the structure. If, e.g., when $S = \{R_0, R_1, f\}$ we may write

$$\mathfrak{A} = (A, R_0^{\mathfrak{A}}, R_1^{\mathfrak{A}}, f^{\mathfrak{A}})$$

or "$\mathfrak{A}$ has domain A with relations $R_0^{\mathfrak{A}}, R_1^{\mathfrak{A}}$ and an operation $f^{\mathfrak{A}}$".

A constant symbol $c \in S$ is interpreted by a 0-ary function $\mathfrak{A}(c)\colon A^0 = \{0\} \to A$ which is defined for the single argument 0 and takes a single value $\mathfrak{A}(c)(0)$ in A. It is natural to identify the function $\mathfrak{A}(c)$ with the constant value $\mathfrak{A}(c)(0)$ and obtain $\mathfrak{A}(c) \in A$.

One often uses the same notation for a structure and its underlying set like in

$$A = (A, R_0^{\mathfrak{A}}, R_1^{\mathfrak{A}}, f^{\mathfrak{A}}).$$

This "overloading" of notation is common in mathematics (and in natural language). Usually a human reader is readily able to detect and "disambiguate" ambiguities introduced by multiple usage. There are techniques in computer science to deal with overloading, e.g., by *typing* of notions. Another common overloading is given by a naive identification of syntax and semantics, i.e., by writing

$$A = (A, R_0, R_1, f) \text{ instead of } A = (A, R_0^{\mathfrak{A}}, R_1^{\mathfrak{A}}, f^{\mathfrak{A}})$$

Since we are particularly interested in the interplay of syntax and semantics we shall try to avoid this particular kind of overloading.

Example 14. Formalize the *ordered field of reals* $\mathbb{R}$ as follows. Define the language of ordered fields

$$S_{\mathrm{OF}} = \{<, +, \cdot, 0, 1\}.$$

Then define the S_{OF}-structure $\mathbb{R}$ by

$$\begin{aligned}
\mathbb{R}(\forall) &= \mathbb{R}\\
\mathbb{R}(<) = <^{\mathbb{R}} &= \{(u,v) \in \mathbb{R}^2 \,|\, u < v\}\\
\mathbb{R}(+) = +^{\mathbb{R}} &= \{(u,v,w) \in \mathbb{R}^3 \,|\, u + v = w\}\\
\mathbb{R}(\cdot) = \cdot^{\mathbb{R}} &= \{(u,v,w) \in \mathbb{R}^3 \,|\, u \cdot v = w\}\\
\mathbb{R}(0) = 0^{\mathbb{R}} &= 0 \in \mathbb{R}\\
\mathbb{R}(1) = 1^{\mathbb{R}} &= 1 \in \mathbb{R}
\end{aligned}$$

This defines the standard structure $\mathbb{R} = (\mathbb{R}, <^{\mathbb{R}}, +^{\mathbb{R}}, \cdot^{\mathbb{R}}, 0^{\mathbb{R}}, 1^{\mathbb{R}})$.

Observe that the symbols could *in principle* be interpreted in completely different, even counterintuitive ways like

$$\begin{aligned}
\mathbb{R}'(\forall) &= \mathbb{N}\\
\mathbb{R}'(<) &= \{(u,v) \in \mathbb{N}^2 \,|\, u > v\}\\
\mathbb{R}'(+) &= \{(u,v,w) \in \mathbb{N}^3 \,|\, u \cdot v = w\}\\
\mathbb{R}'(\cdot) &= \{(u,v,w) \in \mathbb{N}^3 \,|\, u + v = w\}\\
\mathbb{R}'(0) &= 1\\
\mathbb{R}'(1) &= 0
\end{aligned}$$

Example 15. Define the language of *Boolean algebras* by

$$S_{\mathrm{BA}} = \{\wedge, \vee, -, 0, 1\}$$

where $\wedge$ and $\vee$ are binary function symbols for "and" and "or", $-$ is a unary function symbol for "not", and 0 and 1 are constant symbols. A Boolean algebra of particular importance in logic is the algebra $\mathbb{B}$ of *truth values*. Let $B = |\mathbb{B}| = \{\mathbb{F}, \mathbb{T}\}$ with $\mathbb{F} = \mathbb{B}(0)$ and $\mathbb{T} = \mathbb{B}(1)$. Define the operations and $= \mathbb{B}(\wedge)$, or $= \mathbb{B}(\vee)$, and not $= \mathbb{B}(-)$ by *operation tables* in analogy with standard multiplication tables:

and	$\mathbb{F}$	$\mathbb{T}$
$\mathbb{F}$	$\mathbb{F}$	$\mathbb{F}$
$\mathbb{T}$	$\mathbb{F}$	$\mathbb{T}$

,

or	$\mathbb{F}$	$\mathbb{T}$
$\mathbb{F}$	$\mathbb{F}$	$\mathbb{T}$
$\mathbb{T}$	$\mathbb{T}$	$\mathbb{T}$

, and

not	
$\mathbb{F}$	$\mathbb{T}$
$\mathbb{T}$	$\mathbb{F}$

.

Note that we use the non-exclusive "or" instead of the exclusive "either - or".

Exercise 5. Show that every *truth-function* $F\colon B^n \to B$ can be obtained as a composition of the functions *and* and *not*.

The notion of structure leads to derived definitions.

Definition 16. *Let $\mathfrak{A}$ be an S-structure and $\mathfrak{A}'$ be an S'-structure. Then $\mathfrak{A}$ is a* reduct *of $\mathfrak{A}'$, or $\mathfrak{A}'$ is an* expansion *of $\mathfrak{A}$, if $S \subseteq S'$ and $\mathfrak{A}' \upharpoonright (\{\forall\} \cup S) = \mathfrak{A}$.*

According to this definition, the additive group $(\mathbb{R}, +, 0)$ of reals is a reduct of the field $(\mathbb{R}, +, \cdot, 0, 1)$.

Definition 17. *Let $\mathfrak{A}, \mathfrak{B}$ be S-structures. Then $\mathfrak{A}$ is a* substructure *of $\mathfrak{B}$, $\mathfrak{A} \subseteq \mathfrak{B}$, if $\mathfrak{B}$ is a pointwise extension of $\mathfrak{A}$, i.e.,*

a) $A = |\mathfrak{A}| \subseteq |\mathfrak{B}|$;

b) for every n-ary relation symbol $R \in S$ holds $R^{\mathfrak{A}} = R^{\mathfrak{B}} \cap A^n$;

c) for every n-ary function symbol $f \in S$ holds $f^{\mathfrak{A}} = f^{\mathfrak{B}} \upharpoonright A^n$.

Definition 18. *Let $\mathfrak{A}, \mathfrak{B}$ be S-structures and $h\colon |\mathfrak{A}| \to |\mathfrak{B}|$. Then h is a* homomorphism *from $\mathfrak{A}$ into $\mathfrak{B}$, $h\colon \mathfrak{A} \to \mathfrak{B}$, if*

a) for every n-ary relation symbol $R \in S$ and for every $a_0, \ldots, a_{n-1} \in A$

$$R^{\mathfrak{A}}(a_0, \ldots, a_{n-1}) \textit{ implies } R^{\mathfrak{B}}(h(a_0), \ldots, h(a_{n-1}));$$

b) for every n-ary function symbol $f \in S$ and for every $a_0, \ldots, a_{n-1} \in A$

$$f^{\mathfrak{B}}(h(a_0), \ldots, h(a_{n-1})) = h(f^{\mathfrak{A}}(a_0, \ldots, a_{n-1})).$$

h is an embedding *of $\mathfrak{A}$ into $\mathfrak{B}$, $h\colon \mathfrak{A} \hookrightarrow \mathfrak{B}$, if moreover*

a) h is injective;

b) for every n-ary relation symbol $R \in S$ and for every $a_0, \ldots, a_{n-1} \in A$

$$R^{\mathfrak{A}}(a_0, \ldots, a_{n-1}) \textit{ iff } R^{\mathfrak{B}}(h(a_0), \ldots, h(a_{n-1})).$$

If h is also bijective, it is called an isomorphism.

5 The satisfaction relation

> *"What is truth?" Pilate asked.*
> John 18:38

An S-structure interprets the symbols in S. To interpret a formula in a structure, one also has to interpret the (occuring) variables.

Definition 19. *Let S be a language. An S-*model *is a function*

$$\mathfrak{M}\colon \{\forall\} \cup S \cup \mathrm{Var} \to V$$

such that $\mathfrak{M} \upharpoonright \{\forall\} \cup S$ is an S-structure and for all $n \in \mathbb{N}$ holds $\mathfrak{M}(v_n) \in |\mathfrak{M}|$. $\mathfrak{M}(v_n)$ is the interpretation or valuation *of the variable v_n in $\mathfrak{M}$.*

It will be important to modify a model $\mathfrak{M}$ at specific variables. For pairwise distinct variables $x_0, \ldots, x_{r-1}$ and $a_0, \ldots, a_{r-1} \in |\mathfrak{M}|$ define

$$\mathfrak{M}\frac{a_0 \ldots a_{r-1}}{x_0 \ldots x_{r-1}} = (\mathfrak{M} \setminus \{(x_0, \mathfrak{A}(x_0)), \ldots, (x_{r-1}, \mathfrak{A}(x_{r-1}))\}) \cup \{(x_0, a_0), \ldots, (x_{r-1}, a_{r-1})\}.$$

We now define the *semantics* of first-order languages by interpreting terms and formulas in models.

Definition 20. *Let $\mathfrak{M}$ be an S-model. Define the* interpretation $\mathfrak{M}(t) \in |\mathfrak{M}|$ *of a term $t \in T^S$ by recursion on the term calculus:*

a) for t a variable, $\mathfrak{M}(t)$ is already defined;

b) for an n-ary function symbol and terms $t_0, ..., t_{n-1} \in T^S$, let

$$\mathfrak{M}(ft_0....t_{n-1}) = f^{\mathfrak{A}}(\mathfrak{M}(t_0), ..., \mathfrak{M}(t_{n-1})).$$

This explains the interpretation of a term like $v_3^2 + v_{200}^3$ in the reals.

Definition 21. *Let $\mathfrak{M}$ be an S-model. Define the* interpretation $\mathfrak{M}(\varphi) \in \mathbb{B}$ *of a formula $\varphi \in L^S$, where $\mathbb{B} = \{\mathbb{F}, \mathbb{T}\}$ is the Boolean algebra of truth values, by recursion on the formula calculus:*

a) $\mathfrak{M}(\bot) = \mathbb{F}$;

b) for terms $t_0, t_1 \in T^S$: $\mathfrak{M}(t_0 \equiv t_1) = \mathbb{T}$ iff $\mathfrak{M}(t_0) = \mathfrak{M}(t_1)$;

c) for every n-ary relation symbol $R \in S$ and terms $t_0, ..., t_1 \in T^S$

$$\mathfrak{M}(Rt_0...t_{n-1}) = \mathbb{T} \textit{ iff } R^{\mathfrak{M}}(\mathfrak{M}(t_0), ..., \mathfrak{M}(t_{n-1}));$$

d) $\mathfrak{M}(\neg\varphi) = \mathbb{T}$ iff $\mathfrak{M}(\varphi) = \mathbb{F}$;

e) $\mathfrak{M}(\varphi \to \psi) = \mathbb{T}$ iff $\mathfrak{M}(\varphi) = \mathbb{T}$ implies $\mathfrak{M}(\psi) = \mathbb{T}$;

f) $\mathfrak{M}(\forall v_n \varphi) = \mathbb{T}$ iff for all $a \in |\mathfrak{M}|$ holds $\mathfrak{M}\frac{a}{v_n}(\varphi) = \mathbb{T}$.

We write $\mathfrak{M} \vDash \varphi$ instead of $\mathfrak{M}(\varphi) = \mathbb{T}$. We also say that $\mathfrak{M}$ satisfies *φ or that φ* holds *in $\mathfrak{M}$ or that φ is* true *in $\mathfrak{M}$. For $\Phi \subseteq L^S$ write $\mathfrak{M} \vDash \Phi$ iff $\mathfrak{M} \vDash \varphi$ for every $\varphi \in \Phi$.*

Definition 22. *Let S be a language and $\Phi \subseteq L^S$. Φ is* universally valid *if Φ holds in every S-model. Φ is* satisfiable *if there is an S-model $\mathfrak{M}$ such that $\mathfrak{M} \vDash \Phi$.*

The language extension by the (abbreviating) symbols $\vee, \wedge, \leftrightarrow, \exists$ is consistent with the expected meanings of the additional symbols:

Exercise 6. Prove:

a) $\mathfrak{M} \vDash (\varphi \vee \psi)$ iff $\mathfrak{M} \vDash \varphi$ *or* $\mathfrak{M} \vDash \psi$;

b) $\mathfrak{M} \vDash (\varphi \wedge \psi)$ iff $\mathfrak{M} \vDash \varphi$ *and* $\mathfrak{M} \vDash \psi$;

c) $\mathfrak{M} \vDash (\varphi \leftrightarrow \psi)$ iff $\mathfrak{M} \vDash \varphi$ *is equivalent to* $\mathfrak{M} \vDash \psi$;

d) $\mathfrak{M} \vDash \exists v_n \varphi$ iff *there exists* $a \in |\mathfrak{M}|$ such that $\mathfrak{M}\frac{a}{v_n} \vDash \varphi$.

With the notion of $\vDash$ we can now formally define what it means for a structure to be a group or for a function to be differentiable. Before considering examples we make some auxiliary definitions and simplifications.

It is intuitively obvious that the interpretation of a term only depends on the occuring variables, and that satisfaction for a formula only depends on its free, non-bound variables.

Definition 23. *For $t \in T^S$ define* $\mathrm{var}(t) \subseteq \{v_n | n \in \mathbb{N}\}$ *by recursion on (the lengths of) terms:*

– $\mathrm{var}(x) = \{x\}$*;*

- $\mathrm{var}(c) = \emptyset$;
- $\mathrm{var}(f t_0 \ldots t_{n-1}) = \bigcup_{i<n} \mathrm{var}(t_i)$.

Definition 24. *Für $\varphi \in L^S$ define the set of* free variables $\mathrm{free}(\varphi) \subseteq \{v_n | n \in \mathbb{N}\}$ *by recursion on (the lengths of) formulas:*

- $\mathrm{free}(t_0 \equiv t_1) = \mathrm{var}(t_0) \cup \mathrm{var}(t_1)$;
- $\mathrm{free}(R t_0 \ldots t_{n-1}) = \mathrm{var}(\, t_0) \cup \ldots \cup \mathrm{var}(t_{n-1})$;
- $\mathrm{free}(\neg\varphi) = \mathrm{free}(\varphi)$;
- $\mathrm{free}(\varphi \rightarrow \psi) = \mathrm{free}(\varphi) \cup \mathrm{free}(\psi)$.
- $\mathrm{free}(\forall x \varphi) = \mathrm{free}(\varphi) \setminus \{x\}$.

For $\Phi \subseteq L^S$ define the class $\mathrm{free}(\Phi)$ *of* free variables *as*

$$\mathrm{free}(\Phi) = \bigcup_{\varphi \in \Phi} \mathrm{free}(\varphi)\,.$$

Example 25.

$$\begin{aligned}
\mathrm{free}(Ryx \rightarrow \forall y \neg y = z) &= \mathrm{free}(Ryx) \cup \mathrm{free}(\forall y \neg y = z) \\
&= \mathrm{free}(Ryx) \cup (\mathrm{free}(\, \neg y = z) \setminus \{y\}) \\
&= \mathrm{free}(Ryx) \cup (\mathrm{free}(\, y = z) \setminus \{y\}) \\
&= \{y, x\} \cup (\{y, z\} \setminus \{y\}) \\
&= \{y, x\} \cup \{z\} \\
&= \{x, y, z\}.
\end{aligned}$$

Definition 26.

a) For $n \in \mathbb{N}$ let $L_n^S = \{\varphi \in L^S \mid \mathrm{free}(\varphi) \subseteq \{v_0, \ldots, v_{n-1}\}\}$.

*b) $\varphi \in L^S$ is an S-*sentence *if* $\mathrm{free}(\varphi) = \emptyset$; *$L_0^S$ is the class of S-sentences.*

Theorem 27. *Let t be an S-term and let $\mathfrak{M}$ and $\mathfrak{M}'$ be S-models with the same structure $\mathfrak{M} \upharpoonright \{\forall\} \cup S = \mathfrak{M}' \upharpoonright \{\forall\} \cup S$ and $\mathfrak{M} \upharpoonright \mathrm{var}(t) = \mathfrak{M}' \upharpoonright \mathrm{var}(t)$. Then $\mathfrak{M}(t) = \mathfrak{M}'(t)$.*

Theorem 28. *Let t be an S-term and let $\mathfrak{M}$ and $\mathfrak{M}'$ be S-models with the same structure $\mathfrak{M} \upharpoonright \{\forall\} \cup S = \mathfrak{M}' \upharpoonright \{\forall\} \cup S$ and $\mathfrak{M} \upharpoonright \mathrm{free}(\varphi) = \mathfrak{M}' \upharpoonright \mathrm{free}(\varphi)$. Then*

$$\mathfrak{M} \vDash \varphi \quad \textit{iff} \quad \mathfrak{M}' \vDash \varphi.$$

Proof. By induction on formulas.
$\varphi = t_0 \equiv t_1$: Then $\mathrm{var}(t_0) \cup \mathrm{var}(t_1) = \mathrm{free}(\varphi)$ and

$$\begin{aligned}
\mathfrak{M} \vDash \varphi \ &\text{iff}\ \ \mathfrak{M}(t_0) = \mathfrak{M}(t_1) \\
&\text{iff}\ \ \mathfrak{M}'(t_0) = \mathfrak{M}'(t_1) \ \text{ by the previous Theorem,} \\
&\text{iff}\ \ \mathfrak{M}' \vDash \varphi.
\end{aligned}$$

$\varphi = \psi \rightarrow \chi$ and assume the claim to be true for ψ and χ. Then

$$\begin{aligned}
\mathfrak{M} \vDash \varphi \ &\text{iff}\ \ \mathfrak{M} \vDash \psi \text{ implies } \mathfrak{M} \vDash \chi \\
&\text{iff}\ \ \mathfrak{M}' \vDash \psi \text{ implies } \mathfrak{M}' \vDash \chi \ \text{ by the inductive assumption,} \\
&\text{iff}\ \ \mathfrak{M}' \vDash \varphi.
\end{aligned}$$

$\varphi = \forall v_n \psi$ and assume the claim to be true for ψ. Then $\mathrm{free}(\psi) \subseteq \mathrm{free}(\varphi) \cup \{v_n\}$. For all $a \in A = |\mathfrak{M}|$: $\mathfrak{M}\frac{a}{v_n} \restriction \mathrm{free}(\psi) = \mathfrak{M}'\frac{a}{v_n} \restriction \mathrm{free}(\psi)$ and so

$$\begin{aligned}\mathfrak{M} \vDash \varphi \ \text{ iff } & \text{ for all } a \in A \text{ holds } \mathfrak{M}\frac{a}{v_n} \vDash \psi \\ \text{ iff } & \text{ for all } a \in A \text{ holds } \mathfrak{M}'\frac{a}{v_n} \vDash \psi \text{ by the inductive assumption,} \\ \text{ iff } & \mathfrak{M}' \vDash \varphi.\end{aligned}$$

□

This allows further simplifications in notations for $\vDash$:

Definition 29. *Let $\mathfrak{A}$ be an S-structure and let $(a_0, \ldots, a_{n-1})$ be a sequence of elements of A. Let t be an S-term with* $\mathrm{var}(t) \subseteq \{v_0, \ldots, v_{n-1}\}$. *Then define*

$$t^{\mathfrak{A}}[a_0, \ldots, a_{n-1}] = \mathfrak{M}(t),$$

where $\mathfrak{M} \supseteq \mathfrak{A}$ is some (or any) S-model with $\mathfrak{M}(v_0) = a_0, \ldots, \mathfrak{M}(v_{n-1}) = a_{n-1}$.
Let φ be an S-formula with $\mathrm{free}(\varphi) \subseteq \{v_0, \ldots, v_{n-1}\}$. *Then define*

$$\mathfrak{A} \vDash \varphi[a_0, \ldots, a_{n-1}] \quad \textit{iff} \quad \mathfrak{M} \vDash \varphi,$$

where $\mathfrak{M} \supseteq \mathfrak{A}$ is some (or any) S-model with $\mathfrak{M}(v_0) = a_0, \ldots, \mathfrak{M}(v_{n-1}) = a_{n-1}$.

In case $n = 0$ also write $t^{\mathfrak{A}}$ instead of $t^{\mathfrak{A}}[a_0, \ldots, a_{n-1}]$, and $\mathfrak{A} \vDash \varphi$ instead of $\mathfrak{A} \vDash \varphi[a_0, \ldots, a_{n-1}]$. In the latter case we also say: $\mathfrak{A}$ is a model *of φ, $\mathfrak{A}$* satisfies *φ or φ* is true *in $\mathfrak{A}$.*

For $\Phi \subseteq L_0^S$ a class of sentences also write

$$\mathfrak{A} \vDash \Phi \quad \textit{iff for all} \quad \varphi \in \Phi \textit{ holds}: \mathfrak{A} \vDash \varphi.$$

Example 30. *Groups. S_{Gr}*: $=\{\circ, e\}$ with a binary function symbol $\circ$ and a constant symbol e is the *language of groups theory*. The group axioms are

a) $\forall v_0 \forall v_1 \forall v_2 \circ v_0 \circ v_1 v_2 \equiv \circ\circ v_0 v_1 v_2$;

b) $\forall v_0 \circ v_0 e \equiv v_0$;

c) $\forall v_0 \exists v_1 \circ v_0 v_1 \equiv e$.

This defines the axiom set

$$\Phi_{\mathrm{Gr}} = \{\forall v_0 \forall v_1 \forall v_2 \circ v_0 \circ v_1 v_2 \equiv \circ\circ v_0 v_1 v_2, \forall v_0 \circ v_0 e \equiv v_0, \forall v_0 \exists v_1 \circ v_0 v_1 \equiv e\}.$$

An S-structure $\mathfrak{G} = (G, *, k)$ satisfies Φ_{Gr} iff it is a group in the ordinary sense.

Definition 31. *Let S be a language and let $\Phi \subseteq L_0^S$ be a class of S-sentences. Then*

$$\mathrm{Mod}^S \Phi = \{\mathfrak{A} \,|\, \mathfrak{A} \textit{ is an } S\textit{-structure and } \mathfrak{A} \vDash \Phi\}$$

is the model class *of Φ. In case $\Phi = \{\varphi\}$ we also write* $\mathrm{Mod}^S \varphi$ *instead of* $\mathrm{Mod}^S \Phi$. *We also say that Φ is an* axiom system *for* $\mathrm{Mod}^S \Phi$, *or that Φ* axiomatizes *the class* $\mathrm{Mod}^S \Phi$.

Thus $\mathrm{Mod}^{S_{\mathrm{Gr}}} \Phi_{\mathrm{Gr}}$ is the model class of all groups. Model classes are studied in generality within *model theory* which is a branch of mathematical logic. For specific axiom systems Φ the model class $\mathrm{Mod}^S \Phi$ is examined in subfields of mathematics: group theory, ring theory, graph theory, etc. Some typical questions questions are: is $\mathrm{Mod}^S \Phi \neq \emptyset$, i.e., is Φ satisfiable? What are the elements of $\mathrm{Mod}^S \Phi$? Can one classify the isomorphism classes of models? What are the cardinalities of models?

Exercise 7. One may consider $\mathrm{Mod}^S \Phi$ with appropriate morphisms as a category. In certain cases this category has closure properties like closure under products. One can give the categorial definition of cartesian product and show their existence under certain assumptions on Φ .

6 Logical implication and propositional connectives

> *The design of the following treatise is to investigate the fundamental laws of those operations of the mind by which reasoning is performed; to give expression to them in the symbolical language of a Calculus, and upon this foundation to establish the science of Logic and construct its method.*
>
> George Boole, The Laws of Thought

Definition 32. *For a symbol class S and $\Phi \subseteq L^S$ and $\varphi \in L^S$ define that Φ* (logically) implies *φ ($\Phi \vDash \varphi$) iff every S-model $\mathfrak{I} \vDash \Phi$ is also a model of φ.*

Note that logical implication $\vDash$ is a relation between *syntactical* entities which is defined via the *semantic* notion of interpretation. The relation $\Phi \vDash ?$ can be viewed as the central relation in modern axiomatic mathematics: given the assumptions Φ what do they imply? The $\vDash$-relation is usually verified by mathematical *proofs*. These proofs seem to refer to the exploration of some domain of mathematical objects and, in practice, require particular mathematical skills and ingenuity.

We will however show that the logical implication $\vDash$ satisfies certain simple syntactical laws. These laws correspond to ordinary proof methods but are purely formal. Amazingly a finite list of methods will (in principle) suffice for all mathematical proofs. This is Gödel's completeness theorem that we shall prove later.

Theorem 33. *Let S be a language, $t \in T^S$, $\varphi, \psi \in L^S$, and $\Gamma, \Phi \subseteq L^S$. Then*

a) *(Monotonicity) If $\Gamma \subseteq \Phi$ and $\Gamma \vDash \varphi$ then $\Phi \vDash \varphi$.*

b) *(Assumption property) If $\varphi \in \Gamma$ then $\Gamma \vDash \varphi$.*

c) *($\rightarrow$-Introduction) If $\Gamma \cup \varphi \vDash \psi$ then $\Gamma \vDash (\varphi \rightarrow \psi)$.*

d) *($\rightarrow$-Elimination) If $\Gamma \vDash \varphi$ and $\Gamma \vDash (\varphi \rightarrow \psi)$ then $\Gamma \vDash \psi$.*

e) *($\bot$-Introduction) If $\Gamma \vDash \varphi$ and $\Gamma \vDash \neg\varphi$ then $\Gamma \vDash \bot$.*

f) *($\bot$-Elimination) If $\Gamma \cup \{\neg\varphi\} \vDash \bot$ then $\Gamma \vDash \varphi$.*

g) *($\equiv$-Introduction) $\Gamma \vDash t \equiv t$.*

Proof. f) Assume $\Gamma \cup \{\neg\varphi\} \vDash \bot$. Consider an S-model with $\mathfrak{M} \vDash \Gamma$. Assume that $\mathfrak{M} \nvDash \varphi$. Then $\mathfrak{M} \vDash \neg\varphi$. $\mathfrak{M} \vDash \Gamma \cup \{\neg\varphi\}$, and by assumption, $\mathfrak{M} \vDash \bot$. But by the definition of the satisfaction relation, this is false. Thus $\mathfrak{M} \vDash \varphi$. Thus $\Gamma \vDash \varphi$. □

Exercise 8. There are similar rules for the introduction and elimination of junctors like $\wedge$ and $\vee$ that we have introduced as abbreviations:

a) ($\wedge$-Introduction) If $\Gamma \vDash \varphi$ and $\Gamma \vDash \psi$ then $\Gamma \vDash \varphi \wedge \psi$.

b) ($\wedge$-Elimination) If $\Gamma \vDash \varphi \wedge \psi$ then $\Gamma \vDash \varphi$ and $\Gamma \vDash \psi$.

c) ($\vee$-Introduction) If $\Gamma \vDash \varphi$ then $\Gamma \vDash \varphi \vee \psi$ and $\Gamma \vDash \psi \vee \varphi$.

d) ($\vee$-Elimination) If $\Gamma \vDash \varphi \vee \psi$ and $\Gamma \vdash \neg\varphi$ then $\Gamma \vDash \psi$.

7 Substitution and term rules

To prove further rules for equality and quantification, we first have to consider the *substitution* of terms in formulas.

Definition 34. *For a term $s \in T^S$, pairwise distinct variables $x_0, ..., x_{r-1}$ and terms $t_0, ..., t_{r-1} \in T^S$ define the* (simultaneous) substitution

$$s \, \frac{t_0....t_{r-1}}{x_0...x_{r-1}}$$

of $t_0, ..., t_{r-1}$ for $x_0, ..., x_{r-1}$ by recursion:

a) $x \, \frac{t_0....t_{r-1}}{x_0...x_{r-1}} = \begin{cases} x, \text{ if } x \neq x_0, ..., x \neq x_{r-1} \\ t_i\,, \text{ if } x = x_i \end{cases}$ *for all variables x;*

b) $(f s_0 ... s_{n-1}) \, \frac{t_0....t_{r-1}}{x_0...x_{r-1}} = f s_0 \, \frac{t_0....t_{r-1}}{x_0...x_{r-1}} ... s_{n-1} \, \frac{t_0....t_{r-1}}{x_0...x_{r-1}}$ *for all n-ary function symbols $f \in S$.*

Note that the *simultaneous* substitution

$$s \, \frac{t_0....t_{r-1}}{x_0...x_{r-1}}$$

is in general different from a *successive* substitution

$$s \, \frac{t_0}{x_0} \, \frac{t_1}{x_1} ... \frac{t_{r-1}}{x_{r-1}}$$

which depends on the order of substitution. E.g., $x \, \frac{yx}{xy} = y$, $x \, \frac{y}{x} \, \frac{x}{y} = y \, \frac{x}{y} = x$ and $x \, \frac{x}{y} \, \frac{y}{x} = x \, \frac{y}{x} = y$.

Definition 35. *For a formula $\varphi \in L^S$, pairwise distinct variables $x_0, ..., x_{r-1}$ and terms $t_0, ..., t_{r-1} \in T^S$ define the* (simultaneous) substitution

$$\varphi \, \frac{t_0....t_{r-1}}{x_0...x_{r-1}}$$

of $t_0, ..., t_{r-1}$ for $x_0, ..., x_{r-1}$ by recursion:

a) $(s_0 \equiv s_1) \, \frac{t_0....t_{r-1}}{x_0...x_{r-1}} = s_0 \, \frac{t_0....t_{r-1}}{x_0...x_{r-1}} \equiv s_1 \, \frac{t_0....t_{r-1}}{x_0...x_{r-1}}$ *for all terms $s_0, s_1 \in T^S$;*

b) $(R s_0 ... s_{n-1}) \, \frac{t_0....t_{r-1}}{x_0...x_{r-1}} = R s_0 \, \frac{t_0....t_{r-1}}{x_0...x_{r-1}} ... s_{n-1} \, \frac{t_0....t_{r-1}}{x_0...x_{r-1}}$ *for all n-ary relation symbols $R \in s$ and terms $s_0, ..., s_{n-1} \in T^S$;*

c) $(\neg \varphi) \, \frac{t_0....t_{r-1}}{x_0...x_{r-1}} = \neg (\varphi \, \frac{t_0....t_{r-1}}{x_0...x_{r-1}})$*;*

d) $(\varphi \rightarrow \psi) \, \frac{t_0....t_{r-1}}{x_0...x_{r-1}} = (\varphi \, \frac{t_0....t_{r-1}}{x_0...x_{r-1}} \rightarrow \psi \, \frac{t_0....t_{r-1}}{x_0...x_{r-1}})$*;*

e) for $(\forall x \varphi) \, \frac{t_0....t_{r-1}}{x_0...x_{r-1}}$ *we proceed in two steps: let $x_{i_0}, ..., x_{i_{s-1}}$ with $i_0 < ... < i_{s-1}$ be exactly those x_i which are "relevant" for the substitution, i.e., $x_i \in \mathrm{free}(\forall x \varphi)$ and $x_i \neq t_i$.*

- *if x does not occur in $t_{i_0}, ..., t_{i_{s-1}}$, then set*

$$(\forall x \varphi) \, \frac{t_0....t_{r-1}}{x_0...x_{r-1}} = \forall x \, (\varphi \, \frac{t_{i_0}...t_{i_{s-1}}}{x_{i_0}...x_{i_{s-1}}}).$$

- *if x does occur in $t_{i_0}, ..., t_{i_{s-1}}$, then let $k \in \mathbb{N}$ minimal such that v_k does not occur in φ, $t_{i_0}, ..., t_{i_{s-1}}$ and set*

$$(\forall x \varphi) \, \frac{t_0....t_{r-1}}{x_0...x_{r-1}} = \forall v_k \, (\varphi \, \frac{t_{i_0}...t_{i_{s-1}} v_k}{x_{i_0}...x_{i_{s-1}} x}).$$

The following *substitution theorem* shows that syntactic substitution corresponds semantically to a (simultaneous) modification of assignments by interpreted terms. The definition of substitution was intended to make the substitution theorem true. There are variants of the syntactical substitution which could also satisfy the substitution theorem.

Theorem 36. *Consider an S-model $\mathfrak{M}$, pairwise distinct variables $x_0, ..., x_{r-1}$ and terms $t_0, ..., t_{r-1} \in T^S$.*

a) If $s \in T^S$ is a term,

$$\mathfrak{M}(s\,\frac{t_0...t_{r-1}}{x_0...x_{r-1}}) = \mathfrak{M}\,\frac{\mathfrak{M}(t_0)...\mathfrak{M}(t_{r-1})}{x_0...x_{r-1}}(s).$$

b) If $\varphi \in L^S$ is a formula,

$$\mathfrak{M} \vDash \varphi\,\frac{t_0...t_{r-1}}{x_0...x_{r-1}} \text{ iff } \mathfrak{M}\,\frac{\mathfrak{M}(t_0)...\mathfrak{M}(t_{r-1})}{x_0...x_{r-1}} \vDash \varphi.$$

Proof. By induction on the complexities of s and φ.
a) *Case 1*: $s = x$.
Case 1.1: $x \notin \{x_0, ..., x_{r-1}\}$. Then

$$\mathfrak{M}(x\,\frac{t_0...t_{r-1}}{x_0...x_{r-1}}) = \mathfrak{M}(x) = \mathfrak{M}\,\frac{\mathfrak{M}(t_0)...\mathfrak{M}(t_{r-1})}{x_0...x_{r-1}}(x).$$

Case 1.2: $x = x_i$. Then

$$\mathfrak{M}(x\,\frac{t_0...t_{r-1}}{x_0...x_{r-1}}) = \mathfrak{M}(t_i) = \mathfrak{M}\,\frac{\mathfrak{M}(t_0)...\mathfrak{M}(t_{r-1})}{x_0...x_{r-1}}(x_i) = \mathfrak{M}\,\frac{\mathfrak{M}(t_0)...\mathfrak{M}(t_{r-1})}{x_0...x_{r-1}}(x).$$

Case 2: $s = fs_0...s_{n-1}$ where $f \in S$ is an n-ary function symbol and the terms s_0, ..., $s_{n-1} \in T^S$ satisfy the theorem. Then

$$\begin{aligned}
\mathfrak{M}((fs_0...s_{n-1})\,\frac{t_0...t_{r-1}}{x_0...x_{r-1}}) &= \mathfrak{M}(fs_0\,\frac{t_0...t_{r-1}}{x_0...x_{r-1}}...s_{n-1}\,\frac{t_0...t_{r-1}}{x_0...x_{r-1}}) \\
&= \mathfrak{M}(f)(\mathfrak{M}(s_0\,\frac{t_0...t_{r-1}}{x_0...x_{r-1}}), ..., \mathfrak{M}(s_{n-1}\,\frac{t_0...t_{r-1}}{x_0...x_{r-1}})) \\
&= \mathfrak{M}(f)(\mathfrak{M}\,\frac{\mathfrak{M}(t_0)...\mathfrak{M}(t_{r-1})}{x_0...x_{r-1}}(s_0), \\
&\qquad ..., \mathfrak{M}\,\frac{\mathfrak{M}(t_0)...\mathfrak{M}(t_{r-1})}{x_0...x_{r-1}}(s_{n-1})) \\
&= \mathfrak{M}\,\frac{\mathfrak{M}(t_0)....\mathfrak{M}(t_{r-1})}{x_0...x_{r-1}}(fs_0...s_{n-1}).
\end{aligned}$$

Assuming that the substitution theorem is proved for terms, we prove
b) *Case 4*: $\varphi = Rs_0...s_{n-1}$. Then

$$\begin{aligned}
\mathfrak{M} \vDash (Rs_0...s_{n-1})\,\frac{t_0....t_{r-1}}{x_0...x_{r-1}} &\text{ iff } \mathfrak{M} \vDash Rs_0\,\frac{t_0....t_{r-1}}{x_0...x_{r-1}}...s_{n-1}\,\frac{t_0....t_{r-1}}{x_0...x_{r-1}} \\
&\text{ iff } R^{\mathfrak{M}}\left(\mathfrak{M}(s_0\,\frac{t_0....t_{r-1}}{x_0...x_{r-1}}), ..., \mathfrak{M}(s_1\,\frac{t_0....t_{r-1}}{x_0...x_{r-1}})\right) \\
&\text{ iff } R^{\mathfrak{M}}\left(\mathfrak{M}\frac{\mathfrak{M}(t_0)....\mathfrak{M}(t_{r-1})}{x_0...x_{r-1}}(s_0),\right. \\
&\qquad \left. ..., \mathfrak{M}\frac{\mathfrak{M}(t_0)....\mathfrak{M}(t_{r-1})}{x_0...x_{r-1}}(s_{n-1})\right) \\
&\text{ iff } \mathfrak{M}\frac{\mathfrak{M}(t_0)....\mathfrak{M}(t_{r-1})}{x_0...x_{r-1}} \vDash Rs_0...s_{n-1}
\end{aligned}$$

Equations $s_0 \equiv s_1$ can be treated as a special case of the relational *Case 4*. Propositional combinations of formulas by $\bot$, $\neg$ and $\rightarrow$ behave similar to terms; indeed formulas can be viewed as terms whose values are truth values. So we are left with universal quantification.

:

Case 5: $\varphi = (\forall x \psi) \frac{t_0 \dots t_{r-1}}{x_0 \dots x_{r-1}}$, assuming that the theorem holds for ψ.

We proceed according to our definition of syntactic substitution. Let $x_{i_0}, \dots, x_{i_{s-1}}$ with $i_0 < \dots < i_{s-1}$ be exactly those x_i such that $x_i \in \mathrm{free}(\forall x \psi)$ and $x_i \neq t_i$. Since

$$\mathfrak{M} \frac{\mathfrak{M}(t_0) \dots \mathfrak{M}(t_{r-1})}{x_0 \dots x_{r-1}} \vDash \varphi \text{ iff } \mathfrak{M} \frac{\mathfrak{M}(t_{i_0}) \dots \mathfrak{M}(t_{i_{s-1}})}{x_{i_0} \dots x_{i_{s-1}}} \vDash \varphi,$$

we can assume that $(x_0, \dots, x_{r-1}) = (x_{i_0}, \dots, x_{i_{s-1}})$, i.e., every x_i is free in $\forall x \psi$, $x_i \neq x$, and $x_i \neq t_i$. Now follow the two cases in the definition of the substitution:

Case 5.1: The variable x does not occur in $t_0, \dots., t_{r-1}$ and

$$(\forall x \psi) \frac{t_0 \dots. t_{r-1}}{x_0 \dots x_{r-1}} = \forall x \, (\psi \frac{t_0 \dots. t_{r-1}}{x_0 \dots x_{r-1}}).$$

$$\begin{aligned}
\mathfrak{M} \vDash (\forall x \psi) \frac{t_0 \dots t_{r-1}}{x_0 \dots x_{r-1}} \quad & \text{iff } \; \mathfrak{M} \vDash \forall x \, (\psi \frac{t_0 \dots t_{r-1}}{x_0 \dots x_{r-1}}) \\
& \text{iff } \; \text{for all } a \in M \text{ holds } \mathfrak{M} \frac{a}{x} \vDash \psi \frac{t_0 \dots t_{r-1}}{x_0 \dots x_{r-1}} \\
& \qquad \text{(definition of } \vDash) \\
& \text{iff } \; \text{for all } a \in M \text{ holds} \\
& \qquad (\mathfrak{M} \frac{a}{x}) \frac{\mathfrak{M}\frac{a}{x}(t_0) \dots \mathfrak{M}\frac{a}{x}(t_{r-1})}{x_0 \dots x_{r-1}} \vDash \psi \\
& \qquad \text{(by the inductive hypothesis for } \psi) \\
& \text{iff } \; \text{for all } a \in M \text{ holds} \\
& \qquad (\mathfrak{M} \frac{a}{x}) \frac{\mathfrak{M}(t_0) \dots \mathfrak{M}(t_{r-1})}{x_0 \dots x_{r-1}} \vDash \psi \\
& \qquad \text{(since } x \text{ does not occur in } t_i) \\
& \text{iff } \; \text{for all } a \in M \text{ holds} \\
& \qquad \mathfrak{M} \frac{\mathfrak{M}(t_0) \dots \mathfrak{M}(t_{r-1}) \, a}{x_0 \dots x_{r-1} \, x} \vDash \psi \\
& \qquad \text{(since } x \text{ does not occur in } x_0, \dots, x_{r-1}) \\
& \text{iff } \; \text{for all } a \in M \text{ holds} \\
& \qquad (\mathfrak{M} \frac{\mathfrak{M}(t_0) \dots \mathfrak{M}(t_{r-1})}{x_0 \dots x_{r-1}}) \frac{a}{x} \vDash \psi \\
& \qquad \text{(by simple properties of assignments)} \\
& \text{iff } \; \mathfrak{M} \frac{\mathfrak{M}(t_0) \dots \mathfrak{M}(t_{r-1})}{x_0 \dots x_{r-1}} \vDash \forall x \psi
\end{aligned}$$

Case 5.2: The variable x occurs in $t_0, \dots., t_{r-1}$. Then

$$(\forall x \psi) \frac{t_0 \dots. t_{r-1}}{x_0 \dots x_{r-1}} = \forall v_k \, (\psi \frac{t_0 \dots. t_{r-1} v_k}{x_0 \dots x_{r-1} x}),$$

where $k\in\mathbb{N}$ is minimal such that v_k does not occur in φ, $t_{i_0},....,t_{i_{s-1}}$.

$$
\begin{aligned}
\mathfrak{M}\models(\forall x\,\psi)\,\frac{t_0\ldots t_{r-1}}{x_0\ldots x_{r-1}} \quad &\text{iff}\quad \mathfrak{M}\models\forall v_k\,(\psi\,\frac{t_0\ldots.t_{r-1}v_k}{x_0\ldots x_{r-1}x})\\
&\text{iff}\quad \text{for all } a\in M \text{ holds } \mathfrak{M}\frac{a}{v_k}\models\psi\,\frac{t_0\ldots t_{r-1}v_k}{x_0\ldots x_{r-1}x}\\
&\text{iff}\quad \text{for all } a\in M \text{ holds}\\
&\qquad\quad (\mathfrak{M}\frac{a}{v_k})\frac{\mathfrak{M}\frac{a}{v_k}(t_0)\ldots\mathfrak{M}\frac{a}{v_k}(t_{r-1})\mathfrak{M}\frac{a}{v_k}(v_k)}{x_0\ldots x_{r-1}\,x}\models\psi\\
&\qquad\quad \text{(inductive hypothesis for } \psi)\\
&\text{iff}\quad \text{for all } a\in M \text{ holds}\\
&\qquad\quad (\mathfrak{M}\frac{a}{x})\frac{\mathfrak{M}(t_0)\ldots\mathfrak{M}(t_{r-1})a}{x_0\ldots x_{r-1}x}\models\psi\\
&\qquad\quad \text{(since } v_k \text{ does not occur in } t_i)\\
&\text{iff}\quad \text{for all } a\in M \text{ holds}\\
&\qquad\quad \mathfrak{M}\,\frac{\mathfrak{M}(t_0)\ldots\mathfrak{M}(t_{r-1})\,a}{x_0\ldots x_{r-1}\,x}\models\psi\\
&\qquad\quad \text{(since } x \text{ is anyway sent to } a)\\
&\text{iff}\quad \text{for all } a\in M \text{ holds}\\
&\qquad\quad (\,\mathfrak{M}\,\frac{\mathfrak{M}(t_0)\ldots\mathfrak{M}(t_{r-1})}{x_0\ldots x_{r-1}})\,\frac{a}{x}\models\psi\\
&\qquad\quad \text{(by simple properties of assignments)}\\
&\text{iff}\quad \mathfrak{M}\,\frac{\mathfrak{M}(t_0)\ldots\mathfrak{M}(t_{r-1})}{x_0\ldots x_{r-1}}\models\forall x\,\psi
\end{aligned}
$$

□

We can now formulate properties of the $\models$ relation in connection with the treatment of variables.

Theorem 37. *Let S be a language. Let x,y be variables, $t,t'\in T^S$, $\varphi\in L^S$, and $\Gamma\subseteq L^S$. Then:*

a) ($\forall$-Introduction) If $\Gamma\models\varphi\frac{y}{x}$ and $y\notin\operatorname{free}(\Gamma\cup\{\forall x\varphi\})$ then $\Gamma\models\forall x\,\varphi$.

b) ($\forall$-elimination) If $\Gamma\models\forall x\,\varphi$ then $\Gamma\models\varphi\frac{t}{x}$.

c) ($\equiv$-Elimination or substitution) If $\Gamma\models\varphi\frac{t}{x}$ and $\Gamma\models t\equiv t'$ then $\Gamma\models\varphi\frac{t'}{x}$.

Proof. a) Assume $\Gamma\models\varphi\,\frac{y}{x}$ and $y\notin\operatorname{free}(\Gamma\cup\{\forall x\varphi\})$. Consider an S-model $\mathfrak{M}$ with $\mathfrak{M}\models\Gamma$. Let $a\in M=|\mathfrak{M}|$. Since $y\notin\operatorname{free}(\Gamma)$, $\mathfrak{M}\frac{a}{y}\models\Gamma$. By assumption, $\mathfrak{M}\frac{a}{y}\models\varphi\frac{y}{x}$. By the substitution theorem,

$$(\mathfrak{M}\,\frac{a}{y})\,\frac{\mathfrak{M}\frac{a}{y}\,(y)}{x}\models\varphi \text{ and so } (\mathfrak{M}\,\frac{a}{y})\,\frac{a}{x}\models\varphi$$

Case 1: $x=y$. Then $\mathfrak{M}\frac{a}{x}\models\varphi$.

Case 2: $x\neq y$. Then $\mathfrak{M}\,\frac{a\,a}{y\,x}\models\varphi$, and since $y\notin\operatorname{free}(\varphi)$ we have $\mathfrak{M}\frac{a}{x}\models\varphi$.

Since $a\in M$ is arbitrary, $\mathfrak{M}\models\forall x\varphi$. Thus $\Gamma\models\forall x\varphi$.

b) Let $\Gamma\models\forall x\varphi$. Consider an S-model $\mathfrak{M}$ with $\mathfrak{M}\models\Gamma$. For all $a\in M=|\mathfrak{M}|$ holds $\mathfrak{M}\,\frac{a}{x}\models\varphi$. In particular $\mathfrak{M}\,\frac{\mathfrak{M}(t)}{x}\models\varphi$. By the substitution theorem, $\mathfrak{M}\models\varphi\frac{t}{x}$. Thus $\Gamma\models\varphi\,\frac{t}{x}$.

c) Let $\Gamma\models\varphi\frac{t}{x}$ and $\Gamma\models t\equiv t'$. Consider an S-model $\mathfrak{M}$ mit $\mathfrak{M}\models\Gamma$. By assumption $\mathfrak{M}\models\varphi\frac{t}{x}$ and $\mathfrak{M}\models t\equiv t'$. By the substitution theorem

$$\mathfrak{M}\frac{\mathfrak{M}(t)}{x}\models\varphi.$$

Since $\mathfrak{M}(t) = \mathfrak{M}(t')$,

$$\mathfrak{M}\frac{\mathfrak{M}(t')}{x} \vDash \varphi$$

and again by the substitution theorem

$$\mathfrak{M} \vDash \varphi\frac{t'}{x}.$$

Thus $\Gamma \vDash \varphi\frac{t'}{x}$. □

Note that in proving these proof rules we have used corresponding forms of arguments in the language of our discourse. This "circularity" was noted before and is a general feature in formalizations of logic. A particularly important method of proof is the $\forall$-introduction: to prove a universal statement $\forall x\varphi$ it suffices to consider an "arbitrary but fixed" y and prove the claim for y. Formally this corresponds to using a "new" variable $y \notin \mathrm{free}(\Gamma \cup \{\forall x\varphi\})$.

8 A sequent calculus

> *The only way to rectify our reasonings is to make them as tangible as those of the Mathematicians, so that we can find our error at a glance, and when there are disputes among persons, we can simply say: Let us calculate [calculemus], without further ado, to see who is right.* G.W. Leibniz

We can put the rules of implication established in the previous two sections together as a *calculus* which leads from correct implications $\Phi \vDash \varphi$ to further correct implications $\Phi' \vDash \varphi'$. Our *sequent calculus* will work on finite *sequents* $(\varphi_0, ..., \varphi_{n-1}, \varphi_n)$ of formulas, whose intuitive meaning is that $\{\varphi_0, ..., \varphi_{n-1}\}$ implies φ_n . The GÖDEL completeness theorem shows that these rules actually generate the implication relation $\vDash$. Fix a language S for this section.

Definition 38. *A finite sequence $(\varphi_0, ..., \varphi_{n-1}, \varphi_n)$ of S-formulas is called a* sequent. *The initial segment $\Gamma = (\varphi_0, ..., \varphi_{n-1})$ is the* antecedent *and φ_n is the* succedent *of the sequent. We usually write $\varphi_0 ... \varphi_{n-1}\,\varphi_n$ or $\Gamma\,\varphi_n$ instead of $(\varphi_0, ..., \varphi_{n-1}, \varphi_n)$. To emphasize the last element of the antecedent we may also denote the sequent by $\Gamma'\,\varphi_{n-1}\,\varphi_n$ with $\Gamma' = (\varphi_0, ..., \varphi_{n-2})$.*

A sequent $\varphi_0 ... \varphi_{n-1}\,\varphi$ is correct *if $\{\varphi_0 ... \varphi_{n-1}\} \vDash \varphi$.*

Exercise 9. One could also define a sequent to be the concatenation of finitely many formulas

Definition 39. *The* sequent calculus *consists of the following (sequent-)rules:*

- *monotonicity* (MR) $\dfrac{\Gamma \quad \varphi}{\Gamma \;\; \psi \;\; \varphi}$
- *assumption* (AR) $\dfrac{}{\Gamma \;\; \varphi \;\; \varphi}$

– $\rightarrow$*-introduction* ($\rightarrow I$) $\dfrac{\Gamma\ \ \varphi\ \ \psi}{\Gamma\ \ \ \ \varphi\rightarrow\psi}$

– $\rightarrow$*-elimination* ($\rightarrow E$) $\dfrac{\begin{array}{l}\Gamma\ \ \varphi\\ \Gamma\ \ \varphi\rightarrow\psi\end{array}}{\Gamma\ \ \psi}$

– $\perp$*-introduction* ($\perp I$) $\dfrac{\begin{array}{l}\Gamma\ \ \varphi\\ \Gamma\ \ \neg\varphi\end{array}}{\Gamma\ \ \perp}$

– $\perp$*-elimination* ($\perp E$) $\dfrac{\Gamma\ \ \neg\varphi\ \ \perp}{\Gamma\ \ \ \ \varphi}$

– $\forall$*-introduction* ($\forall I$) $\dfrac{\Gamma\ \ \varphi\frac{y}{x}}{\Gamma\ \ \forall x\varphi}$, *if* $y\notin\mathrm{free}(\Gamma\cup\{\forall x\varphi\})$

– $\forall$*-elimination* ($\forall E$) $\dfrac{\Gamma\ \ \forall x\varphi}{\Gamma\ \ \varphi\frac{t}{x}}$, *if* $t\in T^S$

– $\equiv$*-introduction* ($\equiv I$) $\dfrac{}{\Gamma\ \ t\equiv t}$, *if* $t\in T^S$

– $\equiv$*-elimination* ($\equiv E$) $\dfrac{\begin{array}{l}\Gamma\ \ \varphi\frac{t}{x}\\ \Gamma\ \ t\equiv t'\end{array}}{\Gamma\ \ \varphi\frac{t'}{x}}$

The deduction relation *is the smallest subclass* $\vdash\subseteq\mathrm{Seq}(S)$ *of the class of sequents which is closed under these rules. We write* $\varphi_0\ldots\varphi_{n-1}\vdash\varphi$ *instead of* $\varphi_0\ldots\varphi_{n-1}\ \varphi\in\vdash$*. For* Φ *an arbitrary class of formulas define* $\Phi\vdash\varphi$ *iff there are* $\varphi_0, \ldots, \varphi_{n-1}\in\Phi$ *such that* $\varphi_0\ldots\varphi_{n-1}\vdash\varphi$*. We say that* φ *can be* deduced *or* derived *from* $\varphi_0\ldots\varphi_{n-1}$ *or* Φ*, resp. We also write* $\vdash\varphi$ *instead of* $\emptyset\vdash\varphi$ *and say that* φ *is a* tautology*.*

Remark 40. A calculus is a formal system for obtaining (mathematical) results. The usual algorithms for addition and multiplication of decimal numbers are calculi: the results are achieved by symbolic and systematic operations on the decimal symbols $0,\ldots,9$. Such an addition is not an addition in terms of joining together line segments of certain lengths or forming the union of disjoint finite sets. The calculi are however correct in that the interpretation of the decimal numbers obtained correspond to the results of the intuitive operations of joining line segments or disjoint unions.

Mathematics has shown that far more sophisticated operations can also be described by *calculi*. The derivative of a polynomial function

$$p(x)=a_n x^n+a_{n-1}x^{n-1}+\ldots+a_0$$

can be obtained by formal manipulations of exponents and coefficients:

$$p'(x)=n\,a_n x^{n-1}+(n-1)\,a_{n-1}x^{n-2}+\ldots+a_1$$

without explicitly forming limits of difference quotients.

Since many basic results of analysis can be expressed as formal calculi, the word *calculus* is used for basic analysis courses in the English speaking world. Similarly in German one uses the words Differential*rechnung* and Integral*rechnung*. The words *derivation* or *Ableitung* also refer to derivations within a formal calculus.

A formula $\varphi \in L^S$ is derivable from $\Gamma = \varphi_0 \ldots \varphi_{n-1}$ ($\Gamma \vdash \varphi$) iff there is a *derivation* or a *formal proof*

$$(\Gamma_0 \varphi_0, \Gamma_1 \varphi_1, \ldots, \Gamma_{k-1} \varphi_{k-1})$$

of $\Gamma\varphi = \Gamma_{k-1}\varphi_{k-1}$, in which every sequent $\Gamma_i \varphi_i$ is generated by a sequent rule from sequents $\Gamma_{i_0}\varphi_{i_0}, \ldots, \Gamma_{i_{n-1}}\varphi_{i_{n-1}}$ with $i_0, \ldots, i_{n-1} < i$.

We usually write the derivation $(\Gamma_0 \varphi_0, \Gamma_1 \varphi_1, \ldots, \Gamma_{k-1} \varphi_{k-1})$ as a vertical scheme

$$\begin{array}{ll} \Gamma_0 & \varphi_0 \\ \Gamma_1 & \varphi_1 \\ \vdots & \\ \Gamma_{k-1} & \varphi_{k-1} \end{array}$$

where we may also indicate rules and other remarks along the course of the derivation.

In our theorems on the laws of implication we have already shown:

Theorem 41. *The sequent calculus is* correct*, i.e., every rule of the sequent calculus leads from correct sequents to correct sequents. Thus every derivable sequent is correct. This means that*

$$\vdash \subseteq \vDash .$$

The converse inclusion corresponds to

Definition 42. *The sequent calculus is* complete *iff* $\vDash \subseteq \vdash$.

The Gödel completeness theorem proves the completeness of the sequent calculus. The definition of $\vdash$ immediately implies the following *finiteness* or *compactness theorem.*

Theorem 43. *Let* $\Phi \subseteq L^S$ *and* $\varphi \in L^S$*. Then* $\Phi \vdash \varphi$ *iff there is a finite subset* $\Phi_0 \subseteq \Phi$ *such that* $\Phi_0 \vdash \varphi$*.*

After proving the completeness theorem, such structural properties carry over to the implication relation $\vDash$.

9 Derivable sequent rules

The composition of rules of the sequent calculus yields *derived sequent rules* which are again correct. First note:

Lemma 44. *Assume that*

$$\begin{array}{ll} \Gamma & \varphi_0 \\ \vdots & \\ \Gamma & \varphi_{k-1} \\ \hline \Gamma & \varphi_k \end{array}$$

is a derived rule of the sequent calculus. Then

$$\begin{array}{ll} \Gamma_0 & \varphi_0 \\ \vdots & \\ \Gamma_{k-1} & \varphi_{k-1} \\ \hline \Gamma & \varphi_k \end{array} \text{, where } \Gamma_0, \ldots, \Gamma_{k-1} \text{ are initial sequences of } \Gamma$$

is also a derived rule of the sequent calculus.

Proof. This follows immediately from iterated applications of the monotonicity rule. □

We now list several derived rules.

9.1 Auxiliary rules

We write the derivation of rules as proofs in the sequent calculus where the premisses of the derivation are written above the upper horizontal line and the conclusion as last row.

ex falso quodlibet $\dfrac{\Gamma\ \ \bot}{\Gamma\ \ \varphi}$:

$$\begin{array}{llll}
1. & \Gamma & & \bot \\
\hline
2. & \Gamma & \neg\varphi & \bot \\
\hline
3. & \Gamma & & \varphi
\end{array}$$

$\neg$-*Introduction* $\dfrac{\Gamma\ \ \varphi\ \ \bot}{\Gamma\ \ \ \ \neg\varphi}$:

$$\begin{array}{lllll}
1. & \Gamma & \varphi & & \bot \\
\hline
2. & \Gamma & & & \varphi\to\bot \\
3. & \Gamma & \neg\neg\varphi & & \neg\neg\varphi \\
4. & \Gamma & \neg\neg\varphi & \neg\varphi & \neg\varphi \\
5. & \Gamma & \neg\neg\varphi & \neg\varphi & \bot \\
6. & \Gamma & \neg\neg\varphi & & \varphi \\
7. & \Gamma & \neg\neg\varphi & & \bot \\
\hline
8. & \Gamma & & & \neg\varphi
\end{array}$$

$$\begin{array}{ll}
\Gamma & \neg\varphi \\
\hline
\Gamma & \varphi\to\psi
\end{array}$$

$$\begin{array}{ll}
\Gamma & \psi \\
\hline
\Gamma & \varphi\to\psi
\end{array}$$

Cut rule

$$\begin{array}{lll}
\Gamma & & \varphi \\
\Gamma & \varphi & \psi \\
\hline
\Gamma & & \psi
\end{array}$$

Contraposition

$$\begin{array}{lll}
\Gamma & \varphi & \psi \\
\hline
\Gamma & \neg\psi & \neg\varphi
\end{array}$$

9.2 Introduction and elimination of $\vee, \wedge, \ldots$

The (abbreviating) logical symbols $\vee$, $\wedge$, and $\exists$ also possess (derived) introduction and elimination rules. We list the rules and leave their derivations as exercises.

$\vee$-*Introduction*

$$\begin{array}{ll}
\Gamma & \varphi \\
\hline
\Gamma & \varphi\vee\psi
\end{array}$$

$\lor$-*Introduction*

$$\begin{array}{ll} \Gamma & \psi \\ \hline \Gamma & \varphi \lor \psi \end{array}$$

$\lor$-*Elimination*

$$\begin{array}{ll} \Gamma & \varphi \lor \psi \\ \Gamma & \varphi \to \chi \\ \Gamma & \psi \to \chi \\ \hline \Gamma & \chi \end{array}$$

$\land$-*Introduction*

$$\begin{array}{ll} \Gamma & \varphi \\ \Gamma & \psi \\ \hline \Gamma & \varphi \land \psi \end{array}$$

$\land$-*Elimination*

$$\begin{array}{ll} \Gamma & \varphi \land \psi \\ \hline \Gamma & \varphi \end{array}$$

$\land$-*Elimination*

$$\begin{array}{ll} \Gamma & \varphi \land \psi \\ \hline \Gamma & \psi \end{array}$$

$\exists$-*Introduction*

$$\begin{array}{ll} \Gamma & \varphi\frac{t}{x} \\ \hline \Gamma & \exists x \varphi \end{array}$$

$\exists$-*Elimination*

$$\begin{array}{lll} \Gamma & & \exists x \varphi \\ \Gamma & \varphi\frac{y}{x} & \psi \\ \hline \Gamma & & \psi \end{array} \quad \text{where } y \notin \text{free}(\Gamma \cup \{\exists x \varphi, \psi\})$$

9.3 Manipulations of antecedents

We derive rules by which the formulas in the antecedent may be permuted arbitrarily, showing that only the *set* of antecedent formulas is relevant.

Transpositions of premisses

$$\begin{array}{lllll} 1. & \Gamma & \varphi & \psi & \chi \\ \hline 2. & \Gamma & \varphi & & \psi \to \chi \\ 3. & \Gamma & & & \varphi \to (\psi \to \chi) \\ 4. & \Gamma & \psi & & \psi \\ 5. & \Gamma & \psi & \varphi & \varphi \\ 6. & \Gamma & \psi & \varphi & \psi \to \chi \\ \hline 7. & \Gamma & \psi & \varphi & \chi \end{array}$$

Duplication of premisses

$$\begin{array}{lllll} 1. & \Gamma & \varphi & & \psi \\ \hline 2. & \Gamma & \varphi & \varphi & \psi \end{array}$$

Elimination of double premisses

$$\begin{array}{llll} 1. & \Gamma & \varphi & \varphi \;\; \psi \\ \hline 2. & \Gamma & \varphi & \varphi \rightarrow \psi \\ 3. & \Gamma & & \varphi \rightarrow (\varphi \rightarrow \psi) \\ 4. & \Gamma & \varphi & \varphi \\ \hline 5. & \Gamma & \varphi & \psi \end{array}$$

Iterated applications of these rules yield:

Lemma 45. *Let* $\varphi_0 \ldots \varphi_{m-1}$ *and* $\psi_0 \ldots \psi_{n-1}$ *be antecedents such that*

$$\{\varphi_0, \ldots, \varphi_{m-1}\} = \{\psi_0, \ldots, \psi_{n-1}\}$$

and $\chi \in L^S$. *Then*

$$\frac{\varphi_0 \;\; \cdots \;\; \varphi_{m-1} \;\; \chi}{\psi_0 \;\; \cdots \;\; \psi_{n-1} \;\; \chi}$$

is a derived rule.

9.4 Formal proofs about $\equiv$

We give some examples of formal proofs which show that within the proof calculus $\equiv$ is an equivalence relation.

Lemma 46. *We prove the following tautologies:*

a) *Reflexivity:* $\vdash \forall x\, x \equiv x$

b) *Symmetry:* $\vdash \forall x \forall y (x \equiv y \rightarrow y \equiv x)$

c) *Transitivity:* $\vdash \forall x \forall y \forall z (x \equiv y \wedge y \equiv z \rightarrow x \equiv z)$

Proof. a)

$$\begin{array}{l} x \equiv x \\ \hline \forall x\, x \equiv x \end{array}$$

b)

$$\begin{array}{ll} x \equiv y & x \equiv y \\ x \equiv y & x \equiv x \\ x \equiv y & (z \equiv x)\frac{x}{z} \\ x \equiv y & (z \equiv x)\frac{y}{x} \\ x \equiv y & y \equiv x \\ & x \equiv y \rightarrow y \equiv x \\ & \forall y (x \equiv y \rightarrow y \equiv x) \\ \hline & \forall x \forall y (x \equiv y \rightarrow y \equiv x) \end{array}$$

c)

$$\begin{array}{ll} x \equiv y \wedge y \equiv z & x \equiv y \wedge y \equiv z \\ x \equiv y \wedge y \equiv z & x \equiv y \\ x \equiv y \wedge y \equiv z & (x \equiv w)\frac{y}{w} \\ x \equiv y \wedge y \equiv z & y \equiv z \\ x \equiv y \wedge y \equiv z & (x \equiv w)\frac{z}{w} \\ x \equiv y \wedge y \equiv z & x \equiv z \\ & x \equiv y \wedge y \equiv z \rightarrow x \equiv z \\ & \forall z (x \equiv y \wedge y \equiv z \rightarrow x \equiv z) \\ & \forall y \forall z (x \equiv y \wedge y \equiv z \rightarrow x \equiv z) \\ \hline & \forall x \forall y \forall z (x \equiv y \wedge y \equiv z \rightarrow x \equiv z) \end{array}$$

□

We show moreover that $\equiv$ is a *congruence relation* from the perspective of $\vdash$.

Theorem 47. *Let* $\varphi \in L^S$ *and* $t_0, \ldots, t_{n-1}, t_0', \ldots, t_{n-1}' \in T^S$. *Then*

$$\vdash t_0 \equiv t_0' \wedge \ldots \wedge t_{n-1} \equiv t_{n-1}' \rightarrow (\varphi \frac{t_0 \ldots t_{n-1}}{v_0 \ldots v_{n-1}} \leftrightarrow \varphi \frac{t_0' \ldots t_{n-1}'}{v_0 \ldots v_{n-1}}).$$

Proof. Choose pairwise distinct "new" variables $u_0, \ldots, u_{n-1}$. Then

$$\varphi \frac{t_0 \ldots t_{n-1}}{v_0 \ldots v_{n-1}} = \varphi \frac{u_0}{v_0} \frac{u_1}{v_1} \ldots \frac{u_{n-1}}{v_{n-1}} \frac{t_0}{u_0} \frac{t_1}{u_1} \ldots \frac{t_{n-1}}{u_{n-1}}$$

and

$$\varphi \frac{t_0' \ldots t_{n-1}'}{v_0 \ldots v_{n-1}} = \varphi \frac{u_0}{v_0} \frac{u_1}{v_1} \ldots \frac{u_{n-1}}{v_{n-1}} \frac{t_0'}{u_0} \frac{t_1'}{u_1} \ldots \frac{t_{n-1}'}{u_{n-1}}.$$

Thus the simultaneous substitutions can be seen as successive substitutions, and the order of the substitutions $\frac{t_i}{u_i}$ may be permuted without affecting the final outcome. We may use the substitution rule repeatedly:

$$\begin{array}{ll}
\varphi \frac{t_0 \ldots t_{n-1}}{v_0 \ldots v_{n-1}} & \varphi \frac{t_0 \ldots t_{n-1}}{v_0 \ldots v_{n-1}} \\
\varphi \frac{u_0}{v_0} \ldots \frac{u_{n-1}}{v_{n-1}} \frac{t_0}{u_0} \ldots \frac{t_{n-1}}{u_{n-1}} & \varphi \frac{u_0}{v_0} \ldots \frac{u_{n-1}}{v_{n-1}} \frac{t_0}{u_0} \ldots \frac{t_{n-1}}{u_{n-1}} \\
\varphi \frac{u_0}{v_0} \ldots \frac{u_{n-1}}{v_{n-1}} \frac{t_0}{u_0} \ldots \frac{t_{n-1}}{u_{n-1}} \; t_{n-1} \equiv t_{n-1}' & \varphi \frac{u_0}{v_0} \ldots \frac{u_{n-1}}{v_{n-1}} \frac{t_0}{u_0} \ldots \frac{t_{n-1}'}{u_{n-1}} \\
\vdots & \\
\varphi \frac{u_0}{v_0} \ldots \frac{u_{n-1}}{v_{n-1}} \frac{t_0}{u_0} \ldots \frac{t_{n-1}}{u_{n-1}} \; t_{n-1} \equiv t_{n-1}' \ldots t_0 \equiv t_0' & \varphi \frac{u_0}{v_0} \ldots \frac{u_{n-1}}{v_{n-1}} \frac{t_0'}{u_0} \ldots \frac{t_{n-1}'}{u_{n-1}} \\
\varphi \frac{t_0 \ldots t_{n-1}}{v_0 \ldots v_{n-1}} \; t_0 \equiv t_0' \ldots t_{n-1} \equiv t_{n-1}' & \varphi \frac{t_0' \ldots t_{n-1}'}{v_0 \ldots v_{n-1}}.
\end{array}$$

□

10 Consistency

> *Vor Allem aber möchte ich unter den zahlreichen Fragen, welche hinsichtlich der Axiome gestellt werden können, dies als das wichtigste Problem bezeichnen, zu beweisen, daß dieselben untereinander widerspruchslos sind, d.h. daß man auf Grund derselben mittelst einer endlichen Anzahl von logischen Schlüssen niemals zu Resultaten gelangen kann, die miteinander in Widerspruch stehen.*
> David Hilbert

Fix a language S.

Definition 48. *A set* $\Phi \subseteq L^S$ *is* consistent *if* $\Phi \nvdash \bot$. Φ *is* inconsistent *if* $\Phi \vdash \bot$.

We prove some laws of consistency.

Lemma 49. *Let $\Phi \subseteq L^S$ and $\varphi \in L^S$. Then*

a) Φ is inconsistent iff there is $\psi \in L^S$ such that $\Phi \vdash \psi$ and $\Phi \vdash \neg\psi$.

b) $\Phi \vdash \varphi$ iff $\Phi \cup \{\neg\varphi\}$ is inconsistent.

c) If Φ is consistent, then $\Phi \cup \{\varphi\}$ is consistent or $\Phi \cup \{\neg\varphi\}$ is consistent (or both).

d) Let $\mathcal{F}$ be a family of consistent sets which is linearly ordered by inclusion, i.e., for all $\Phi, \Psi \in \mathcal{F}$ holds $\Phi \subseteq \Psi$ or $\Psi \subseteq \Phi$. Then

$$\Phi^* = \bigcup_{\Phi \in \mathcal{F}} \Phi$$

is consistent.

Proof. a) Assume $\Phi \vdash \bot$. Then by the *ex falso* rule, $\Phi \vdash \psi$ and $\Phi \vdash \neg\psi$.

Conversely assume that $\Phi \vdash \psi$ and $\Phi \vdash \neg\psi$ for some $\psi \in L^S$. Then $\Phi \vdash \bot$ by $\bot$-introduction.

b) Assume $\Phi \vdash \varphi$. Take $\varphi_0, \ldots, \varphi_{n-1} \in \Phi$ such that $\varphi_0 \ldots \varphi_{n-1} \vdash \varphi$. Then we can extend a derivation of $\varphi_0 \ldots \varphi_{n-1} \vdash \varphi$ as follows

$$\begin{array}{lllll}
\varphi_0 & \cdots & \varphi_{n-1} & & \varphi \\
\varphi_0 & \cdots & \varphi_{n-1} & \neg\varphi & \neg\varphi \\
\varphi_0 & \cdots & \varphi_{n-1} & \neg\varphi & \bot
\end{array}$$

and $\Phi \cup \{\neg\varphi\}$ is inconsistent.

Conversely assume that $\Phi \cup \{\neg\varphi\} \vdash \bot$ and take $\varphi_0, \ldots, \varphi_{n-1} \in \Phi$ such that $\varphi_0 \ldots \varphi_{n-1} \neg \varphi \vdash \bot$. Then $\varphi_0 \ldots \varphi_{n-1} \vdash \varphi$ and $\Phi \vdash \varphi$.

c) Assume that $\Phi \cup \{\varphi\}$ and $\Phi \cup \{\neg\varphi\}$ are inconsistent. Then there are $\varphi_0, \ldots, \varphi_{n-1} \in \Phi$ such that $\varphi_0 \ldots \varphi_{n-1} \vdash \varphi$ and $\varphi_0 \ldots \varphi_{n-1} \vdash \neg\varphi$. By the introduction rule for $\bot$, $\varphi_0 \ldots \varphi_{n-1} \vdash \bot$. Thus Φ is inconsistent.

d) Assume that Φ^* is inconsistent. Take $\varphi_0, \ldots, \varphi_{n-1} \in \Phi^*$ such that $\varphi_0 \ldots \varphi_{n-1} \vdash \bot$. Take $\Phi_0, \ldots \Phi_{n-1} \in \mathcal{F}$ such that $\varphi_0 \in \Phi_0, \ldots, \varphi_{n-1} \in \Phi_{n-1}$. Since $\mathcal{F}$ is linearly ordered by inclusion there is $\Phi \in \{\Phi_0, \ldots \Phi_{n-1}\}$ such that $\varphi_0, \ldots, \varphi_{n-1} \in \Phi$. Then Φ is inconsistent, contradiction. □

The proof of the completeness theorem will be based on the relation between consistency and satisfiability.

Lemma 50. *Assume that $\Phi \subseteq L^S$ is satisfiable. Then Φ is consistent.*

Proof. Assume that $\Phi \vdash \bot$. By the correctness of the sequent calculus, $\Phi \vDash \bot$. Assume that Φ is satisfiable and let $\mathfrak{M} \vDash \Phi$. Then $\mathfrak{M} \vDash \bot$. This contradicts the definition of the satisfaction relation. Thus Φ is not satisfiable. □

We shall later show the converse of this Lemma, since:

Theorem 51. *The sequent calculus is complete iff every consistent $\Phi \subseteq L^S$ is satisfiable.*

Proof. Assume that the sequent calculus is complete. Let $\Phi \subseteq L^S$ be consistent, i.e., $\Phi \nvdash \bot$. By completeness, $\Phi \nvDash \bot$, and we can take an S-model $\mathfrak{M} \vDash \Phi$ such that $\mathfrak{M} \nvDash \bot$. Thus Φ is satisfiable.

Conversely, assume that every consistent $\Phi \subseteq L^S$ is satisfiable. Assume $\Psi \vDash \psi$. Assume for a contradiction that $\Psi \nvdash \psi$. Then $\Psi \cup \{\neg\psi\}$ is consistent. By assumption there is an S-model $\mathfrak{M} \vDash \Psi \cup \{\neg\psi\}$. $\mathfrak{M} \vDash \Psi$ and $\mathfrak{M} \nvDash \psi$, which contradicts $\Psi \vDash \psi$. Thus $\Psi \vdash \psi$. □

11 Term models and HENKIN sets

The following constructions will assume that the class of all terms of some language is a set. In view of the previous lemma, we strive to construct interpretations for given sets $\Phi \subseteq L^S$ of S-formulas. Since we are working in great generality and abstractness, the only material available for the construction of structures is the language L^S itself. We shall build a model out of S-terms.

Definition 52. *Let S be a language and let $\Phi \subseteq L^S$ be consistent. The* term model $\mathfrak{T}^\Phi$ *of Φ is the following S-model:*

a) Define a relation $\sim$ on T^S,

$$t_0 \sim t_1 \text{ iff } \Phi \vdash t_0 \equiv t_1 .$$

$\sim$ is an equivalence relation on T^S.

b) For $t \in T^S$ let $\bar{t} = \{s \in T^S \mid s \sim t\}$ be the equivalence class of t.

c) The underlying set $T^\Phi = \mathfrak{T}^\Phi(\forall)$ of the term model is the set of $\sim$-equivalence classes

$$T^\Phi = \{\bar{t} \mid t \in T^S\}.$$

d) For an n-ary relation symbol $R \in S$ let $R^{\mathfrak{T}^\Phi}$ on T^Φ be defined by

$$(\bar{t}_0, ..., \bar{t}_{n-1}) \in R^{\mathfrak{T}^\Phi} \text{ iff } \Phi \vdash R t_0 ... t_{n-1} .$$

e) For an n-ary function symbol $f \in S$ let $f^{\mathfrak{T}^\Phi}$ on T^Φ be defined by

$$f^{\mathfrak{T}^\Phi}(\bar{t}_0, ..., \bar{t}_{n-1}) = \overline{f t_0 ... t_{n-1}} .$$

f) For $n \in \mathbb{N}$ define the variable interpretation $\mathfrak{T}^\Phi(v_n) = \overline{v_n}$.

The term model is well-defined.

Lemma 53. *In the previous construction the following holds:*

a) $\sim$ is an equivalence relation on T^S.

b) The definition of $R^{\mathfrak{T}^\Phi}$ is independent of representatives.

c) The definition of $f^{\mathfrak{T}^\Phi}$is independent of representatives.

Proof. a) We derived the axioms of equivalence relations for $\equiv$:

- $\vdash \forall x\, x \equiv x$
- $\vdash \forall x \forall y\, (x \equiv y \rightarrow y \equiv x)$
- $\vdash \forall x \forall y \forall z\, (x \equiv y \wedge y \equiv z \rightarrow x \equiv z)$

Consider $t \in T^S$. Then $\vdash t \equiv t$. Thus for all $t \in T^S$ holds $t \sim t$.

Consider $t_0, t_1 \in T^S$ with $t_0 \sim t_1$. Then $\vdash t_0 \equiv t_1$. Also $\vdash t_0 \equiv t_1 \rightarrow t_1 \equiv t_0$, $\vdash t_1 \equiv t_0$, and $t_1 \sim t_0$. Thus for all $t_0, t_1 \in T^S$ with $t_0 \sim t_1$ holds $t_1 \sim t_0$.

The transitivity of $\sim$ follows similarly.

b) Let $\bar{t}_0, ..., \bar{t}_{n-1} \in T^\Phi$, $\bar{t}_0 = \bar{s}_0, ..., \bar{t}_{n-1} = \bar{s}_{n-1}$ and $\Phi \vdash R t_0 ... t_{n-1}$. Then $\vdash t_0 \equiv s_0$, ... , $\vdash t_{n-1} \equiv s_{n-1}$. Repeated applications of the substitution rule yield $\Phi \vdash R s_0 ... s_{n-1}$. Hence $\Phi \vdash R t_0 ... t_{n-1}$ implies $\Phi \vdash R s_0 ... s_{n-1}$. By the symmetry of the argument, $\Phi \vdash R t_0 ... t_{n-1}$ iff $\Phi \vdash R s_0 ... s_{n-1}$.

c) Let $\bar{t}_0, ..., \bar{t}_{n-1} \in T^\Phi$ and $\bar{t}_0 = \bar{s}_0, ..., \bar{t}_{n-1} = \bar{s}_{n-1}$. Then $\vdash t_0 \equiv s_0$, ... , $\vdash t_{n-1} \equiv s_{n-1}$. Repeated applications of the substitution rule to $\vdash f t_0 ... t_{n-1} \equiv f t_0 ... t_{n-1}$ yield

$$\vdash f t_0 ... t_{n-1} \equiv f s_0 ... s_{n-1}$$

and $\overline{ft_0...t_{n-1}} = \overline{fs_0...s_{n-1}}$. □

We aim to obtain $\mathfrak{T}^\Phi \vDash \Phi$. The initial cases of an induction over the complexity of formulas is given by

Theorem 54.

a) For terms $t \in T^S$ holds $\mathfrak{T}^\Phi(t) = \bar{t}$.

b) For atomic formulas $\varphi \in L^S$ holds

$$\mathfrak{T}^\Phi \vDash \varphi \textit{ iff } \Phi \vdash \varphi.$$

Proof. a) By induction on the term calculus. The initial case $t = v_n$ is obvious by the definition of the term model. Now consider a term $t = ft_0...t_{n-1}$ with an n-ary function symbol $f \in S$, and assume that the claim is true for $t_0, ..., t_{n-1}$. Then

$$\begin{aligned}\mathfrak{T}^\Phi(ft_0...t_{n-1}) &= f^{\mathfrak{T}^\Phi}(\mathfrak{T}^\Phi(t_0), ..., \mathfrak{T}^\Phi(t_{n-1})) \\ &= f^{\mathfrak{T}^\Phi}(\bar{t_0}, ..., \overline{t_{n-1}}) \\ &= \overline{ft_0...t_{n-1}}.\end{aligned}$$

b) Let $\varphi = Rt_0...t_{n-1}$ with an n-ary relation symbol $R \in S$ and $t_0, ..., t_{n-1} \in T^S$. Then

$$\begin{aligned}\mathfrak{T}^\Phi \vDash Rt_0...t_{n-1} &\text{ iff } R^{\mathfrak{T}^\Phi}(\mathfrak{T}^\Phi(t_0), ..., \mathfrak{T}^\Phi(t_{n-1})) \\ &\text{ iff } R^{\mathfrak{T}^\Phi}(\bar{t_0}, ..., \overline{t_{n-1}}) \\ &\text{ iff } \Phi \vdash Rt_0...t_{n-1}.\end{aligned}$$

Let $\varphi = t_0 \equiv t_1$ with $t_0, t_1 \in T^S$. Then

$$\begin{aligned}\mathfrak{T}^\Phi \vDash t_0 \equiv t_1 &\text{ iff } \mathfrak{T}^\Phi(t_0) = \mathfrak{T}^\Phi(t_1) \\ &\text{ iff } \bar{t_0} = \bar{t_1} \\ &\text{ iff } t_0 \sim t_1 \\ &\text{ iff } \Phi \vdash t_0 \equiv t_1.\end{aligned}$$

□

To extend the lemma to complex S-formulas, Φ has to satisfy some recursive properties.

Definition 55. *A set $\Phi \subseteq L^S$ of S-formulas is a* HENKIN set *if it satisfies the following properties:*

a) Φ is consistent;

b) Φ is (derivation) complete*, i.e., for all $\varphi \in L^S$*

$$\Phi \vdash \varphi \textit{ or } \Phi \vdash \neg\varphi;$$

c) Φ contains witnesses*, i.e., for all $\forall x\varphi \in L^S$ there is a term $t \in T^S$ such that*

$$\Phi \vdash \neg\forall x\varphi \to \neg\varphi\frac{t}{x}.$$

Lemma 56. *Let $\Phi \subseteq L^S$ be a* HENKIN *set. Then for all $\chi, \psi \in L^S$ and variables x:*

a) $\Phi \nvdash \chi$ iff $\Phi \vdash \neg\chi$.

b) $\Phi \vdash \chi$ implies $\Phi \vdash \psi$, iff $\Phi \vdash \chi \to \psi$.

c) *For all* $t \in T^S$ *holds* $\Phi \vdash \chi \frac{t}{u}$ *iff* $\Phi \vdash \forall x \chi$.

Proof. a) Assume $\Phi \nvdash \chi$. By derivation completeness, $\Phi \vdash \neg\chi$. Conversely assume $\Phi \vdash \neg\chi$. Assume for a contradiction that $\Phi \vdash \chi$. Then Φ is inconsistent. Contradiction. Thus $\Phi \nvdash \chi$.
b) Assume $\Phi \vdash \chi$ implies $\Phi \vdash \psi$.
Case 1. $\Phi \vdash \chi$. Then $\Phi \vdash \psi$ and by an easy derivation $\Phi \vdash \chi \rightarrow \psi$.
Case 2. $\Phi \nvdash \chi$. By the derivation completeness of Φ holds $\Phi \vdash \neg\chi$. And by an easy derivation $\Phi \vdash \chi \rightarrow \psi$.

Conversely assume that $\Phi \vdash \chi \rightarrow \psi$. Assume that $\Phi \vdash \chi$. By $\rightarrow$-elimination, $\Phi \vdash \psi$. Thus $\Phi \vdash \chi$ implies $\Phi \vdash \psi$.
c) Assume that for all $t \in T^S$ holds $\Phi \vdash \chi \frac{t}{u}$. Assume that $\Phi \nvdash \forall x \chi$. By a), $\Phi \vdash \neg\forall x \chi$. Since Φ contains witnesses there is a term $t \in T^S$ such that $\Phi \vdash \neg\forall x \chi \rightarrow \neg\chi \frac{t}{u}$. By $\rightarrow$-elimination, $\Phi \vdash \neg\chi \frac{t}{u}$. Contradiction. Thus $\Phi \vdash \forall x \chi$. The converse follows from the rule of $\forall$-elimination. □

Theorem 57. *Let* $\Phi \subseteq L^S$ *be a* Henkin *set. Then*

a) *For all formulas* $\chi \in L^S$, *pairwise distinct variables* $\vec{x}$ *and terms* $\vec{t} \in T^S$

$$\mathfrak{T}^\Phi \vDash \chi \frac{\vec{t}}{\vec{x}} \text{ iff } \Phi \vdash \chi \frac{\vec{t}}{\vec{x}} .$$

b) $\mathfrak{T}^\Phi \vDash \Phi$.

Proof. b) follows immediately from a). a) is proved by induction on the formula calculus. The atomic case has already been proven. Consider the non-atomic cases:
i) $\chi = \bot$. Then $\bot \frac{\vec{t}}{\vec{x}} = \bot$. $\mathfrak{T}^\Phi \vDash \bot \frac{\vec{t}}{\vec{x}}$ is false by definition of the satisfaction relation $\vDash$, and $\Phi \vdash \chi \frac{\vec{t}}{\vec{x}}$ is false since Φ is consistent. Thus $\mathfrak{T}^\Phi \vDash \bot \frac{\vec{t}}{\vec{x}}$ iff $\Phi \vdash \bot \frac{\vec{t}}{\vec{x}}$.
ii.) $\chi = \neg\varphi \frac{\vec{t}}{\vec{x}}$ and assume that the claim holds for φ. Then

$$\begin{aligned} \mathfrak{T}^\Phi \vDash \neg\varphi \frac{\vec{t}}{\vec{x}} & \text{ iff not } \mathfrak{T}^\Phi \vDash \varphi \frac{\vec{t}}{\vec{x}} \\ & \text{ iff not } \Phi \vdash \varphi \frac{\vec{t}}{\vec{x}} \text{ by the inductive assumption} \\ & \text{ iff } \Phi \vdash \neg\varphi \frac{\vec{t}}{\vec{x}} \text{ by a) of the previous lemma.} \end{aligned}$$

iii.) $\chi = (\varphi \rightarrow \psi) \frac{\vec{t}}{\vec{x}}$ and assume that the claim holds for φ and ψ. Then

$$\begin{aligned} \mathfrak{T}^\Phi \vDash (\varphi \rightarrow \psi) \frac{\vec{t}}{\vec{x}} & \text{ iff } \mathfrak{T}^\Phi \vDash \varphi \frac{\vec{t}}{\vec{x}} \text{ implies } \mathfrak{T}^\Phi \vDash \psi \frac{\vec{t}}{\vec{x}} \\ & \text{ iff } \Phi \vdash \varphi \frac{\vec{t}}{\vec{x}} \text{ implies } \Phi \vdash \psi \frac{\vec{t}}{\vec{x}} \text{ by the inductive assumption} \\ & \text{ iff } \Phi \vdash \varphi \frac{\vec{t}}{\vec{x}} \rightarrow \psi \frac{\vec{t}}{\vec{x}} \text{ by a) of the previous lemma} \\ & \text{ iff } \Phi \vdash (\varphi \rightarrow \psi) \frac{\vec{t}}{\vec{x}} \text{ by the definition of substitution.} \end{aligned}$$

iv.) $\chi = (\forall x \varphi) \frac{t_0 \ldots t_{r-1}}{x_0 \ldots x_{r-1}}$ and assume that the claim holds for φ. By definition of the substitution χ is of the form

$$\forall u \, (\varphi \frac{t_0 \ldots t_{r-1} \, u}{x_0 \ldots x_{r-1} \, x}) \text{ oder } \forall u \, (\varphi \frac{t_1 \ldots t_{r-1} \, u}{x_1 \ldots x_{r-1} \, x})$$

with a suitable variable u. Without loss of generality assume that χ is of the first form. Then

$$\begin{aligned}
\mathfrak{T}^{\Phi}\vDash(\forall x\varphi)\frac{\vec{t}}{\vec{x}} \ & \text{iff} \ \ \mathfrak{T}^{\Phi}\vDash\exists u\,(\varphi\frac{t_0\ldots t_{r-1}\,u}{x_0\ldots x_{r-1}\,x})\\
& \text{iff} \ \ \text{for all } t\in T^S \text{ holds } \mathfrak{T}^{\Phi}\frac{\bar{t}}{u}\vDash\varphi\frac{t_0\ldots t_{r-1}\,u}{x_0\ldots x_{r-1}\,x}\\
& \text{iff} \ \ \text{for all } t\in T^S \text{ holds } \mathfrak{T}^{\Phi}\frac{\mathfrak{I}^{\Phi}(t)}{u}\vDash\varphi\frac{t_0\ldots t_{r-1}\,u}{x_0\ldots x_{r-1}\,x} \ \ \text{by a previous lemma}\\
& \text{iff} \ \ \text{for all } t\in T^S \text{ holds } \mathfrak{T}^{\Phi}\vDash(\varphi\frac{t_0\ldots t_{r-1}}{x_0\ldots x_{r-1}})\frac{t}{u} \ \ \text{by the substitution lemma}\\
& \text{iff} \ \ \text{for all } t\in T^S \text{ holds } \mathfrak{T}^{\Phi}\vDash\varphi\frac{t_0\ldots t_{r-1}\,t}{x_0\ldots x_{r-1}\,x} \ \ \text{by successive substitutions}\\
& \text{iff} \ \ \text{for all } t\in T^S \text{ holds } \Phi\vdash\varphi\frac{t_0\ldots t_{r-1}\,t}{x_0\ldots x_{r-1}\,x} \ \ \text{by the inductive assumption}\\
& \text{iff} \ \ \text{for all } t\in T^S \text{ holds } \Phi\vdash(\varphi\frac{t_0\ldots t_{r-1}\,u}{x_0\ldots x_{r-1}\,x})\frac{t}{u} \ \ \text{by successive substitutions}\\
& \text{iff} \ \ \Phi\vdash\forall u\,(\varphi\frac{t_0\ldots t_{r-1}\,u}{x_0\ldots x_{r-1}\,x}) \ \ \text{by c) of the previous lemma}\\
& \text{iff} \ \ \Phi\vdash(\forall x\varphi)\frac{\vec{t}}{\vec{x}}\,.
\end{aligned}$$

□

12 Constructing HENKIN sets

We shall show that every consistent set of formulas can be extended to a HENKIN set by “adding witnesses” and then ensuring negation completeness. We first consider witnesses.

Theorem 58. *Let $\Phi\subseteq L^S$ be consistent. Let $\varphi\in L^S$ and let z be a variable which does not occur in $\Phi\cup\{\varphi\}$. Then the set*

$$\Phi\cup\{\neg\forall x\varphi\rightarrow\neg\varphi\frac{z}{x}\}$$

is consistent.

Proof. Assume for a contradiction that $\Phi\cup\{(\neg\exists x\varphi\vee\varphi\frac{z}{x})\}$ is inconsistent. Take $\varphi_0,\ldots,\varphi_{n-1}\in\Phi$ such that

$$\varphi_0\ldots\varphi_{n-1}\ \neg\forall x\varphi\rightarrow\neg\varphi\frac{z}{x}\vdash\bot\,.$$

Set $\Gamma=(\varphi_0,\ldots,\varphi_{n-1})$. Then continue the derivation as follows:

1.	Γ	$\neg\forall x\varphi\rightarrow\neg\varphi\frac{z}{x}$	$\bot$
2.	Γ	$\neg\neg\forall x\varphi$	$\neg\neg\forall x\varphi$
3.	Γ	$\neg\neg\forall x\varphi$	$\neg\forall x\varphi\rightarrow\neg\varphi\frac{z}{x}$
4.	Γ	$\neg\neg\forall x\varphi$	$\bot$
5.	Γ		$\neg\forall x\varphi$
6.	Γ	$\neg\varphi\frac{z}{x}$	$\neg\varphi\frac{z}{x}$
7.	Γ	$\neg\varphi\frac{z}{x}$	$\neg\forall x\varphi\rightarrow\neg\varphi\frac{z}{x}$
8.	Γ	$\neg\varphi\frac{z}{x}$	$\bot$
9.	Γ		$\varphi\frac{z}{x}$
10.	Γ		$\forall x\varphi$
11.	Γ		$\bot$

Hence Φ is inconsistent, contradiction. □

This means that "unused" variables may be used as HENKIN witnesses. Since "unused" constant symbols behave much like unused variables, we get:

Theorem 59. *Let $\Phi \subseteq L^S$ be consistent. Let $\varphi \in L^S$ and let $c \in S$ be a constant symbol which does not occur in $\Phi \cup \{\varphi\}$. Then the set*

$$\Phi \cup \{\neg\forall x\varphi \to \neg\varphi\frac{c}{x}\}$$

is consistent.

Proof. Assume that $\Phi \cup \{(\neg\exists x\varphi \vee \varphi\frac{c}{x})\}$ is inconsistent. Take a derivation

$$\begin{array}{c} \Gamma_0\varphi_0 \\ \Gamma_1\varphi_1 \\ \vdots \\ \Gamma_{n-1}\,\varphi_{n-1} \\ \Gamma_n\ (\neg\forall x\varphi \to \neg\varphi\frac{c}{x})\ \bot \end{array} \tag{1}$$

with $\Gamma_n \subseteq \Phi$. Choose a variable z, which does not occur in the derivation. For a formula ψ define ψ' by replacing each occurence of c by z, and for a sequence $\Gamma = (\psi_0, ..., \psi_{k-1})$ of formulas let $\Gamma' = (\psi'_0, ..., \psi'_{k-1})$. Replacing each occurence of c by z in the deriavation we get

$$\begin{array}{c} \Gamma'_0\varphi'_0 \\ \Gamma'_1\varphi'_1 \\ \vdots \\ \Gamma'_{n-1}\,\varphi'_{n-1} \\ \Gamma_n\ (\neg\forall x\varphi \to \neg\varphi\frac{z}{x})\ \bot \end{array} \tag{2}$$

The particular form of the final sequence is due to the fact that c does not occur in $\Phi \cup \{\varphi\}$. To show that (2) is again a derivation in the sequent calculus we show that the replacement $c \mapsto z$ transforms every instance of a sequent rule in (1) into an instance of a (derivable) rule in (2). This is obvious for all rules except possibly the quantifyer rules.

So let

$$\frac{\Gamma\ \ \psi\frac{y}{x}}{\Gamma\ \ \forall x\psi}\ \text{, with } y \notin \text{free}(\Gamma \cup \{\forall x\psi\})$$

be an $\forall$-introduction in (1). Then $(\psi\frac{y}{x})' = \psi'\frac{y}{x}$, $(\forall x\psi)' = \forall x\psi'$, and $y \notin \text{free}(\Gamma' \cup \{(\forall x\psi)'\})$. Hence

$$\frac{\Gamma'\ \ (\psi\frac{y}{x})'}{\Gamma'\ \ (\forall x\psi)'}$$

is a justified $\forall$-introduction.

Now consider an $\forall$-elimination in (1):

$$\frac{\Gamma\ \ \forall x\psi}{\Gamma\ \ \psi\frac{t}{x}}$$

Then $(\forall x\psi)' = \forall x\psi'$ and $(\psi\frac{t}{x})' = \psi'\frac{t'}{x}$ where t' is obtained from t by replacing all occurences of c by z. Hence

$$\frac{\Gamma'\ \ (\forall x\psi)'}{\Gamma'\ \ (\psi\frac{t}{x})'}$$

is a justified ∀-elimination.

The derivation (2) proves that

$$\Phi \cup \left\{ \left(\neg \forall x \varphi \rightarrow \neg \varphi \frac{z}{x} \right) \right\} \vdash \bot,$$

which contradicts the preceding lemma. □

We shall now show that any consistent set of formulas can be consistently expanded to a set of formulas which contains witnesses.

Theorem 60. *Let S be a language and let $\Phi \subseteq L^S$ be consistent. Then there is a language S^ω and $\Phi^\omega \subseteq L^{S^\omega}$ such that*

a) S^ω extends S by constant symbols, i.e., $S \subseteq S^\omega$ and if $s \in S^\omega \setminus S$ then s is a constant symbol;

b) $\Phi^\omega \supseteq \Phi$;

c) Φ^ω is consistent;

d) Φ^ω contains witnesses;

e) if L^S is countable then so are L^{S^ω} and Φ^ω.

Proof. For every a define a "new" distinct constant symbol c_a, which does not occur in S, e.g., $c_a = ((a, S), 1, 0)$. Extend S by constant symbols c_ψ for $\psi \in L^S$:

$$S^+ = S \cup \{c_\psi | \psi \in L^S\}.$$

Then set

$$\Phi^+ = \Phi \cup \{\neg \forall x \varphi \rightarrow \neg \varphi \frac{c_{\forall x \varphi}}{x} | \forall x \varphi \in L^S\}.$$

Φ^+ contains witnesses for all universal formulas of S.
(1) $\Phi^+ \subseteq L^{S^+}$ is consistent.
Proof: Assume instead that Φ^+ is inconsistent. Choose a finite sequence $\forall x_0 \varphi_0$, ..., $\forall x_{n-1} \varphi_{n-1} \in L^S$ of pairwise distinct universal formulas such that

$$\Phi \cup \{\neg \forall x_0 \varphi_0 \rightarrow \neg \varphi_0 \frac{c_{\forall x_0 \varphi_0}}{x_0}, \ldots, \neg \forall x_{n-1} \varphi_{n-1} \rightarrow \neg \varphi_{n-1} \frac{c_{\forall x_{n-1} \varphi_{n-1}}}{x_{n-1}}\}$$

is inconsistent. By the previous theorem one can inductively show that for all $i < n$ the set

$$\Phi \cup \{\neg \forall x_0 \varphi_0 \rightarrow \neg \varphi_0 \frac{c_{\forall x_0 \varphi_0}}{x_0}, \ldots, \neg \forall x_{n-1} \varphi_{n-1} \rightarrow \neg \varphi_{n-1} \frac{c_{\forall x_{i-1} \varphi_{\mathrm{ni}-1}}}{x_{i-1}}\}$$

is consistent. Contradiction. *qed*(1)

We iterate the +-operation through the integers. Define recursively

$$\begin{aligned}
\Phi^0 &= \Phi \\
S^0 &= S \\
S^{n+1} &= (S^n)^+ \\
\Phi^{n+1} &= (\Phi^n)^+ \\
S^\omega &= \bigcup_{n \in \mathbb{N}} S^n \\
\Phi^\omega &= \bigcup_{n \in \mathbb{N}} \Phi^n.
\end{aligned}$$

S^{ω} is an extension of S by constant symbols. For $n \in \mathbb{N}$, Φ^n is consistent by induction. Φ^{ω} is consistent by the lemma on unions of consistent sets.
(2) Φ^{ω} contains witnesses.
Proof. Let $\forall x \varphi \in L^{S^{\omega}}$. Let $n \in \mathbb{N}$ such that $\forall x \varphi \in L^{S^n}$. Then $\neg \forall x \varphi \rightarrow \neg \varphi \frac{c_{\forall x \varphi}}{x} \in \Phi^{n+1} \subseteq \Phi^{\omega}$. *qed*(2)
(3) Let L^S be countable. Then $L^{S^{\omega}}$ and Φ^{ω} are countable.
Proof. Since L^S is countable, there can only be countably many symbols in the alphabet of $S^0 = S$. The alphabet of S^1 is obtained by adding the countable set $\{c_{\psi} | \psi \in L^S\}$; the alphabet of S^1 is countable as the union of two countable sets. The set of words over a countable alphabet is countable, hence L^{S^1} and $\Phi^1 \subseteq L^{S^1}$ are countable.

Inductive application of this argument show that for any $n \in \mathbb{N}$, the sets L^{S^n} and Φ^n are countable. Since countable unions of countable sets are countable, $L^{S^{\omega}} = \bigcup_{n \in \mathbb{N}} L^{S^n}$ and also $\Phi^{\omega} \subseteq L^{S^{\omega}}$ are countable. □

Exercise 10. Let S be a countable language, let $\Phi \subseteq L^S$ be consistent, and let $\mathrm{Var} \setminus \mathrm{Var}(\Phi)$ be infinite. Then there exists $\Phi^{\omega} \subseteq L^S$ such that

a) $\Phi^{\omega} \supseteq \Phi$;

b) Φ^{ω} is consistent;

c) Φ^{ω} contains witnesses.

To get Henkin sets we have to ensure derivation completeness.

Theorem 61. *Let S be a language and let $\Phi \subseteq L^S$ be consistent. Then there is a consistent $\Phi^* \subseteq L^S$, $\Phi^* \supseteq \Phi$ which is derivation complete.*

Proof. Define the partial order $(P, \subseteq)$ by

$$P = \{\Psi \subseteq L^S \,|\, \Psi \supseteq \Phi \text{ and } \Psi \text{ is consistent}\}.$$

$P \neq \emptyset$ since $\Phi \in P$. P is *inductively ordered* by a previous lemma: if $\mathcal{F} \subseteq P$ is linearly ordered by inclusion, i.e., for all $\Psi, \Psi' \in \mathcal{F}$ holds $\Psi \subseteq \Psi'$ or $\Psi' \subseteq \Psi$ then

$$\bigcup_{\Psi \in \mathcal{F}} \Psi \in P.$$

Hence $(P, \subseteq)$ satisfies the conditions of Zorn's lemma. Let Φ^* be a maximal element of $(P, \subseteq)$. By the definition of P, $\Phi^* \subseteq L^S$, $\Phi^* \supseteq \Phi$, and Φ^* is consistent. Derivation completeness follows from the following claim.
(1) For all $\varphi \in L^S$ holds $\varphi \in \Phi^*$ or $\neg \varphi \in \Phi^*$.
Proof. Φ^* is consistent. By a previous lemma, $\Phi^* \cup \{\varphi\}$ or $\Phi^* \cup \{\neg \varphi\}$ are consistent.
Case 1. $\Phi^* \cup \{\varphi\}$ is consistent. By the $\subseteq$-maximality of Φ^*, $\Phi^* \cup \{\varphi\} = \Phi^*$ and $\varphi \in \Phi^*$.
Case 2. $\Phi^* \cup \{\neg \varphi\}$ is consistent. By the $\subseteq$-maximality of Φ^*, $\Phi^* \cup \{\neg \varphi\} = \Phi^*$ and $\neg \varphi \in \Phi^*$. □

The proof uses Zorn's lemma. In case L^S is countable one can work without Zorn's lemma.

Proof. (For countable L^S) Let $L^S = \{\varphi_n | n \in \mathbb{N}\}$ be an enumeration of L^S. Define a sequence $(\Phi_n | n \in \mathbb{N})$ by recursion on n such that

i. $\Phi \subseteq \Phi_n \subseteq \Phi_{n+1} \subseteq L^S$;

ii. Φ_n is consistent.

For $n = 0$ set $\Phi_0 = \Phi$. Assume that Φ_n is defined according to i. and ii.
Case 1. $\Phi_n \cup \{\varphi_n\}$ is consistent. Then set $\Phi_{n+1} = \Phi_n \cup \{\varphi_n\}$.
Case 2. $\Phi_n \cup \{\varphi_n\}$ is inconsistent. Then $\Phi_n \cup \{\neg \varphi_n\}$ is consistent by a previous lemma, and we define $\Phi_{n+1} = \Phi_n \cup \{\neg \varphi_n\}$.

Let

$$\Phi^* = \bigcup_{n \in \mathbb{N}} \Phi_n .$$

Then Φ^* is a consistent superset of Φ. By construction, $\varphi \in \Phi^*$ or $\neg\varphi \in \Phi^*$, for all $\varphi \in L^S$. Hence Φ^* is derivation complete. □

According to Theorem 60 a given consistent set Φ can be extended to $\Phi^\omega \subseteq L^{S^\omega}$ containing witnesses. By Theorem 61 Φ^ω can be extended to a derivation complete $\Phi^* \subseteq L^{S^\omega}$. Since the latter step does not extend the language, Φ^* contains witnesses and is thus a henkin set:

Theorem 62. *Let S be a language and let $\Phi \subseteq L^S$ be consistent. Then there is a language S^* and $\Phi^* \subseteq L^{S^*}$ such that*

a) $S^* \supseteq S$ is an extension of S by constant symbols;

b) $\Phi^* \supseteq \Phi$ is a Henkin *set;*

c) if L^S is countable then so are L^{S^*} and Φ^*.

13 The completeness theorem

> *The development of mathematics towards greater precision has led, as is well known, to the formalization of large tracts of it, so that one can prove any theorem using nothing but a few mechanical rules.*
> Kurt Gödel, 1941

We can now combine our technical preparations to show the fundamental theorems of first-order logic. Combining Theorems 62 and 57, we obtain a general and a countable model existence theorem:

Theorem 63. (Henkin model existence theorem) *Let $\Phi \subseteq L^S$. Then Φ is consistent iff Φ is satisfiable.*

By Lemma 51, Theorem 63 the model existence theorems imply the main theorem.

Theorem 64. (Gödel completeness theorem) *The sequent calculus is complete, i.e., $\vDash = \vdash$.*

The Gödel completeness theorem is the *fundamental theorem of mathematical logic*. It connects syntax and semantics of formal languages in an optimal way. Before we continue the mathematical study of its consequences we make some general remarks about the wider impact of the theorem:

- The completeness theorem gives an *ultimate correctness criterion* for mathematical proofs. A proof is correct if it can (in principle) be reformulated as a formal derivation. Although mathematicians prefer semi-formal or informal arguments, this criterion could be applied in case of doubt.

- Checking the correctness of a formal proof in the above sequent calculus is a syntactic task that can be carried out by computer. We shall later consider a prototypical *proof checker* `Naproche` which uses a formal language which is a subset of natural english.
- By systematically running through all possible formal proofs, *automatic theorem proving* is in principle possible. In this generality, however, algorithms immediately run into very high algorithmic complexities and become practically infeasable.
- Practical automatic theorem proving has become possible in restricted situations, either by looking at particular kinds of axioms and associated intended domains, or by restricting the syntactical complexity of axioms and theorems.
- Automatic theorem proving is an important component of *artificial intelligence* (AI) where a system has to obtain logical consequences from conditions formulated in first-order logic. Although there are many difficulties with artificial intelligence this approach is still being followed with some success.
- Another special case of automatic theorem proving is given by *logic programming* where programs consist of logical statements of some restricted complexity and a run of a program is a systematic search for a solution of the given statements. The original and most prominent logic programming language is `Prolog` which is still widely used in linguistics and AI.
- There are other areas which can be described formally and where syntax/semantics constellations similar to first-order logic may occur. In the theory of algorithms there is the syntax of programming languages versus the (mathematical) meaning of a program. Since programs crucially involve time alternative logics with time have to be introduced. Now in all such generalizations, the GÖDEL completeness theorem serves as a pattern onto which to model the syntax/semantics relation.
- The success of the formal method in mathematics makes mathematics a leading *formal science*. Several other sciences also strive to present and justify results formally, like computer science and parts of philosophy.
- The completeness theorem must not be confused with the famous GÖDEL *incompleteness theorems*: they say that certain axiom systems like PEANO arithmetic are incomplete in the sense that they do not imply some formulas which hold in the standard model of the axiom system.

14 The compactness theorem

The equality of $\vDash$ and $\vdash$ and the compactness theorem 43 for $\vdash$ imply

Theorem 65. (Compactness theorem) *Let $\Phi \subseteq L^S$ and $\varphi \in \Phi$. Then*

a) $\Phi \vDash \varphi$ iff there is a finite subset $\Phi_0 \subseteq \Phi$ such that $\Phi_0 \vDash \varphi$.

b) Φ is satisfiable iff every finite subset $\Phi_0 \subseteq \Phi$ is satisfiable.

This theorem is often to construct (unusual) models of first-order theories. It is the basis of a field of logic called *Model Theory*.

We present a number theoretic application of the compactness theorem. The language of arithmetic can be naturally interpreted in the structure $\mathbb{N} = (\mathbb{N}, +, \cdot, 0, 1)$. This structure obviously satisfies the following axioms:

Definition 66. *The axiom system* PA $\subseteq L^{S_{\mathrm{AR}}}$ *of* PEANO arithmetic *consists of the following sentences*

- $\forall x\, x+1 \neq 0$
- $\forall x \forall y\; x+1 = y+1 \rightarrow x = y$
- $\forall x\, x+0 = x$
- $\forall x \forall y\; x+(y+1) = (x+y)+1$
- $\forall x\, x \cdot 0 = 0$
- $\forall x \forall y\; x \cdot (y+1) = x \cdot y + x$
- *Schema of induction: for every formula* $\varphi(x_0, ..., x_{n-1}, x_n) \in L^{S_{\mathrm{AR}}}$*:*

$$\forall x_0 ... \forall x_{n-1} (\varphi(x_0, ..., x_{n-1}, 0) \wedge \forall x_n (\varphi \rightarrow \varphi(x_0, ..., x_{n-1}, x_n + 1)) \rightarrow \forall x_n\, \varphi)$$

The theory PA is allows to prove a lot of number theoretic properties, e.g., about divisibility and prime numbers. On the other hand the first *incompleteness theorem* of GÖDEL shows that there are arithmetic sentences φ which are not decided by PA although they are true in the standard model $\mathbb{N}$ of PA. Therefore PA is *not* complete.

If φ and $\neg\varphi$ are both not derivable from PA then PA $+ \neg\varphi$ and PA $+ \varphi$ are consistent. By the model existence theorem, there are models $\mathfrak{M}^-$ and $\mathfrak{M}^+$ such that $\mathfrak{M}^- \vDash$ PA $+ \neg\varphi$ and $\mathfrak{M}^+ \vDash$ PA $+ \varphi$. $\mathfrak{M}^-$ and $\mathfrak{M}^+$ are not isomorphic. So there exist models of PA which are not isomorphic to the standard model $\mathbb{N}$.

We can also use the compactness theorem to obtain nonstandard models of theories. Define the S_{AR}-terms $\bar{n}$ for $n \in \mathbb{N}$ recursively by

$$\begin{aligned} \bar{0} &= 0, \\ \overline{n+1} &= (\bar{n} + 1). \end{aligned}$$

Define divisibility by the S_{AR}-formula $\delta(x, y) = \exists z\; x \cdot z \equiv y$.

Theorem 67. *There is a model* $\mathfrak{M} \vDash$ PA *which contains an element* $\infty \in M$*,* $\infty \neq \bar{0}^{\mathfrak{M}}$ *such that* $\mathfrak{M} \vDash \delta(\bar{n}, \infty)$ *for every* $n \in \mathbb{N} \setminus \{0\}$ *(we use* $\mathfrak{M} \vDash \delta(\bar{n}, \infty)$ *as an intuitive abbreviation for* $\mathfrak{M} \vDash \delta(\bar{n}, v_0)[\infty]$*).*

So "from the outside", ∞ is divisible by every positive natural number. This implies that $\mathfrak{M}$ is a nonstandard model with $\mathfrak{M} \ncong \mathbb{N}$.

Proof. Consider the theory

$$\Phi = \mathrm{PA} \cup \{\delta(\bar{n}, v_0) \mid n \in \mathbb{N} \setminus \{0\}\} \cup \{\neg v_0 \equiv \bar{0}\}$$

(1) Φ is satisfiable.
Proof. We use the compactness theorem 65(b). Let $\Phi_0 \subseteq \Phi$ be finite. It suffices to show that Φ_0 is satisfiable. Take a finite number $n_0 \in \mathbb{N}$ such that

$$\Phi_0 \subseteq \mathrm{PA} \cup \{\delta(\bar{n}, v_0) \mid n \in \mathbb{N}, 1 \leqslant n \leqslant n_0\}.$$

Let $N = n!$. Then

$$\mathbb{N} \vDash \mathrm{PA} \text{ and } \mathbb{N} \vDash \delta(\bar{n}, N) \text{ for } 1 \leqslant n \leqslant n_0.$$

So $\mathbb{N}\frac{N}{v_0} \vDash \Phi_0$. *qed*(1)

By (1), let $\mathfrak{M}' \vDash \Phi$. Let $\infty = \mathfrak{M}'(v_0) \in |\mathfrak{M}'|$. Let $\mathfrak{M}$ be the S_{AR}-structure which extends to the model $\mathfrak{M}'$, i.e., $\mathfrak{M} = \mathfrak{M}' \upharpoonright \{\forall\} \cup S_{\mathrm{AR}}$. Then $\mathfrak{M}$ is a structure satisfying the theorem. □

This indicates that the model class of PA is rather complicated and rich. Indeed there is a subfield of model theory which studies models of Peano arithmetic.

We define notions which allow to examine the axiomatizability of classes of structures.

Definition 68. *Let S be a language and $\mathcal{K}$ be a class of S-structures.*

a) *$\mathcal{K}$ ist* elementary *or* finitely axiomatizable *if there is an S-sentence φ with $\mathcal{K} = \mathrm{Mod}^S\varphi$.*

b) *$\mathfrak{K}$ is* Δ-elementary *or* axiomatizable*, if there is a set Φ of S-sentences with $\mathcal{K} = \mathrm{Mod}^S\Phi$.*

We state simple properties of the Mod-operator:

Theorem 69. *Let S be a language.Then*

a) *For $\Phi \subseteq \Psi \subseteq L_0^S$ holds $\mathrm{Mod}^S\Phi \supseteq \mathrm{Mod}^S\Psi$.*

b) *For $\Phi, \Psi \subseteq L_0^S$ holds $\mathrm{Mod}^S(\Phi \cup \Psi) = \mathrm{Mod}^S\Phi \cap \mathrm{Mod}^S\Psi$.*

c) *For $\Phi \subseteq L_0^S$ holds $\mathrm{Mod}^S\Phi = \bigcap_{\varphi \in \Phi} \mathrm{Mod}^S\varphi$.*

d) *For $\varphi_0, \ldots, \varphi_{n-1} \in L_0^S$ holds $\mathrm{Mod}^S\{\varphi_0, \ldots, \varphi_{n-1}\} = \mathrm{Mod}^S(\varphi_0 \wedge \ldots \wedge \varphi_{n-1})$.*

e) *For $\varphi \in L_0^S$ holds $\mathrm{Mod}^S(\neg\varphi) = \mathrm{Mod}^S(\emptyset) \setminus \mathrm{Mod}^S(\varphi)$.*

c) explains the denotation "Δ-elementary", since $\mathrm{Mod}^S\Phi$ is the intersection ("**D**urchschnitt") of all single $\mathrm{Mod}^S\varphi$.

Theorem 70. *Let S be a language and $\mathcal{K}, \mathcal{L}$ be classes of S-structures with*

$$\mathcal{L} = \mathrm{Mod}^S\emptyset \setminus \mathcal{K}.$$

Then if $\mathcal{K}$ and $\mathcal{L}$ are axiomatizable, they are finitely axiomatizable.

Proof. Take axiom systems Φ_K and Φ_L such that $\mathfrak{K} = \mathrm{Mod}^S\Phi_K$ and $\mathfrak{L} = \mathrm{Mod}^S\Phi_L$. Assume that $\mathfrak{K}$ is not finitely axiomatizable.
(1) Let $\Phi_0 \subseteq \Phi_K$ be finite. Then $\Phi_0 \cup \Phi_L$ is satisfiable.
Proof: $\mathrm{Mod}^S\Phi_0 \supseteq \mathrm{Mod}^S\Phi_K$. Since $\mathfrak{K}$ is not finitely axiomatizable, $\mathrm{Mod}^S\Phi_0 \neq \mathrm{Mod}^S\Phi_K$. Then $\mathrm{Mod}^S\Phi_0 \cap \mathfrak{L} \neq \emptyset$. Take a model $\mathfrak{A} \in \mathfrak{L}$, $\mathfrak{A} \in \mathrm{Mod}^S\Phi_0$. Then $\mathfrak{A} \vDash \Phi_0 \cup \Phi_L$. *qed*(1)
(2) $\Phi_K \cup \Phi_L$ is satisfiable.
Proof: By the compactness theorem 65 it suffices to show that every finite $\Psi \subseteq \Phi_K \cup \Phi_L$ is satsifiable. By (1), $(\Psi \cap \Phi_K) \cup \Phi_L$ is satisfiable. Thus $\Psi \subseteq (\Psi \cap \Phi_K) \cup \Phi_L$ is satisfiable. *qed*(2)

By (2), $\mathrm{Mod}^S\Phi_K \cap \mathrm{Mod}^S\Phi_L \neq \emptyset$. But the classes $\mathfrak{K}$ and $\mathfrak{L}$ are complements, contradiction. Thus $\mathfrak{K}$ is finitely axiomatizable. □

15 The Löwenheim-Skolem theorems

Definition 71. *An S-structure $\mathfrak{A}$ is* finite, infinite, countable*, or* uncountable*, resp., iff the underlying set $|\mathfrak{A}|$ is finite, infinite, countable, or uncountable, resp..*

If the language S is countable, i.e., finite or countably infinite, and it $\Phi \subseteq L^S$ is a *countable* consistent set of formulas then an inspection of the above construction of a term model for Φ shows the following theorem.

Theorem 72. (Downward LÖWENHEIM-SKOLEM theorem) *Let $\Phi \subseteq L^S$ be a* countable *consistent set of formulas. Then Φ possesses a model $\mathfrak{M} = (\mathfrak{A}, \beta) \vDash \Phi$, $\mathfrak{A} = (A, \ldots)$ with a* countable *underlying set A.*

The word "downward" emphasises the existence of models of "small" cardinality. We shall soon consider an "upward" LÖWENHEIM-SKOLEM theorem.

Theorem 73. *Assume that $\Phi \subseteq L^S$ has arbitrarily large finite models. Then Φ has an infinite model.*

Proof. For $n \in \mathbb{N}$ define the sentence

$$\varphi_{\geqslant n} = \exists v_0, \ldots, v_{n-1} \bigwedge_{i<j<n} \neg v_i \equiv v_j ,$$

where the big conjunction is defined by

$$\bigwedge_{i<j<n} \psi_{ij} = \psi_{0,1} \wedge \ldots \wedge \psi_{0,n-1} \wedge \psi_{1,2} \wedge \ldots \wedge \psi_{1,n-1} \wedge \ldots \wedge \psi_{n-1,n-1} .$$

For any model $\mathfrak{M}$

$$\mathfrak{M} \vDash \varphi_{\geqslant n} \text{ iff } A \text{ has at least } n \text{ elements.}$$

Now set

$$\Phi' = \Phi \cup \{\varphi_{\geqslant n} \,|\, n \in \mathbb{N}\}.$$

(1) Φ' has a model.
Proof. By the compactness theorem 65b it suffices to show that every finite $\Phi_0 \subseteq \Phi$ has a model. Let $\Phi_0 \subseteq \Phi$ be finite. Take $n_0 \in \mathbb{N}$ such that

$$\Phi_0 \subseteq \Phi \cup \{\varphi_{\geqslant n} \,|\, n \leqslant n_0\}.$$

By assumption Φ has a model with at least n_0 elements. Thus $\Phi \cup \{\varphi_{\geqslant n} \,|\, n \leqslant n_0\}$ and Φ_0 have a model. *qed*(1)

Let $\mathfrak{M} \vDash \Phi'$. Then $\mathfrak{M}$ is an infinite model of Φ. □

Theorem 74. (Upward LÖWENHEIM-SKOLEM theorem) *Let $\Phi \subseteq L^S$ have an infinite S-model and let X be an arbitrary set. Then Φ has a model into which X can be embedded injectively.*

Proof. Let $\mathfrak{M}$ be an infinite model of Φ. Choose a sequence $(c_x \,|\, x \in X)$ of pairwise distinct constant symbols which do not occur in S, e.g., setting $c_x = ((x, S), 1, 0)$. Let $S' = S \cup \{c_x \,|\, x \in X\}$ be the extension of S by the new constant symbols. Set

$$\Phi' = \Phi \cup \{\neg c_x \equiv c_y \,|\, x, y \in X, x \neq y\}.$$

(1) Φ' has a model.
Proof. It suffices to show that every finite $\Phi_0 \subseteq \Phi'$ has a model. Let $\Phi_0 \subseteq \Phi'$ be finite. Take a finite set $X_0 \subseteq X$ such that

$$\Phi_0 \subseteq \Phi \cup \{\neg c_x \equiv c_y \,|\, x, y \in X_0, x \neq y\}.$$

Since $|\mathfrak{M}|$ is infinite we can choose an injective sequence $(a_x | x \in X_0)$ of elements of $|\mathfrak{M}|$ such that $x \neq y$ implies $a_x \neq a_y$. For $x \in X \setminus X_0$ choose $a_x \in |\mathfrak{M}|$ arbitrarily. Then in the extended model

$$\mathfrak{M}' = \mathfrak{M} \cup \{(c_x, a_x) | x \in X\} \vDash \Phi \cup \{\neg c_x \equiv c_y \,|\, x, y \in X_0, x \neq y\} \supseteq \Phi_0 .$$

qed(1)

By (1), choose a model $\mathfrak{M}' \vDash \Phi'$. Then the map

$$i: X \to |\mathfrak{M}'|, x \mapsto \mathfrak{M}'(c_x)$$

is injective. The reduct $\mathfrak{M}'' = \mathfrak{M}' \upharpoonright \{\forall\} \cup S$ is as required. □

Theorem 75. *Let S be a language.*

a) *The class of all finite S-structures is not axiomatizable.*

b) *The class of all infinite S-structures is axiomatizable but not finitely axiomatizable.*

Proof. a) is immediate by Theorem 73.
b) The class of infinite S-structures is axiomatized by

$$\Phi = \{\varphi_{\geqslant n} \,|\, n \in \mathbb{N}\}.$$

If that class were *finitely* axiomatizable then the complementary class of finite S-structures would also be (finitely) axiomatizable, contradicting a). □

16 Normal forms

There are many motivations to transform formulas into equivalent *normal forms*. The motivation here will be that normal forms are important for *automated theorem proving* and for *logic programming*.

We are particularly interested in transforming formulas ψ into formulas ψ' such that ψ is consistent iff ψ' is consistent. This relates to provability as follows: $\Phi \vdash \varphi$ iff $\Phi \cup \{\neg\varphi\}$ is not satisfiable/inconsistent. So a check for provability can be based on inconsistency checks.

Work in some fixed language S.

Definition 76.

a) *An S-formula is a* literal *if it is atomic or the negation of an atomic formula.*

b) *Define the* dual *of the literal L as*

$$\bar{L} = \begin{cases} \neg L, \text{ if } L \text{ is an atomic formula;} \\ K, \text{ if } L \text{ is of the form } \neg K. \end{cases}$$

c) *A formula φ is in* disjunctive normal form *if it is of the form*

$$\varphi = \bigvee_{i<m} \Big(\bigwedge_{j<n_i} L_{ij} \Big)$$

where each L_{ij} is a literal.

d) *A formula φ is in* conjunctive normal form *if it is of the form*

$$\varphi = \bigwedge_{i<m} \Big(\bigvee_{j<n_i} L_{ij} \Big)$$

where each L_{ij} is a literal. Sometimes a disjunctive normal form is also written in set notion as

$$\varphi = \{\{L_{00}, \ldots, L_{0n_0-1}\}, \ldots, \{L_{m-1,0}, \ldots, L_{m-1,n_{m-1}}\}\}.$$

Theorem 77. *Let φ be a formula without quantifiers. Then φ is equivalent to some φ' in disjunctive normal form and to some φ'' in conjunctive normal form.*

Proof. By induction on the complexity of φ. Clear for φ atomic. The $\neg$ step follows from the de Morgan laws:

$$\begin{aligned}\neg \bigvee_{i<m} (\bigwedge_{j<n_i} L_{ij}) &\leftrightarrow \bigwedge_{i<m} \neg(\bigwedge_{j<n_i} L_{ij}) \\ &\leftrightarrow \bigwedge_{i<m} (\bigvee_{j<n_i} \neg L_{ij}).\end{aligned}$$

The $\wedge$-step is clear for conjunctive normal forms. For disjunctive normal forms the associativity rules yield

$$\bigvee_{i<m} (\bigwedge_{j<n_i} L_{ij}) \wedge \bigvee_{i<m'} (\bigwedge_{j<n_i'} L_{ij}') \leftrightarrow \bigvee_{i<m, i'<m'} (\bigwedge_{j<n_i} L_{ij} \wedge \bigwedge_{j<n_i'} L_{ij}')$$

which is also in conjunctive normal form. □

Definition 78. *A formula φ is in* prenex normal form *if it is of the form*

$$\varphi = Q_0\, x_0\, Q_1\, x_1 ... Q_{m-1}\, x_{m-1}\, \psi$$

where each Q_i is either the quantifier $\forall$ or $\exists$ and ψ is quantifier-free. Then the quantifier string $Q_0\, x_0\, Q_1\, x_1 ... Q_{m-1}\, x_{m-1}$ is called the prefix *of φ and the formula ψ is the* matrix *of φ.*

Theorem 79. *Let φ be a formula. Then φ is equivalent to a formula φ' in prenex normal form.*

Proof. By induction on the complexity of φ. Clear for atomic formulas. If

$$\varphi \leftrightarrow Q_0\, x_0\, Q_1\, x_1 ... Q_{m-1}\, x_{m-1}\, \psi$$

with quantifier-free ψ then by the de Morgan laws for quantifiers

$$\neg\varphi \leftrightarrow \bar{Q}_0\, x_0\, \bar{Q}_1\, x_1 ...\, \bar{Q}_{m-1} x_{m-1}\, \neg\psi$$

where the dual quantifier $\bar{Q}$ is defined by $\bar{\exists} = \forall$ and $\bar{\forall} = \exists$.

For the $\wedge$-operation consider another formula

$$\varphi' \leftrightarrow Q_0'\, x_0'\, Q_1'\, x_1' ... Q_{m'-1}'\, x_{m'-1}'\, \psi'$$

with quantifier-free ψ'. We may assume that the bound variables of of the prenex normal forms are disjoint. Then

$$\varphi \wedge \varphi' \leftrightarrow Q_0\, x_0\, Q_1\, x_1 ... Q_{m-1}\, x_{m-1} Q_0'\, x_0'\, Q_1'\, x_1' ... Q_{m'-1}'\, x_{m'-1}'\, (\psi \wedge \psi').$$

(semantic argument). □

Definition 80. *A formula φ is* universal *if it is of the form*

$$\varphi = \forall x_0\, \forall x_1 ... \forall x_{m-1}\, \psi$$

where ψ is quantifier-free. A formula φ is existential *if it is of the form*

$$\varphi = \exists x_0\, \exists x_1 ... \exists x_{m-1}\, \psi$$

where ψ is quantifier-free.

We show a quasi-equivalence with respect to universal (and existential) formulas which is not a logical equivalence but concerns the consistency or satisfiability of formulas.

Theorem 81. *Let φ be an S-formula. Then there is a canonical extension S^* of the language S and a canonical universal $\varphi^* \in L^{S^*}$ such that*

$$\varphi \text{ is consistent iff } \varphi^* \text{ is consistent.}$$

The formula φ^* is called the Skolem normal form *of φ.*

Proof. By a previous theorem we may assume that φ is in prenex normal form. We prove the theorem by induction on the number of existential quantifiers in φ. If φ does not contain an existential quantifier we are done. Otherwise let

$$\varphi = \forall x_1 \dots \forall x_m \exists y \psi$$

where $m < \omega$ may also be 0. Introduce a new m-ary function symbol f (or a constant symbol in case $m = 0$) and let

$$\varphi' = \forall x_1 \dots \forall x_m \, \psi \frac{f x_1 \dots x_m}{y}.$$

By induction it suffices to show that φ is consistent iff φ' is consistent.
(1) $\varphi' \rightarrow \varphi$.
Proof. Assume φ'. Consider $x_1, \dots, x_m$. Then $\psi \frac{f x_1 \dots x_m}{y}$. Then $\exists y \psi$. Thus $\forall x_1 \dots \forall x_m \exists y \psi$. *qed*(1)
(2) If φ' is consistent then φ is consistent.
Proof. If $\varphi \rightarrow \bot$ then by (1) $\varphi' \rightarrow \bot$. *qed*(2)
(3) If φ is consistent then φ' is consistent.
Proof. Let φ be consistent and let $\mathcal{M} = (M, \dots) \vDash \varphi$. Then

$$\forall a_1 \in M \dots \forall a_m \in M \, \exists b \in M \mathcal{M} \frac{\vec{a} \; b}{\vec{x} \; y} \vDash \psi.$$

Using the axiom of choice there is a function $h: M^m \rightarrow M$ such that

$$\forall a_1 \in M \dots \forall a_m \in M \mathcal{M} \frac{\vec{a} \;\; h(a_1, \dots, a_m)}{\vec{x} \;\; y} \vDash \psi.$$

Expand the structure $\mathcal{M}$ to $\mathcal{M}' = \mathcal{M} \cup \{(f, h)\}$ where the symbol f is interpreted by the function h. Then $h(a_1, \dots, a_m) = \mathcal{M}' \frac{\vec{a}}{\vec{x}} (f x_1 \dots x_m)$ and

$$\forall a_1 \in M \dots \forall a_m \in M \mathcal{M}' \frac{\vec{a} \;\; \mathcal{M}'\frac{\vec{a}}{\vec{x}}(f x_1 \dots x_m)}{\vec{x} \;\; y} = \mathcal{M}' \frac{\vec{a}}{\vec{x}} \frac{\mathcal{M}'\frac{\vec{a}}{\vec{x}}(f x_1 \dots x_m)}{y} \vDash \psi.$$

By the substitution theorem this is equivalent to

$$\forall a_1 \in M \dots \forall a_m \in M \, \mathcal{M}' \frac{\vec{a}}{\vec{x}} \vDash \psi \frac{f x_1 \dots x_m}{y}.$$

Hence

$$\mathcal{M}' \vDash \forall x_1 \dots \forall x_m \, \psi \frac{f x_1 \dots x_m}{y} = \varphi'.$$

Thus φ' is consistent. □

17 Herbrand's theorem

By the previous chapter we can reduce the question whether a given finite set of formulas is inconsistent to the question whether some universal formula is inconsistent. By the following theorem this can be answered rather concretely.

Theorem 82. *Let S be a language which contains at least one constant symbol. Let*

$$\varphi = \forall x_0 \forall x_1 ... \forall x_{m-1} \psi$$

be a universal S-sentence with quantifier-free matrix ψ. Then φ is inconsistent if there are variable-free S-terms ("constant terms")

$$t_0^0, ..., t_{m-1}^0, ..., t_0^{N-1}, ..., t_{m-1}^{N-1}$$

such that

$$\varphi' = \bigwedge_{i<N} \psi \frac{t_0^i, ..., t_{m-1}^i}{x_0, ..., x_{m-1}} = \psi \frac{t_0^0, ..., t_{m-1}^0}{x_0, ..., x_{m-1}} \wedge ... \wedge \psi \frac{t_0^{N-1}, ..., t_{m-1}^{N-1}}{x_0, ..., x_{m-1}}$$

is inconsistent.

Proof. All sentences φ', for various choices of constant terms, are logical consequences of φ. So φ is consistent, all φ' are consistent.

Conversely assume that all φ' are consistent. Then by the compactness theorem

$$\Phi = \{\psi \frac{t_0, ..., t_{m-1}}{x_0, ..., x_{m-1}} \,|\, t_0, ..., t_{m-1} \text{ are constant } S\text{-terms}\}$$

is consistent. Let $\mathcal{M} = (M, ...) \vDash \Phi$. Let

$$H = \{\mathcal{M}(t) \,|\, t \text{ is a constant } S\text{-term}\}.$$

Then $H \neq \emptyset$ since S contains a constant symbol. By definition, H is closed under the functions of $\mathcal{M}$. So we let $\mathcal{H} = (H, ...) \subseteq \mathcal{M}$ be the substructure of $\mathcal{M}$ with domain H.
(1) $\mathcal{H} \vDash \varphi$.
Proof. Let $\mathcal{M}(t_0)$, ..., $\mathcal{M}(t_{m-1}) \in H$ where t_0, ..., t_{m-1} are constant S-terms. Then $\psi \frac{t_0, ..., t_{m-1}}{x_0, ..., x_{m-1}} \in \Phi$, $\mathcal{M} \vDash \psi \frac{t_0, ..., t_{m-1}}{x_0, ..., x_{m-1}}$, and by the substitution theorem

$$\mathcal{M} \frac{\mathcal{M}(t_0), ..., \mathcal{M}(t_{m-1})}{x_0, ..., x_{m-1}} \vDash \psi.$$

Since ψ is quantifier-free this transfers to $\mathcal{H}$:

$$\mathcal{H} \frac{\mathcal{M}(t_0), ..., \mathcal{M}(t_{m-1})}{x_0, ..., x_{m-1}} \vDash \psi.$$

Thus

$$\mathcal{H} \vDash \forall x_0 \forall x_1 ... \forall x_{m-1} \psi = \varphi.$$

qed(1)
Thus φ is consistent. □

In case that the formula ψ does not contain the equality sign $\equiv$ checking for inconsistency of

$$\varphi' = \bigwedge_{i<N} \psi \frac{t_0^i, ..., t_{m-1}^i}{x_0, ..., x_{m-1}} = \psi \frac{t_0^0, ..., t_{m-1}^0}{x_0, ..., x_{m-1}} \wedge ... \wedge \psi \frac{t_0^{N-1}, ..., t_{m-1}^{N-1}}{x_0, ..., x_{m-1}}$$

is in principle a straightforward finitary problem. φ' contains finitely many constant S-terms. φ' is consistent iff the relation symbols can be interpreted on appropriate tuples of the occuring S-terms to make φ' true. There are finitely many possibilities for the assignments of truth values of relations. This leads to the following (theoretical) algorithm for automatic proving for formulas without $\equiv$:

Let $\Omega \subseteq L^S$ be finite and $\chi \in L^S$. To check whether $\Omega \vdash \chi$:

1. Form $\Phi = \Omega \cup \{\neg\chi\}$ and let $\varphi = \forall(\bigwedge \Phi)$ be the universal closure of $\bigwedge \Phi$. Then $\Omega \vdash \chi$ iff $\Phi = \Omega \cup \{\neg\chi\}$ is inconsistent iff $(\bigwedge \Phi) \vdash \bot$ iff $\forall(\bigwedge \Phi) \vdash \bot$.

2. Transform φ into universal form $\varphi^{\forall} = \forall x_0 \forall x_1 ... \forall x_{m-1} \psi$ (Skolemization).
3. (Systematically) search for constant S-terms
$$t_0^0, ..., t_{m-1}^0, ..., t_0^{N-1}, ..., t_{m-1}^{N-1}$$
such that
$$\varphi' = \bigwedge_{i<N} \psi \frac{t_0^i, ..., t_{m-1}^i}{x_0, ..., x_{m-1}} = \psi \frac{t_0^0, ..., t_{m-1}^0}{x_0, ..., x_{m-1}} \wedge ... \wedge \psi \frac{t_0^{N-1}, ..., t_{m-1}^{N-1}}{x_0, ..., x_{m-1}}$$
is inconsistent.
4. If an inconsistent φ' is found, output "yes", otherwise carry on.

Obviously, if "yes" is output then $\Omega \vdash \chi$. This is the *correctness* of the algorithm. On the other hand, Herbrand's theorem ensures that if $\Omega \vdash \chi$ then an appropriate φ' will be found, and "yes" will be output, i.e., the algorithm is *complete*.

Example 83. We demonstrate the procedure with a small example. Let
$$\chi = \exists x \forall y (D(x) \rightarrow D(y))$$
be the well-known *drinker's paradox*: there is somebody called x such that everybody drinks provided x drinks. To prove χ we follow the above steps.

1. χ is valid iff $\neg\chi$ is inconsistent. $\neg\chi$ is equivalent to $\forall x \exists y (D(x) \wedge \neg D(y))$.
2. The Skolemization of that formula is $\forall x (D(x) \wedge \neg D(f_y(x)))$.
3. Ground terms without free variables can be formed from a new constant symbol c and the unary function symbol f_y: $c, f_y(c), f_y(f_y(c)), ...$. We form the corresponding ground instances of the kernel $D(x) \wedge \neg D(f_y(x))$:
$$D(c) \wedge \neg D(f_y(c)), D(f_y(c)) \wedge \neg D(f_y(f_y(c))), D(f_y(f_y(c))) \wedge \neg D(f_y(f_y(f_y(c)))), ...$$
This leads to a sequence of conjunctions of ground instances:
   - $D(c) \wedge \neg D(f_y(c))$ is consistent since the conjunction does not contain dual literals;
   - $D(c) \wedge \neg D(f_y(c)) \wedge D(f_y(c)) \wedge \neg D(f_y(f_y(c)))$ is **inconsistent** since the conjunction contains the dual literals $\neg D(f_y(c))$ and $D(f_y(c))$.

This concludes the proof of the drinker's paradox via Herbrand's theorem.

18 Simple automatic theorem proving

The syntax of first-order logic works with finite strings of symbols and is amenable to computer implementation. The *Handbook of Practical Logic and Automated Reasoning* by John Harrison, Cambridge University Press 2009, contains working programs which define basic notions of first-order logic including the above proof method based on Herbrand's theorem. The programs are written in the functional programming language OCaml. OCaml programs consist of commands which are similar to mathematical definitions of constants and functions. The OCaml programs of the *Handbook* are available via the website `http://www.cl.cam.ac.uk/~jrh13/atp/index.html`.

Harrison defines the type (or class or set) of terms recursively as:

```
type term = Var of string
          | Fn of string * term list;;
```

(Relational) atomic formulas are given by

```
type fol = R of string * term list;;
```

The atomic formula $D(x)$ of the drinker's formula is thus

```
R("D",[Var "x"])
```

Formulas are defined generally over any type `('a)` of atomic formulas

```
type ('a)formula = False
                 | True
                 | Atom of 'a
                 | Not of ('a)formula
                 | And of ('a)formula * ('a)formula
                 | Or of ('a)formula * ('a)formula
                 | Imp of ('a)formula * ('a)formula
                 | Iff of ('a)formula * ('a)formula
                 | Forall of string * ('a)formula
                 | Exists of string * ('a)formula;;
```

First-order formulas in particular are those of type `fol formula` where `fol` has been defined above. The drinker's paradox is expressed in this language as

```
Exists("x",Forall("y",Imp(Atom(R("D",[Var "x"])),Atom(R("D",[Var "y"])))))
```

To improve readability, a pretty printer for the type `fol formula` is defined and automatically invoked. Entering the drinker's formula into an interactive OCaml terminal results in the following dialogue:

```
# Exists("x",Forall("y",Imp(Atom(R("D",[Var "x"])),Atom(R("D",[Var
"y"])))));;
   - : fol formula = <<exists x.  forall y.  D(x) ==> D(y)>>
```

The result of the query is output in the bottom line: the input is recognized by type inference as a first-order formula; the value of the input is then reprinted readably by the pretty printer; the `<< >>` quotes signal the "pretty format". Formulas can also be input in the pretty format. We define the drinker's formula χ by:

```
# let chi = <<exists x.  forall y.  D(x) ==> D(y)>>;;
val chi : fol formula = <<exists x.  forall y.  D(x) ==> D(y)>>
```

The definition of χ is acknowledged by stating that the value of `chi` is now the drinker's formula. In the Herbrand procedure, a formula to be proved is generalized, negated and skolemized. These operations are implemented by certain functions defined by HARRISON:

```
# generalize chi;;
- : fol formula = <<exists x.  forall y.  D(x) ==> D(y)>>
# Not (generalize chi);;
- : fol formula = <<exists x.  forall y.  D(x) ==> D(y)>>
# skolemize (Not (generalize chi));;
- : fol formula = <<D(x) /\ ~D(f_y(x))>>
```

The Paul Gilmore in the 1950's was one of the first to implement the Herbrand procedure. Gilmore's algorithm is implemented in the *Handbook*:

```
let gilmore_loop =
  let mfn djs0 ifn djs =
    filter (non trivial) (distrib (image (image ifn) djs0) djs) in
  herbloop mfn (fun djs -> djs <> []);;

let gilmore fm =
  let sfm = skolemize(Not(generalize fm)) in
  let fvs = fv sfm and consts,funcs = herbfuns sfm in
  let cntms = image (fun (c,_) -> Fn(c,[])) consts in
  length(gilmore_loop (simpdnf sfm) cntms funcs fvs 0 [[]] [] []);;
```

After generalizing, negating and skolemizing, constants and function symbols are extracted and used to generate ground terms and ground instances of the matrix of the skolemized formula in some recursive loop. Then an inconsistency is searched by transforming formulas into disjunctive normal form and checking for inconsistent literals. Details of these algorithms are explained in the the *Handbook*. We prove the drinker's paradox by applying the function gilmore to the formula chi:

```
# gilmore chi;;
0 ground instances tried; 1 items in list
0 ground instances tried; 1 items in list
1 ground instances tried; 1 items in list
1 ground instances tried; 1 items in list
- : int = 2
```

The output gives some information about the progress of the search procedure, in particular that it finishes with "search length" 2, and thus that the claim has been proved.

The Gilmore procedure is able to prove some simple, yet interesting results. The Russell paradox says that there is no x such that $x = \{y | y \notin y\}$, i.e., such that $\forall y(y \in x \leftrightarrow y \notin y)$. We can define and prove the Russell paradox:

```
# let russell = <<~(exists x.  forall y.  Elem(y,x) <=> ~Elem(y,y))>>;;
val russell : fol formula =
<<~(exists x.  forall y.  Elem(y,x) <=> ~Elem(y,y))>>
# gilmore russell;;
0 ground instances tried; 1 items in list
0 ground instances tried; 1 items in list
- : int = 1
```

Exercise 11. Apply the Herbrand procedure to the Russell paradox "by hand".

19 Resolution

The Gilmore procedure is theoretically complete: a formula is provable iff the procedure terminates. Termination can however take very long so that a proof will not be found in practice. Also there is an enormous amount of data to be stored which may cause the program to crash. E.g., the disjunctive normal forms in the `gilmore` program which can simply be checked for inconsistency seem to double in length with each iteration of the algorithm. Practical automatic theorem proving requires more efficient algorithms in order to narrow down the search space for inconsistencies and to keep data sizes small.

We shall now present another method based on *conjunctive normal forms*. We assume that the quantifier-free formula ψ is a conjunction of clauses $\psi = c_0 \wedge c_1 \wedge \ldots \wedge c_{l-1}$. Then $\forall x_0 \forall x_1 \ldots \forall x_{m-1} \psi$ is inconsistent iff the set

$$\{c_i \frac{t_0, \ldots, t_{m-1}}{x_0, \ldots, x_{m-1}} \,|\, t_0, \ldots, t_{m-1} \text{ are constant } S\text{-terms}\}$$

is inconsistent.

The method of *resolution* gives an efficient method for showing the inconsistency of sets of clauses. Let us assume until further notice, that the formulas considered do not contain the symbol $\equiv$.

Definition 84. *Let $c^+ = \{K_0, \ldots, K_{k-1}\}$ and $c^- = \{L_0, \ldots, L_{l-1}\}$ be clauses with literals K_i and L_j. Note that $\{K_0, \ldots, K_{k-1}\}$ stands for the disjunction $K_0 \vee \ldots \vee K_{k-1}$. Assume that K_0 and L_0 are* dual*, i.e., $L_0 = \overline{K_0}$. Then the disjunction*

$$\{K_1, \ldots, K_{k-1}\} \cup \{L_1, \ldots, L_{l-1}\}$$

is a resolution of c^+ and c^-.

Resolution is related to the application of modus ponens: $\varphi \rightarrow \psi$ and φ correspond to the clauses $\{\neg\varphi, \psi\}$ and $\{\varphi\}$. $\{\psi\}$ is a resolution of $\{\neg\varphi, \psi\}$ and $\{\varphi\}$.

Theorem 85. *Let C be a set of clauses and let c be a resolution of two clauses $c^+, c^- \in C$. Then if $C \cup \{c\}$ is inconsistent then C is inconsistent.*

Proof. Let $c^+ = \{K_0, \ldots, K_{k-1}\}$, $c^- = \{\neg K_0, L_1 \ldots, L_{l-1}\}$, and $c = \{K_1, \ldots, K_{k-1}\} \cup \{L_1, \ldots, L_{l-1}\}$. Assume that $\mathcal{M} \vDash C$ is a model of C.
Case 1. $\mathcal{M} \vDash K_0$. Then $\mathcal{M} \vDash c^-$, $\mathcal{M} \vDash \{L_1 \ldots, L_{l-1}\}$, and

$$\mathcal{M} \vDash \{K_1, \ldots, K_{k-1}\} \cup \{L_1, \ldots, L_{l-1}\} = c.$$

Case 2. $\mathcal{M} \vDash \neg K_0$. Then $\mathcal{M} \vDash c^+$, $\mathcal{M} \vDash \{K_1 \ldots, K_{k-1}\}$, and

$$\mathcal{M} \vDash \{K_1, \ldots, K_{k-1}\} \cup \{L_1, \ldots, L_{l-1}\} = c.$$

Thus $\mathcal{M} \vDash C \cup \{c\}$. □

Theorem 86. *Let C be a set of clauses closed under resolution. Then C is inconsistent iff $\emptyset \in C$. Note that the empty clause $\{\}$ is logically equivalent to $\perp$.*

Proof. If $\emptyset \in C$ then C is clearly inconsistent.

Assume that the converse implication is false. Consider an set C of clauses such that

$$(*) \quad C \text{ is inconsistent and closed under resolution, but } \emptyset \notin C.$$

By the compactness theorem there is a finite set of atomic formulas $\{\varphi_0, ..., \varphi_{n-1}\}$ such that

$$C' = \{c \in C | \text{for every literal } L \text{ in } c \text{ there exists } i < n \text{ such that } L = \varphi_i \text{ or } L = \neg\varphi_i\},$$

is also inconsistent. Since resolution only *deletes* atomic formulas, C' is also closed under resolution, and of course $\emptyset \notin C'$. So we may assume right away that there is only a finite set $\{\varphi_0, ..., \varphi_{n-1}\}$ of atomic formulas occuring in C, and that n with that property is chosen minimally.

From $n = 0$ atomic formulas one can only build the empty clause $\emptyset$. Since C is inconsistent, we must have $C \neq \emptyset$. Thus $C = \{\emptyset\}$ and $\emptyset \in C$, which contradicts $(*)$.

So we have $n = m + 1 > 0$. Let

$$C^+ = \{c \in C | \neg\varphi_m \notin c\}, C^- = \{c \in C | \varphi_m \notin c\}$$

and

$$C_0^+ = \{c \setminus \{\varphi_m\} | c \in C^+\}, C_0^- = \{c \setminus \{\neg\varphi_m\} | c \in C^-\}.$$

(1) C_0^+ and C_0^- are closed under resolution.
Proof. Let d'' be a resolution of $d, d' \in C_0^+$. Let $d = c \setminus \{\varphi_m\}$ and $d' = c' \setminus \{\varphi_m\}$ with c, $c' \in C^+$. The resolution d'' was based on some atomic formula $\varphi_i \neq \varphi_m$. Then we can also resolve c, c' by the same atomic formula φ_i. Let c'' be that resolution of c, c'. Since C is closed under resolution, $c'' \in C$, $c'' \in C^+$, and $d'' = c'' \setminus \{\varphi_m\} \in C_0^+$. *qed*(1)
(2) $\emptyset \notin C_0^+$ or $\emptyset \notin C_0^-$.
Proof. If $\emptyset \in C_0^+$ and $\emptyset \in C_0^-$, and since $\emptyset \notin C$ we have $\{\varphi_m\} \in C^+$ and $\{\neg\varphi_m\} \in C^-$. But then the resolution $\emptyset$ of $\{\varphi_m\}$ and $\{\neg\varphi_m\}$ would be in C, contradiction. *qed*(2)
Case 1. $\emptyset \notin C_0^+$. Since C_0^+ is formed by *removing* the atomic formula φ_m, C_0^+ only contains atomic formulas from $\{\varphi_0, ..., \varphi_{m-1}\}$. By the minimality of n and by (1), C_0^+ is consistent.

Let $\mathcal{M} \vDash C_0^+$. By the proof of the model existence theorem we may assume that $\mathcal{M}$ is a term model. Since the the equality sign $\equiv$ does not occur in C the term model can be formed without factoring the term in T^S by some equivalence relation. This means that different terms are interpreted by different elements of $|\mathcal{M}|$.

We can assume that Let the atomic formula φ_m be of the form $rt_0...t_{s-1}$ where r is an n-ary relation symbol and $t_0, ..., t_{s-1} \in T^S$. Since the formula $rt_0...t_{s-1}$ does not occur within C_0^+, we can modify the model $\mathcal{M}$ to a model $\mathcal{M}'$ by only modifying the interpretation $\mathcal{M}(r)$ exactly at $(\mathcal{M}(t_0), ..., \mathcal{M}(t_{s-1}))$. So let $\mathcal{M}'(r)(\mathcal{M}(t_0), ..., \mathcal{M}(t_{s-1}))$ be *false*. Then $\mathcal{M}' \vDash \neg\varphi_m$. We show that $\mathcal{M}' \vDash C$.

Let $c \in C$. If $\neg\varphi_m \in c$ then $\mathcal{M}' \vDash c$. So assume that $\neg\varphi_m \notin c$. Then $c \in C^+$ and $c \setminus \{\varphi_m\} \in C_0^+$. Then $\mathcal{M} \vDash c \setminus \{\varphi_m\}$, $\mathcal{M}' \vDash c \setminus \{\varphi_m\}$, and $\mathcal{M}' \vDash c$. But then C is consistent, contradiction.
Case 2. $\emptyset \notin C_0^-$. We can then proceed analogously to case 1, arranging that $\mathcal{M}'(\mathcal{M}(t_0), ..., \mathcal{M}(t_{s-1}))$ be *true*. So we get a contradiction again. □

This means that the inconsistency check in the automatic proving algorithm can be carried out even more systematically: produce all relevant resolution instances until the empty clause is generated. Again we have correctness and completeness for the algorithm with resolution.

Here is an implementation of resolution by Harrison; given a literal p, all pairs of clauses that contain p positively and negatively resp. are treated by resolution with respect to p.

```
let resolve_on p clauses =
  let p' = negate p and pos,notpos = partition (mem p) clauses in
  let neg,other = partition (mem p') notpos in
  let pos' = image (filter (fun l -> l <> p)) pos
```

```
    and neg' = image (filter (fun l -> l <> p')) neg in
    let res0 = allpairs union pos' neg' in
    union other (filter (non trivial) res0);;
```

The Davis-Putnam procedure uses resolution (and some other methods) to search for inconsisties.

```
  let davisputnam fm =
    let sfm = skolemize(Not(generalize fm)) in
    let fvs = fv sfm and consts,funcs = herbfuns sfm in
    let cntms = image (fun (c,_) -> Fn(c,[])) consts in
    length(dp_loop (simpcnf sfm) cntms funcs fvs 0 [] [] []);;
```

Like in the Gilmore procedure, the given formula is prepared and a loop is initiated to search for inconsistencies. Here the formula is put in *conjunctive normal form* and resolution is performed in some loop. One readily finds examples solvable by Davis-Putnam which Gilmore cannot prove in a short time.

20 Unifikation

Resolution is one of the main mechanisms behind the *logic programming language* Prolog. Prolog programs can be viewed as conjunctions of universally quantified clauses. A universally quantified clause stands for all clauses that can be reached by substituting into the free variable of the clause. Prolog searches systematically for clauses that can be resolved after substitution. Prolog uses “minimal” substitutions (“unifications”) for those resolutions and keeps track of the required substitutions. The composition of all those substitutions is the computational result of the program: a minimal substitution to reach inconsistency.

To demonstrate how one can compute in Prolog let us consider the addition problem “2 + 2 = ?”. Represent natural numbers by terms in a language with the constant symbol zero and the successor function succ. The ground terms of the language are:

$$\mathrm{zero}, \mathrm{succ}(\mathrm{zero}), \mathrm{succ}(\mathrm{succ}(\mathrm{zero})), \ldots$$

Addition is represented as a ternary predicate

$$\mathrm{add}(X, Y, Z) \leftrightarrow X + Y = Z.$$

The following universal sentences axiomatize addition:

$$\begin{array}{ll} A1. & \forall X.\mathrm{add}(X, \mathrm{zero}, X) \\ A2. & \forall X, Y, Z.(\mathrm{add}(X, Y, Z) \rightarrow \mathrm{add}(X, \mathrm{succ}(Y), \mathrm{succ}(Z))) \end{array}$$

Computing $2 + 2$ can be viewed as an inconsistency problem:

$$4 = \mathrm{succ}(\mathrm{succ}(\mathrm{succ}(\mathrm{succ}(\mathrm{zero}))))$$

is the unique term t of the language such that the axioms A1 and A2 are inconsistent with

$$\neg\mathrm{add}(\mathrm{succ}(\mathrm{succ}(\mathrm{zero})), \mathrm{succ}(\mathrm{succ}(\mathrm{zero})), t).$$

So the aim is to find a possibly iterated substitution for the variable V such that A1 and A2 are inconsitent with

$$\neg\mathrm{add}(\mathrm{succ}(\mathrm{succ}(\mathrm{zero})), \mathrm{succ}(\mathrm{succ}(\mathrm{zero})), V).$$

We can write these formulas in clausal form by omitting quantifyers.

$$
\begin{array}{ll}
A1. & \{\mathrm{add}(X,\mathrm{zero},X)\} \\
A2. & \{\neg\mathrm{add}(X,Y,Z),\mathrm{add}(X,\mathrm{succ}(Y),\mathrm{succ}(Z))\} \\
A3. & \{\neg\mathrm{add}(\mathrm{succ}(\mathrm{succ}(\mathrm{zero})),\mathrm{succ}(\mathrm{succ}(\mathrm{zero})),V)\}
\end{array}
$$

All variables are understood to be universally quantified. So we can rename variables freely, and we shall do so in order to avoid variable clashes.

In `Prolog` notation, the program to compute $2+2$ can be written as follows, where the implication in A2 is indicated by “`:-`”:

```
add(X,zero,X).
add(X,succ(Y),succ(Z) :- add(X,Y,Z).
?- add(succ(succ(zero)),succ(succ(zero)),V).
```

Execution of this program means to find substitutions and resolutions leading to inconsistency: we begin with the clauses

1. $\mathrm{add}(X,\mathrm{zero},X)$
2. $\neg\mathrm{add}(X,Y,Z),\mathrm{add}(X,\mathrm{succ}(Y),\mathrm{succ}(Z))$
3. $\neg\mathrm{add}(\mathrm{succ}(\mathrm{succ}(\mathrm{zero})),\mathrm{succ}(\mathrm{succ}(\mathrm{zero})),V)$

The clauses 2 and 3 can be resolved by making the literals add(X, succ(Y), succ(Z)) and ¬add(succ(succ(zero)), succ(succ(zero)), V) dual using the **substitutions** $X\colon =$ succ(succ(zero)), $Y\colon=$succ(zero), $V\colon=$succ(Z). This yields the resolution:

4. $\neg\mathrm{add}(\mathrm{succ}(\mathrm{succ}(\mathrm{zero})),\mathrm{succ}(\mathrm{zero}),Z)$

This should again resolve against 2. To avoid variable clashes, we first rename the (universal) variables in 2:

5. $\neg\mathrm{add}(X1,Y1,Z1),\mathrm{add}(X1,\mathrm{succ}(Y1),\mathrm{succ}(Z1))$

4 and 5 can be resolved by making the literals add($X1$, succ($Y1$), succ($Z1$)) and ¬ add(succ(succ(zero)), succ(zero), Z) dual using the **substitutions** $X1\colon=$succ(succ(zero)), $Y1\colon=$zero, $Z\colon=$succ($Z1$). This yields the resolution:

6. $\neg\mathrm{add}(\mathrm{succ}(\mathrm{succ}(\mathrm{zero}),\mathrm{zero},Z1)$

This should resolve against 1. To avoid variable clashes, we first rename the (universal) variables in 1 by “new” variables:

7. $\mathrm{add}(X2,\mathrm{zero},X2)$.

6 and 7 can be resolved by the substitutions $X2\colon$ =succ(succ(zero)), $Z1\colon$ =$X2$. This yields the “false” resolution, as required:

8. $\{\}$

The combined substitution for V which lead to this contradiction is obtained by “chasing” through the substitutions:

$$V=\mathrm{succ}(Z)=\mathrm{succ}(\mathrm{succ}(Z1))=\mathrm{succ}(\mathrm{succ}(X2))=\mathrm{succ}(\mathrm{succ}(\mathrm{succ}(\mathrm{succ}(\mathrm{zero})))).$$

Thus 2+2=4!

Exercise 12. Addition and multiplication on the natural numbers can be formalized in Prolog by the following program.

```
add(X,zero,X).
add(X,succ(Y),succ(Z) :- add(X,Y,Z).
mult(X,zero,zero).
mult(X,succ(Y),Z) :- mult(X,Y,W), add(W,X,Z).
```

The question $2\times 2=?$ is expressed by the query:

```
?- mult(succ(succ(zero)),succ(succ(zero)),V).
```

Please describe with pen and paper how `Prolog` calculates this product.

In the Prolog example, clauses with variables were brought into agreement by substitution of variables by terms. Then resolution was applied by cancelling out complementary literals. So far the substitutions used yielded ground instances, i.e., all variables were instantiated by constant terms. On the other hand the resolution method works for arbitrary substitutions. $\{\varphi(x)\}$ and $\{\neg\varphi(y), \psi(y)\}$ can be resolved into $\{\psi(y)\}$ by first transforming $\{\varphi(x)\}$ into $\{\varphi(y)\}$.

Definition 87. *Let* $\mathrm{Var} = \{v_n | n < \omega\}$ *be the set of first-order variables. A* substitution *is a map* $\sigma\colon \mathrm{Var} \to T^S$ *into the set of S-terms. If only a finite part of the substitution σ is relevant, it is usually written in the form* $\frac{\sigma(v_0)\dots\sigma(v_{n-1})}{v_0\dots v_{n-1}}$. *The application of a substitution to a term t or a formula φ is defined as before and written in the form $t\sigma$ or $\varphi\sigma$. Consider a finite set* $c = \{L_0, \dots, L_{l-1}\}$ *of literals. Define the substitution* $c\sigma = \{L_0\,\sigma, \dots, L_{l-1}\,\sigma\}$

a) *A substitution σ is a* unifier *for* $\{L_0, \dots, L_{l-1}\}$ *if* $L_0\,\sigma = \dots = L_{l-1}\,\sigma$.

b) $\{L_0, \dots, L_{l-1}\}$ *is* unifiable *if there is a unifier for* $\{L_0, \dots, L_{l-1}\}$.

c) *A unifier σ for* $\{L_0, \dots, L_{l-1}\}$ *is a* most general unifier *for* $\{L_0, \dots, L_{l-1}\}$ *if every unifier τ factors by σ, i.e., there is another substitution ρ such that $\tau = \rho \circ \sigma$. Here the composition of substitutions is defined by*

$$\rho \circ \sigma(v_n) = \sigma(v_n)\,\rho.$$

Theorem 88. *Let* $\{L_0, \dots, L_l\}$ *be a finite* unifiable *set of literals. Then* $\{L_0, \dots, L_{l-1}\}$ *possesses a most general unifier.*

Proof. Define a sequence $\sigma_0, \dots, \sigma_N$ of substitutions by recursion. Set $\sigma_0 = \mathrm{id} \upharpoonright \mathrm{Var}$.

Assume that σ_i is defined. If $\{L_0\,\sigma_i, \dots, L_l\,\sigma_i\}$ consists of one element then set $N = i$ and stop the recursion.

Now assume that $\{L_0\,\sigma_i, \dots, L_l\,\sigma_i\}$ consists of more then one element. Let p be minimal such that there are substituted literals $L_j\,\sigma_i$ and $L_k\,\sigma_i$ which differ in their p^{th} position (as sequences of symbols). Let $s_j \neq s_k$ be the p^{th} element of $L_j\,\sigma_i$ and $L_k\,\sigma_i$ respectively.
Case 1. $s_j, s_k \notin \mathrm{Var}$. Then set $N = i$ and stop the recursion.
Case 2. $s_j \in \mathrm{Var}$ or $s_k \in \mathrm{Var}$. Without loss of generality we may assume that $s_j \in \mathrm{Var}$, and we set $x = s_j$. Let t be the subterm of $L_k\,\sigma_i$ which starts at the p^{th} position with the symbol s_k.
Case 2.1. $x \in \mathrm{var}(t)$. Then set $N = i$ and stop the recursion.
Case 2.2. $x \notin \mathrm{var}(t)$. Then set

$$\sigma_{i+1} = \frac{t}{x} \circ \sigma_i$$

and continue the recursion.
(1) The recursion stops eventually.
Proof. σ_{i+1} can only be defined via *Case 2.2.* There, the variable x does not occur in t. Applying the substitution $\frac{t}{x}$ to $\{L_0\,\sigma_i, \dots, L_l\,\sigma_i\}$ removes the variable x from

$$\{L_0\,\sigma_{i+1}, \dots, L_l\,\sigma_{i+1}\} = \{L_0\,\sigma_i\frac{t}{s_j}, \dots, L_l\,\sigma_i\frac{t}{s_j}\}.$$

So the number of variables in $\{L_0\,\sigma_i, \dots, L_l\,\sigma_i\}$ goes down by at least 1 in each step of the recursion. Therefore the recursion must stop. *qed*(1)

Now let τ be any unifier for $\{L_0, \dots, L_l\}$: $L_0\,\tau = \dots = L_l\,\tau$.
(2) For $i = 0, \dots, N$ there is a substitution τ_i such that $\tau = \tau_i \circ \sigma_i$.
Proof. Define τ_i by recursion on i. Set $\tau_0 = \tau$. Then $\tau = \tau \circ (\mathrm{id} \upharpoonright \mathrm{Var}) = \tau_0 \circ \sigma_0$.

Assume that τ_i is defined such that $\tau = \tau_i \circ \sigma_i$ and that $i < N$. Then σ_{i+1} is defined according to *Case 2.2*. With the notations of that case: $\sigma_{i+1} = \frac{t}{x} \circ \sigma_i$. Since $\tau = \tau_i \circ \sigma_i$ is a unifier for $\{L_0, ..., L_l\}$ then τ_i is a unifier for $\{L_0 \sigma_i, ..., L_l \sigma_i\}$. Thus the variable x and the term t are unified by τ_i: $x\tau_i = \tau_i(x) = t\tau_i$. Set

$$\tau_{i+1} = (\tau_i \setminus \{(x, \tau_i(x))\}) \cup \{(x, x)\}.$$

We show that $\tau_{i+1} \circ \frac{t}{x} = \tau_i$: if $y \neq x$ then

$$y \frac{t}{x} \tau_{i+1} = y\tau_{i+1} = y\tau_i \,,$$

if $y = x$ then

$$y \frac{t}{x} \tau_{i+1} = t\tau_{i+1} = t\tau_i \text{ (since } x \text{ does not occur in } t) = x\tau_i = y\tau_i \,.$$

Then

$$\begin{aligned} \tau_{i+1} \circ \sigma_{i+1} &= \tau_{i+1} \circ (\frac{t}{x} \circ \sigma_i) \\ &= (\tau_{i+1} \circ \frac{t}{x}) \circ \sigma_i \\ &= \tau_i \circ \sigma_i \\ &= \tau \,. \end{aligned}$$

□

We can now define first-order resolution.

Definition 89. *Let c' and c'' be clauses. Let the substitutions σ':* Var ↔ Var *and σ'':* Var ↔ Var *be* renamings of variables *so that $c'\sigma'$ and $c''\sigma''$ do not have common variables. Let $\{L_1, ..., L_m\} \subseteq c'\sigma'$ and $\{K_1, ..., K_n\} \subseteq c''\sigma''$ be sets of literals such that*

$$\{L_1, ..., L_m, \bar{K}_1, ..., \bar{K}_n\}$$

is unifiable where $m, n \geqslant 1$. Let σ be a most general unifier of $\{L_1, ..., L_m, \bar{K}_1, ..., \bar{K}_n\}$. Then the clause

$$c = [(c'\sigma' \setminus \{L_1, ..., L_m\}) \cup (c''\sigma'' \setminus \{K_1, ..., K_n\})]\, \sigma$$

is a (first-order) resolution *of c' and c''.*

Given the clauses c' and c'' one just has to find parts which are unifiable and compute c. It is not necessary to "find" ground instances of the clauses. On the other hand, resolution with ground instances can be gotten from first-order resolution by lifting-techniques.

Theorem 90. *Let c' and c'' be clauses and let c_0' and c_0'' be ground instances of c' and c'' which are resolvable. Let c_0 be a resolution of c_0' and c_0''. Then there is a first-order resolution c of c' and c'' such that c_0 is a ground instance of c.*

Proof. First let σ': Var ↔ Var and σ'': Var ↔ Var be *renamings of variables* so that $c'\sigma'$ and $c''\sigma''$ do not have common variables. Since c_0' and c_0'' are ground instances of c' and c'' they are also ground instances of $c'\sigma'$ and $c''\sigma''$. Let

$$c_0' = c'\sigma'\tau' \text{ and } c_0'' = c''\sigma''\tau''.$$

Since $c'\sigma'$ and $c''\sigma''$ do not have common variables we can assume that τ' and τ'' substitute disjoint sets of variables. Letting $\tau = \tau' \circ \tau''$ we get

$$c_0' = c'\sigma'\tau \text{ and } c_0'' = c''\sigma''\tau.$$

Let the resolution c_0 of c_0' and c_0'' be based on the literal L: $L \in c_0'$ and $\bar{L} \in c_0''$ and

$$c_0 = (c_0' \setminus \{L\}) \cup (c_0'' \setminus \{\bar{L}\}).$$

The literal L is a ground instance of possibly several literals $L_1, ..., L_m \in c'\sigma'$ by the ground substitution τ. Similarly the literal $\bar{L}$ is a ground instance of possibly several literals $K_1, ...,$ $K_n \in c''\sigma''$ by the ground substitution τ. Now τ unifies $\{L_1, ..., L_m, \bar{K}_1, ..., \bar{K}_n\}$ into L. By the theorem on the existence of most general unifiers let σ be a most general unifier for

$$\{L_1, ..., L_m, \bar{K}_1, ..., \bar{K}_n\}.$$

Then

$$c = [(c'\sigma' \setminus \{L_1, ..., L_m\}) \cup (c''\sigma'' \setminus \{K_1, ..., K_n\})]\,\sigma$$

is a *(first-order) resolution* of c' and c''. Since σ is most general, take another substitution ρ such that $\tau = \rho \circ \sigma$. Then

$$\begin{aligned} c_0 &= (c_0' \setminus \{L\}) \cup (c_0'' \setminus \{\bar{L}\}) \\ &= (c'\sigma'\tau \setminus \{L\}) \cup (c''\sigma''\tau \setminus \{\bar{L}\}) \\ &= [(c'\sigma' \setminus \{L_1, ..., L_m\}) \cup (c''\sigma'' \setminus \{K_1, ..., K_n\})]\tau \\ &= [(c'\sigma' \setminus \{L_1, ..., L_m\}) \cup (c''\sigma'' \setminus \{K_1, ..., K_n\})]\sigma\rho \\ &= c\rho \end{aligned}$$

is a ground instance of c. □

Theorem 91. *Let C be a set of clauses and let $c_0 = c\sigma_0$ be a ground instance of c. Then $C \vdash c_0$ by resolution with ground clauses iff there is are substitutions σ and τ such that $C \vdash c\sigma$ can be shown by first-order resolution and $c_0 = c\sigma\tau$.*

21 Set theory

Almost all mathematical notions can be defined set-theoretically. GEORG CANTOR, the founder of set theory, gave the following definition or description:

> Unter einer Menge verstehen wir jede Zusammenfassung M von bestimmten, wohlunterschiedenen Objekten m unsrer Anschauung oder unseres Denkens (welche die "Elemente" von M genannt werden) zu einem Ganzen.

Felix Hausdorff begins the *Grundzüge der Mengenlehre* with a concise description, which seems less dependent on human minds:

> Eine Menge ist eine Zusammenfassung von Dingen zu einem Ganzen, d.h. zu einem neuen Ding.

The notion of set is adequately formalized in first-order axiom systems introduced by ZERMELO, FRAENKEL and others. Together with the GÖDEL completeness theorem for first-order logic this constitutes a "formalistic" answer to the question "what is mathematics": mathematics consists of formal proofs from the axioms of ZERMELO-FRAENKEL set theory.

Definition 92. *Let $\in$ be a binary infix relation symbol; read $x \in y$ as "x is an* element *of y". The* language of set theory *is the language $\{\in\}$. The formulas in $L^{\{\in\}}$ are called* set theoretical formulas *or $\in$-formulas. We write $L^{\in}$ instead of $L^{\{\in\}}$.*

The naive notion of *set* is intuitively understood and was used extensively in previous chapters. The following axioms describe properties of naive sets. Note that the axiom system is an infinite *set* of axioms. It seems unavoidable that we have to go back to some previously given set notions to be able to define the collection of set theoretical axioms - another example of the frequent circularity in foundational theories.

Definition 93. *The axiom system* ST *of* set theory *consists of the following axioms:*

a) The axiom of extensionality (Ext)*:*

$$\forall x \forall y (\forall z (z \in x \leftrightarrow z \in y) \rightarrow x \equiv y)$$

- a set is determined by its elements, sets having the same elements are identical.

b) The pairing axiom (Pair)*:*

$$\forall x \forall y \exists z \forall w\, (w \in z \leftrightarrow w \equiv x \vee w \equiv y).$$

- z is the unordered pair of x and y.

c) The union axiom (Union)*:*

$$\forall x \exists y \forall z (z \in y \leftrightarrow \exists w (w \in x \wedge z \in w))$$

- y is the union of all elements of x.

d) The powerset axiom (Pow)*:*

$$\forall x \exists y \forall z (z \in y \leftrightarrow \forall w (w \in z \rightarrow w \in x))$$

- y consists of all subsets of x.

e) The separation schema (Sep) *postulates for every $\in$-formula $\varphi(z, x_1, ..., x_n)$:*

$$\forall x_1 ... \forall x_n \forall x \exists y \forall z\, (z \in y \leftrightarrow z \in x \wedge \varphi(z, x_1, ..., x_n))$$

- this is an infinite scheme of axioms, the set z consists of all elements of x which satisfy φ.

f) The replacement schema (Rep) *postulates for every $\in$-formula $\varphi(x, y, x_1, ..., x_n)$:*

$$\forall x_1 ... \forall x_n (\forall x \forall y \forall y' ((\varphi(x, y, x_1, ..., x_n) \wedge \varphi(x, y', x_1, ..., x_n)) \rightarrow y \equiv y') \rightarrow \\ \forall u \exists v \forall y\, (y \in v \leftrightarrow \exists x (x \in u \wedge \varphi(x, y, x_1, ..., x_n))))$$

- v is the image of u under the map defined by φ.

g) The foundation schema (Found) *postulates for every $\in$-formula $\varphi(x, x_1, ..., x_n)$:*

$$\forall x_1 ... \forall x_n (\exists x \varphi(x, x_1, ..., x_n) \rightarrow \exists x (\varphi(x, x_1, ..., x_n) \wedge \forall x' (x' \in x \rightarrow \neg \varphi(x', x_1, ..., x_n))))$$

- if φ is satisfiable then there are $\in$-minimal elements satisfying φ.

The axiom of *extensionality* expresses that a set is only determined by its elements. There is no further structure in a set; the order or multiplicity of elements does not matter. The axiom of extensionality can also be seen as a definition of $\equiv$ in terms of $\in$:

$$\forall x \forall y (x \equiv y \leftrightarrow \forall z (z \in x \leftrightarrow z \in y)).$$

The axioms $b) - d)$ describe the basic set theoretic operations of forming two-element sets, unions, and power sets. The separation schema ("Aussonderung") is the crucial axiom of Zermelo set theory. Gottlob Frege had used the more liberal comprehension schema

$$\forall x_1 ... \forall x_n \exists y \forall z\, (z \in y \leftrightarrow \varphi(z, x_1, ..., x_n))$$

without restricting the variable z to some x on the right hand side. This however lead to the famous RUSSELL paradox and was thus inconsistent. ZERMELO's restriction apparently avoids contradiction.

The replacement schema was added by ABRAHAM FRAENKEL to postulate that functional images of sets are sets.

The foundation schema by MIRIMANOFF allows to carry out induction on the binary relation $\in$. To prove a universal property by contradiction one can look at a minimal counterexample and argue that the property is inherited from the elements of a set to the set. The schema is used seldomly in mathematical practice, but it is very convenient for the development of set theory.

Note that the axioms of ST do not require the existence of infinite sets, and indeed one can easily build a canonical model of ST consisting only of finite sets. Such a model can be defined over the structure $\mathbb{N} = (\mathbb{N}, +, \cdot, 0, 1)$. The theory ST has the same strength as first-order Peano arithmetic (PA).

The theory would become much stronger, if the *axiom of infinity (Inf)* was added:

$$\exists x (\exists y\, (y \in x \wedge \forall z \neg z \in y) \wedge \forall y (y \in x \rightarrow \exists z (z \in x \wedge \forall w (w \in z \leftrightarrow w \in y \vee w \equiv y)))).$$

Intuitively, the closure properties of x ensure that x is infinite. The strengthened theory is ZERMELO-FRAENKEL set theory (without the axiom of choice), which is usually taken as the universal foundation of mathematics. We work with the weaker theory ST, since we want to show the GÖDEL incompleteness theorems for ST, which are another version of the usual incompleteness theorems for PA.

21.1 Class terms

Most of the axioms have a form like

$$\forall \vec{x} \exists y \forall z\, (z \in y \leftrightarrow \varphi).$$

Intuitively, y is the collection of sets z which satisfy φ. The common notation for that set is

$$\{z | \varphi\}.$$

This is to be seen as a term, which assigns to the other parameters in φ the value $\{z|\varphi\}$. Since the result of such a term is not necessarily a set we call such terms *class terms*. It is very convenient to employ class terms *within* $\in$-formulas. We view this notation as an abbreviation for "pure" $\in$-formulas.

Definition 94. *A* class term *is of the form $\{x|\varphi\}$ where x is a variable and $\varphi \in L^{\in}$. If $\{x|\varphi\}$ and $\{y|\psi\}$ are class terms then*

- $u \in \{x|\varphi\}$ *stands for* $\varphi\frac{u}{x}$ *;*
- $u = \{x|\varphi\}$ *stands for* $\forall v\, (v \in u \leftrightarrow \varphi\frac{v}{x})$*;*
- $\{x|\varphi\} = u$ *stands for* $\forall v\, (\varphi\frac{v}{x} \leftrightarrow v \in u)$*;*
- $\{x|\varphi\} = \{y|\psi\}$ *stands for* $\forall v\, (\varphi\frac{v}{x} \leftrightarrow \psi\frac{v}{y})$*;*
- $\{x|\varphi\} \in u$ *stands for* $\exists v (v \in u \wedge v = \{x|\varphi\}$*;*
- $\{x|\varphi\} \in \{y|\psi\}$ *stands for* $\exists v (\psi\frac{v}{y} \wedge v = \{x|\varphi\}$.

In this notation, the separation schema becomes:

$$\forall x_1 ... \forall x_n \forall x \exists y\, y = \{z | z \in x \wedge \varphi(z, x_1, ..., x_n)\}.$$

We shall further extend this notation, first by giving specific names to important formulas and class terms.

Definition 95.

a) $\emptyset := \{x|x \neq x\}$ *is the* empty set*;*

b) $V := \{x|x = x\}$ *is the* universe.

We work in the theory ZF for the following propositions.

Proposition 96.

a) $\emptyset \in V.$

b) $V \notin V$ *(*Russell*'s antinomy).*

Proof. a) $\emptyset \in V$ abbreviates the formula

$$\exists v(v = v \wedge v = \emptyset).$$

This is equivalent to $\exists v\, v = \emptyset$ which again is an abbreviation for

$$\exists v \forall w\, (w \in v \leftrightarrow w \neq w).$$

Consider an arbitrary set x. Then the formula is equivalent to

$$\exists v \forall w\, (w \in v \leftrightarrow w \in x \wedge w \neq w).$$

This follows from the instance

$$\forall x \exists y \forall z\, (z \in y \leftrightarrow z \in x \wedge z \neq z)$$

of the separation schema for the formula $z \neq z$.
b) Assume that $V \in V$. By the schema of separation

$$\exists y\, y = \{z|z \in V \wedge z \notin z\}.$$

Let $y = \{z|z \in V \wedge z \notin z\}$. Then

$$\forall z\, (z \in y \leftrightarrow z \in V \wedge z \notin z).$$

This is equivalent to

$$\forall z\, (z \in y \leftrightarrow z \notin z).$$

Instantiating the universal quantifier with y yields

$$y \in y \leftrightarrow y \notin y$$

which is a contradiction. □

We introduce further abbreviations. By a *term* we understand a class term or a variable, i.e., those terms which may occur in an extended $\in$-formula. We also introduce *bounded quantifiers* to simplify notation.

Definition 97. *Let A be a term. Then $\forall x \in A\, \varphi$ stands for $\forall x(x \in A \rightarrow \varphi)$ and $\exists x \in A\, \varphi$ stands for $\exists x\, (x \in A \wedge \varphi)$.*

Definition 98. *Let $x, y, z, \ldots$ be variables and $X, Y, Z, \ldots$ be class terms. Define*

a) $X \subseteq Y := \forall x \in X\, x \in Y$*, X is a* subclass *of Y;*

b) $X \cup Y := \{x|x \in X \vee x \in Y\}$ *is the* union *of X and Y;*

c) $X\cap Y := \{x|x\in X\wedge x\in Y\}$ *is the* intersection *of* X *and* Y*;*

d) $X\setminus Y := \{x|x\in X\wedge x\notin Y\}$ *is the* difference *of* X *and* Y*;*

e) $\bigcup X := \{x|\exists y\in X\; x\in y\}$ *is the* union *of* X*;*

f) $\bigcap X := \{x|\forall y\in X\; x\in y\}$ *is the* intersection *of* X*;*

g) $\mathcal{P}(X) = \{x|x\subseteq X\}$ *is the* power class *of* X*;*

h) $\{X\} = \{x|x = X\}$ *is the* singleton set *of* X*;*

i) $\{X, Y\} = \{x|x = X\vee x = Y\}$ *is the* (unordered) pair *of* X *and* Y*;*

j) $\{X_0, ..., X_{n-1}\} = \{x|x = X_0\vee ...\vee x = X_{n-1}\}$.

One can prove the well-known boolean properties for these operations. We only give a few examples.

Proposition 99. $X\subseteq Y\wedge Y\subseteq X\rightarrow X = Y.$

Proposition 100. $\bigcup\{x, y\} = x\cup y.$

Proof. We show the equality by two inclusions:
($\subseteq$). Let $u\in\bigcup\{x, y\}$. $\exists v(v\in\{x, y\}\wedge u\in v)$. Let $v\in\{x, y\}\wedge u\in v$. $(v = x\vee v = y)\wedge u\in v$.
Case 1. $v = x$. Then $u\in x$. $u\in x\vee u\in y$. Hence $u\in x\cup y$.
Case 2. $v = y$. Then $u\in y$. $u\in x\vee u\in y$. Hence $u\in x\cup y$.

Conversely let $u\in x\cup y$. $u\in x\vee u\in y$.
Case 1. $u\in x$. Then $x\in\{x, y\}\wedge u\in x$. $\exists v(v\in\{x, y\}\wedge u\in v)$ and $u\in\bigcup\{x, y\}$.
Case 2. $u\in y$. Then $x\in\{x, y\}\wedge u\in x$. $\exists v(v\in\{x, y\}\wedge u\in v)$ and $u\in\bigcup\{x, y\}$. □

We can now reformulate the ZF axioms using class terms; for brevity we omit initial universal quantifiers.

a) Extensionality: $x\subseteq y\wedge y\subseteq x\rightarrow x = y$.

b) Pairing: $\{x, y\}\in V$.

c) Union: $\bigcup x\in V$.

d) Powerset: $\mathcal{P}(x)\in V$.

e) Separation schema: for all terms A with free variables $x_0, ..., x_{n-1}$

$$x\cap A\in V.$$

f) Replacement: see later.

g) Foundation: for all terms A with free variables $x_0, ..., x_{n-1}$

$$A\neq\emptyset\rightarrow\exists x\in A\; x\cap A = \emptyset.$$

Also the axiom of infinity can be written as

$$\exists x\;(\emptyset\in x\wedge\forall u\in x\; u\cup\{u\}\in x).$$

22 Relations and functions

Ordered pairs are the basis for the theory of relations.

Definition 101. $(x,y)=\{\{x\},\{x,y\}\}$ *is the* ordered pair *of* x *and* y*.*

Remark 102. There are sometimes discussions whether $(x,\, y)$ *is* the ordered pair of x and y , or to what degree it agrees with the intuitive notion of an ordered pair. ...

Proposition 103. $(x,y)\in V$*.*
$(x,y)=(x',y')\rightarrow x=y\wedge x'=y'$*.*

Definition 104. *Let* A,B,R *be terms. Define*

a) $A\times B=\{z|\exists a\in A\exists b\in B\, z=(a,b)\}$ *is the* cartesian product *of* A *and* B*.*

b) R *is a* (binary) relation *if* $R\subseteq V\times V$*.*

c) If R *is a binary relation write* aRb *instead of* $(a,b)\in R$*.*

We can now introduce the usual notions for relations:

Definition 105.

a) $\operatorname{dom}(R)=\{x|\exists y\,(x,y)\in R\}$ *is the* domain *of* R*.*

b) $\operatorname{ran}(R)=\{y|\exists x\,(x,y)\in R\}$ *is the* range *of* R*.*

c) $R\upharpoonright A=\{z|z\in R\wedge\exists x\exists y((x,y)=z\wedge x\in A)\}$ *is the* restriction *of* R *to* A*.*

d) $R[A]=\{y|\exists x\in A\, xRy\}$ *is the* image *of* A *under* R*.*

e) $R^{-1}=\{z|\exists x\exists y\,(xRy\wedge z=(y,x))\}$ *is the* inverse *of* R*.*

f) $R^{-1}[B]=\{x|\exists y\in B\, xRy\}$ *is the* preimage *of* B *under* R*.*

One can prove the usual properties for these notions in ZF. One can now formalize the types of relations, like equivalence relations, partial and linear orders, etc. We shall only consider notions which are relevant for our short introduction to set theory.

Definition 106. *Let* F,A,B *be terms. Then*

a) F *is a* function *if* $\forall x\forall y,y'\,(xFy\wedge xFy'\rightarrow y=y')$*.*

b) $F\colon A\rightarrow B$ *if* F *is a function*$\wedge\operatorname{dom}(F)=A\wedge\operatorname{ran}(F)\subseteq B$*. The sequence notions* $(F(x)|x\in A)$ *or* $(F(x))_{x\in A}$ *are just other ways to write* $F\colon A\rightarrow V$*.*

c) $F(x)=\{v|\exists y\,(xFy\wedge\forall y'\,(xFy'\rightarrow y=y')\rightarrow\exists y\,(xFy\wedge v\in y)\}$ *is the* value *of* F *at* x*.*

Note that if $F\colon A\rightarrow B$ and $x\in A$ then $xFF(x)$. If there is no unique y such that xFy then $F(x)=V$ which we may read as $F(x)$ is “undefined”.

Using functional notations we may now write the replacement schema as

$$F \text{ is a function} \rightarrow F[x]\in V.$$

One could now develop the usual theory of functions and formalize notions like surjective, injective, bijective, compositions, etc.

23 Natural numbers and complete induction

It is natural to formalize the integer n in set theory by some set with n elements. This intuitive plan will be implemented in the sequel. For every ordinary natural number we define a set-theoretical version.

Definition 107.

$$\begin{array}{lcl} 0 & = & \emptyset \\ 1 & = & \{0\} \\ 2 & = & \{0,1\} \\ & \vdots & \\ n+1 & = & \{0,1,\ldots,n\}=\{0,1,\ldots,n-1\}\cup\{n\}=n\cup\{n\} \\ & \vdots & \end{array}$$

For $x\in V$ define $x+1=x\cup\{x\}$ and $x+2=(x+1)+1$.

We would like to set $\mathbb{N}=\{0,1,2,\ldots\}$, but such infinitary definitions are not possible in first-order logic. Instead we look for a characteristic property to define a class which is able to serve as the collection of set-theoretic integers. Every ordinary integer n is either of the form $n=0$ or $n=m+1$ for some other integer.

Definition 108. *A set n is a* natural number *if*

$$\forall i\in n+1(i=0\vee\exists j\in n.i=j+1).$$

Let $\mathbb{N}$ be the class of natural numbers. We shall often use letters like i, j, k, l, m, n as variables for natural numbers.

We show that this class is an adequate formalization of the intuitive notion of "natural number".

Lemma 109.

a) $0\in\mathbb{N}$*;*

b) $\forall n\in\mathbb{N}.\,n+1\in\mathbb{N}$*;*

c) $0,1,2,\ldots\in\mathbb{N}$*.*

Proof. *a)* is trivial. *b)* Assume $n\in\mathbb{N}$.To show that $n+1\in\mathbb{N}$ consider $i\in(n+1)+1=(n+1)\cup\{n+1\}$. If $i\in n+1$ then we have by assumption that $i=0\vee\exists j\in n.i=j+1$ and so $i=0\vee\exists j\in n+1.i=j+1$. If $i=n+1$ then $n\in n+1$ satisfies $\exists j\in n+1.i=j+1$. □

$\mathbb{N}$ is the $\subseteq$-smallest class which contains 0 and is closed under $+1$.

Lemma 110. *Let A be a term such that $0\in A$ and $\forall x\in A.x+1\in A$. Then $\mathbb{N}\subseteq A$.*

Proof. Let $n\in\mathbb{N}$. Assume for a contradiction that $n\notin A$. By foundation let $i\in n+1$ be $\in$-minimal such that $i\notin A$. Then $i\neq 0$. By the definition of $\mathbb{N}$ there is $j\in n$ such that $i=j+1$. Then $j\in i$ and by the $\in$-minimality of i we have $j\in A$. But then $i=j+1\in A$. Contradiction. □

This immediately implies the principle of (complete) induction theorem for natural numbers.

Theorem 111. *Let $\varphi(n,\vec{p})$ be an $\in$-formula. Assume that $\varphi(0,\vec{p})$ and that $\forall n\in\mathbb{N}.\varphi(n,\vec{p})\rightarrow\varphi(n+1,\vec{p})$. Then $\forall n\in\mathbb{N}.\varphi(n,\vec{p})$.*

Proof. Apply the previous lemma with $A=\{x\in\mathbb{N}\mid\varphi(x,\vec{p})\}$. □

Theorem 112. $\mathbb{N}$ *is a set if the axiom of infinity holds.*

Proof. If the axiom of infinity holds, take a set x such that

$$(\exists y\,(y\in x\wedge\forall z\neg z\in y)\wedge\forall y(y\in x\rightarrow\exists z(z\in x\wedge\forall w(w\in z\leftrightarrow w\in y\vee w\equiv y)))).$$

This means that $\emptyset\in x\wedge\forall y\in x.y+1\in x$. By Lemma 110, $\mathbb{N}\subseteq x$. Then $\mathbb{N}=x\cap\mathbb{N}$ is a set by separation.

Conversely assume that $\mathbb{N}\in V$. Then $\mathbb{N}$ obviously witnesses the axiom of infinity. □

So the axiom of infinity can be written briefly as

$$\mathbb{N}\in V.$$

The natural numbers are linearly ordered by $\in$. To prove this we draw a useful consequence of foundation, which also shows that the $\in$-relation is strict on $\mathbb{N}$.

Lemma 113. *There is* no *finite sequence $x_0,x_1,\ldots,x_n$ which forms an $\in$-cycle with*

$$x_0\in x_1\in\ldots\in x_n\in x_0.$$

In particular $\forall x\,x\notin x$.

Proof. Assume that $x_0\in x_1\in\ldots\in x_n\in x_0$. Let $A=\{x_0,\ldots,x_n\}$. $A\neq\emptyset$ since $x_0\in A$. By foundation, take $x\in A$ such that $x\cap A=\emptyset$.
Case 1. $x=x_0$. Then $x_n\in x\cap A\neq\emptyset$, contradiction.
Case 2. $x=x_i$ for some $1\leqslant i\leqslant n$. Then $x_{i-1}\in x\cap A=\emptyset$, contradiction. □

Theorem 114. *$\forall m,n\in\mathbb{N}.m\in n\vee m=n\vee n\in m$.*

Proof. By complete induction on $m\in\mathbb{N}$. The property holds for $m=0$
(1) $\forall n\in\mathbb{N}.0\in n\vee 0=n\vee n\in 0$.
Proof. By complete induction on $n\in\mathbb{N}$. The initial case $n=0$ is trivial:
(1.1) $0\in 0\vee 0=0\vee 0\in 0$.
(1.2) Assume that $0\in n\vee 0=n\vee n\in 0$. Then $0\in n+1\vee 0=n+1\vee n+1\in 0$.
Proof. We have $0\in n\vee 0=n$. Then $0\in n+1$. *qed*(1.2, 1)
(2) Assume $\forall n\in\mathbb{N}.m\in n\vee m=n\vee n\in m$. Then $\forall n\in\mathbb{N}.m+1\in n\vee m+1=n\vee n\in m+1$.
Proof. We prove the conclusion by induction on $n\in\mathbb{N}$.
(2.1) $m+1\in 0\vee m+1=0\vee 0\in m+1$.
Proof. By assumption $m\in 0\vee m=0\vee 0\in m$. Then $0\in m\vee 0=m$. Thus $0\in m+1$. *qed*(2.1)
(2.2) Assume that $m+1\in n\vee m+1=n\vee n\in m+1$. Then $m+1\in n+1\vee m+1=n+1\vee n+1\in m+1$.
Proof. If $m+1\in n\vee m+1=n$ then $m+1\in n+1$ as required. So assume that $n\in m+1$. If $n=m$ then $m+1=n+1$ as required.

This leaves the case $n\in m$. By the assumption of (2) we have $m\in n+1\vee m=n+1\vee n+1\in m$. We obtain a contradiction from the case $m\in n+1$: if $m\in n$ then $m\in n\in m$, contracting the previous lemma; if $m=n$ then $m\in m$ which again contradicts the lemma. So we are left with $m=n+1\vee n+1\in m$ which implies $n+1\in m+1$. *qed*(2.2, 2) □

Definition 115. *Let $<=\{(m,n)|\,m\in n\}$ be the natural strict linear order on $\mathbb{N}$.*

A natural number is the set of smaller numbers.

Lemma 116. *For $n\in\mathbb{N}$, $n\subseteq\mathbb{N}$. Therefore $n=\{m\in\mathbb{N}\,|\,m<n\}$.*

Proof. By induction on $n\in\mathbb{N}$. Obviously $\emptyset\subseteq\mathbb{N}$. Assume that $n\subseteq\mathbb{N}$. Then $n+1=n\cup\{n\}\subseteq\mathbb{N}$. □

So, intuitively, we have the desired $n = \{0, 1, ..., n-1\}$.

24 Complete recursion and arithmetic

Recursion, often called induction as well, over the natural numbers is a ubiquitous method for defining mathematical objects.

Theorem 117. *Let $a \in V$ and $G: V \to V$. Then there is a canonically defined class term F such that*

$$F: \mathbb{N} \to V,\ F(0) = a \text{ and } \forall n \in \mathbb{N}\ F(n+1) = G(F(n)).$$

We then say that F is defined by recursion *over $\mathbb{N}$ by the recursion equations $F(0) = a$ and $\forall n \in \mathbb{N}\ F(n+1) = G(F(n))$. Moreover, the function F is uniquely determined: if $F': \mathbb{N} \to V$ satisfies the recursion equations $F'(0) = a$ and $\forall n \in \mathbb{N}\ F'(n+1) = G(F'(n))$ then $F = F'$.*

Proof. To "compute" the value of F at some $k \in \mathbb{N}$ we would intuitively form a sequence of $(F(0), F(1), ..., F(k))$ according to the recursion equations. This finite sequence is uniquely determined by k.
(1) For all $k \in \mathbb{N}$ there exists a unique $f: k+1 \to V$ such that $f(0) = a$ and $\forall n < k. f(n+1) = G(f(n))$.
Proof. By induction on k. For $k = 0$ set $f = \{(0, a)\}: 1 \to V$. Obviously, this f is uniquely determined.

Assume that $k \in \mathbb{N}$ and that $f: k+1 \to V$ is the uniquely determined function with $f(0) = a$ and $\forall n < k. f(n+1) = G(f(n))$. Define

$$f' = f \cup \{(k+1, G(f(k)))\}: k+2 \to V.$$

Then $f' \in V$ is a set by pairing and union. Also $f'(0) = a$ and $\forall n < k+1. f'(n+1) = G(f'(n))$ by the assumptions on f and by the definition of f'.

Consider some $f'': k+2 \to V$ such that $f''(0) = a$ and $\forall n < k+1. f''(n+1) = G(f''(n))$. By the uniqueness assumption at k we must have $f'' \upharpoonright k+1 = f = f' \upharpoonright k+1$. Also

$$f''(k+1) = G(f''(k)) = G(f(k)) = G(f'(k)) = f'(k+1).$$

Thus $f'' = f'$ which proves uniqueness at $k+1$. *qed*(1)

We can now define

$$F = \{(k, f(k)) | k \in \mathbb{N} \wedge f: k+1 \to V \wedge f(0) = a \wedge \forall n < k. f(n+1) = G(f(n))\}.$$

By (1), $F: \mathbb{N} \to V$, and we have to check the recursive equations. $F(0) = f(0) = a$ for some appropriate f. For $n \in \mathbb{N}$ choose $f: n+2 \to V$ with $f(0) = a \wedge \forall k < n. f(k+1) = G(f(k))$. As before, $f \upharpoonright n+1$ is the unique function used to define $F(n)$. Then

$$F(n+1) = f(n+1) = G(f(n)) = G((f \upharpoonright n+1)(n)) = G(F(n)).$$

Finally, to show the uniqueness of F let $F': \mathbb{N} \to V$ satisfy the recursion equations $F'(0) = a$ and $\forall n \in \mathbb{N}\ F'(n+1) = G(F'(n))$.
(2) $\forall n \in \mathbb{N}. F(n) = F'(n)$.
Proof. By induction. $F(0) = a = F'(0)$. For the induction step assume that $F(n) = F'(n)$. Then $F(n+1) = G(F(n)) = G(F'(n)) = F'(n+1)$. □

We can now define arithmetical operations on natural numbers, using familiar recursive properties.

Definition 118. *Define the* sum $\mathrm{add}(m,n)$ *of* $m,n\in\mathbb{N}$ *by recursion on the variable* n *(taking* m *as a parameter) such that*

$$\mathrm{add}(m,n)=\begin{cases} m, \textit{ if } n=0\\ \mathrm{add}(m,k)+1, \textit{ if } n=k+1\end{cases}$$

Also write $m+n$ *instead of* $\mathrm{add}(m,n)$*. Then the recursive equation can be written as*

$$\begin{array}{rcl} m+0 &=& m\\ m+(n+1) &=& (m+n)+1. \end{array}$$

One can show that addition satisfies the expected properties.

Proposition 119.

a) $m+n\in\mathbb{N}$ *.*

b) $m+0=0+m=m$ *.*

c) $(i+j)+k=i+(j+k)$*.*

d) $m+n=n+m$ *.*

Proof. By induction. □

Definition 120. *Define the* product $m\cdot n$ *of* $m,n\in\mathbb{N}$*:*

$$\begin{array}{rcl} m\cdot 0 &=& 0\\ m\cdot(n+1) &=& (m\cdot n)+m \end{array}$$

Multiplication satisfies natural properties.

Proposition 121.

a) $m\cdot n\in\mathbb{N}$ *.*

b) $m\cdot 0=0\cdot m=0$ *.*

c) $(k\cdot m)\cdot l=k\cdot(m\cdot l)$*.*

d) $k\cdot(l+m)=(k\cdot l)+(k\cdot m)$*.*

e) $m\cdot n=n{\cdot}m$ *.*

Proof. By induction. □

Definition 122. *Define the* power m^n *of* $m,n\in\mathbb{N}$ *recursively:*

$$\begin{array}{rcl} m^0 &=& 1\\ m^{n+1} &=& (m^n)\cdot m \end{array}$$

Again one can prove the usual arithmetic laws for exponentiation. We shall also need that obvious instances of these laws are provable in ST. E.g.,

a) $\mathrm{ST}\vdash 0<1$, $\mathrm{ST}\vdash 0<2$, $\mathrm{ST}\vdash 1<2$, etc.. To interpret these statements observe that the formulas on the right-hand side of $\vdash$ have to be $\in$-formulas. Hence the "numbers" occuring in those formulas must be abstraction terms, i.e., they are the set-theoretic numbers defined by $0=\emptyset$ and $n+1=n\cup\{n\}$, and the $<$-symbol is the set-theoretic $\in$. To prove that, e.g., $\mathrm{ST}\vdash 3<5$ requires in principle to unravel the abstraction term notation into $\in$-notation. Although these proof are schematic, $\mathrm{ST}\vdash 3<5$ will be proved by a different and much shorter proof than $\mathrm{ST}\vdash 121<122$.

b) In the metatheory, where we are constructing statements and proofs, these consideration can be expressed by:

   if m and n are ordinary natural numbers with $m < n$ then $\mathrm{ST} \vdash m < n$.

   This metatheoretic statement deserves some comment: it is a universal statement in the metatheory, quantifying over all ordinary natural numbers, i.e., the intuitive mathematical numbers. In the metatheory we can form the set-theoretic statement $m < n$ for all ordinary natural numbers m and n. The statement $\mathrm{ST} \vdash m < n$ contains a metatheoretic existential quantification expressing that a formal proof exists.

25 Formalizing the metatheory in the object theory

We are currently "reflecting" parts of our metatheory, i.e., common mathematical argumentation, into our object theory ST. So far this has been done for class operations, ordered pairs, relations, functions, and natural numbers. We have given ST-definitions of certain notions and we have proved propositions showing that the ST-definitions satisfy ST-properties corresponding to properties of the original metatheoretic notions. If the ST-definitions are chosen efficiently then the proofs of the propositions should be straightforward, corresponding to the usual intuitions and arguments.

We shall continue to formalize the syntax of first-order logic within ST. To avoid confusion, it is helpful to distinguish between metatheoretic notions and their ST-formalizations. We introduce "Gödel brackets" to denote formalizations.

> If A is a metatheoretic notion that has been formalized in ST we may use $\ulcorner A \urcorner$ to denote its formalization.

Sometimes $\ulcorner A \urcorner$ is called the *Gödelization* of A . If we would formalize into number theory instead of set theory, and if then $\ulcorner A \urcorner$ is a natural number, one often calls $\ulcorner A \urcorner$ the *Gödel number* of A. Analogously we may call $\ulcorner A \urcorner$ the *Gödel set* of A. This turns properties of mathematical objects into properties of sets.

So in defining the natural numbers within ST we could have been more careful by defining

$$\begin{array}{lcl} \ulcorner 0 \urcorner & = & \emptyset \\ \ulcorner 1 \urcorner & = & \{0\} \\ \ulcorner 2 \urcorner & = & \{0,1\} \\ & \vdots & \\ \ulcorner n+1 \urcorner & = & \{0,1,...,n\} = \{0,1,...,n-1\} \cup \{n\} = n \cup \{n\} \\ & \vdots & \end{array}$$

Then the "correctness" statements from the last chapter look like:

a) $\mathrm{ST} \vdash \ulcorner 0 \urcorner < \ulcorner 1 \urcorner$, $\mathrm{ST} \vdash \ulcorner 0 \urcorner < \ulcorner 2 \urcorner$, $\mathrm{ST} \vdash \ulcorner 1 \urcorner < \ulcorner 2 \urcorner$, etc.. That $<$ is the set-theoretic $<$-relation, namely $\in$, can still be infered from the context. Of course one could also introduce further notation to distinguish the set-theoretic relation from the metatheoretic $<$-relation. We shall, however, try to restrict our notation to the necessary.

b) if m and n are ordinary natural numbers with $m < n$ then $\mathrm{ST} \vdash \ulcorner m \urcorner < \ulcorner n \urcorner$.

There are many more correctness statements in the context of our formalizations. If the formalization was to be carried out in some automatic system, one should introduce mechanisms to prove most of these schematically and automatically. Here some examples:

a)

if $k \cdot l = m$ then $\mathrm{ST} \vdash \ulcorner k \urcorner \cdot \ulcorner l \urcorner = \ulcorner m \urcorner$
if $k \cdot l \neq m$ then $\mathrm{ST} \vdash \ulcorner k \urcorner \cdot \ulcorner l \urcorner \neq \ulcorner m \urcorner$

The implications of a) cannot be reversed (at the moment). We cannot argue that

warning: if $\mathrm{ST} \vdash \ulcorner k \urcorner \cdot \ulcorner l \urcorner = \ulcorner m \urcorner$ then $k \cdot l = m$.

The reason is that ST, for all we know, might be an inconsistent theory so that ST would be able to prove every statement of set theory. There are difficult relations between truth ("$m < n$") and provability ("$\mathrm{ST} \vdash \ulcorner m < n \urcorner$") which also concern the mechanics behind the Gödel incompleteness theorems.

26 Finite sequences

For the incompleteness results we want to formalize first-order syntax *within* ST. Syntax is very much about finite sequences: words are finite sequences of symbols, sequents are finite sequences of formulas, formal derivations are finite sequences of sequents. Carrying out syntax within ST requires a good theory of finite sequences within ST. For the next definition observe that a set-theoretical natural number is equal to the set of all smaller numbers.

Definition 123. *w is a* finite sequence *if $w: n \rightarrow V$ for some $n \in \mathbb{N}$. We call n the* length *of w and write* $\mathrm{length}(w) = n$ *or $|s| = n$. For $i < |s|$ write w_i instead of $w(i)$. The sequence s will also be denoted by $(w_i)_{i<n}$, $(w_0, \ldots, w_{n-1})$ or $w_0 \ldots w_{n-1}$.*

Let V^* be the class of all finite sequences. For sequences $w = w_0 \ldots w_{m-1}$ and $w'_0 \ldots w'_{n-1}$ define the concatenation $w^\frown w'$ *to be the unique sequence with* $\mathrm{length}(w^\frown w') = m + n$ *and*

$$(w^\frown w')_i = \begin{cases} w_i\,, \text{ for } i < m \\ w'_{i-m}\,, \text{ for } m \leqslant i < m + n \end{cases}$$

We also write $w_0 \ldots w_{m-1} w'_0 \ldots w'_{n-1}$ or $w w'$ instead of $w^\frown w'$.

Lemma 124. *V^* with $^\frown$ has the properties of a* monoid with cancellation*:*

a) $^\frown$ is associative: $(w w') w'' = w (w' w'')$.

b) $\emptyset$ is a neutral element for $^\frown$: $\emptyset w = w \emptyset = w$.

c) $^\frown$ satisfies cancelation: if $u w = u' w$ then $u = u'$; if $w u = w u'$ then $u = u'$.

Exercise 13. Prove these laws by induction.

The associative law allows us to omit brackets in multiple concatenations and to write $w w' w''$ instead of $(w w')\, w''$. We can also view "single" sets s as sequences of length 1 , writing simply s instead of (s); then $w s$ stands for $w^\frown(s)$.

Finite sequences can be used to define "smallest" classes closed under given functions. Let us first introduce multi-argument functions.

Definition 125. *For a term A and $n \in \mathbb{N}$ let*

$$A^n = \{w \mid w: n \rightarrow A\}$$

be the n-fold cartesian product *of A. $F: A^n \rightarrow V$ is called an n-*ary function *on A.*

Note that $\emptyset\colon 0\to V$ *and* $A^0=\{\emptyset\}$. *So a* 0*-ary function* $F\colon V^0\to V$ *can be identified with its constant value* $F(\emptyset)$.

Exercise 14. Show in ST that $x^n\in V$.

26.1 Calculi

We have defined terms and formulas of first-order logic as *calculi* that produced, e.g., formulas out of other formulas and out of other syntactic material. We had used the meta-mathematical definition:

Definition 126. *The class* L^S *of all* S-formulas *is the smallest subclass of* S^* *such that*

a) $\bot\in L^S$ *(the false formula);*

b) $t_0\equiv t_1\in L^S$ *for all* S*-terms* $t_0,t_1\in T^S$ *(equality);*

c) $Rt_0\ldots t_{n-1}\in L^S$ *for all* n*-ary relation symbols* $R\in S$ *and all* S*-terms* $t_0,\ldots,t_{n-1}\in T^S$ *(relational formula);*

d) $\neg\varphi\in L^S$ *for all* $\varphi\in L^S$ *(negation);*

e) $(\varphi\to\psi)\in L^S$ *for all* $\varphi,\psi\in L^S$ *(implication);*

f) $\forall x\varphi\in L^S$ *for all* $\varphi\in L^S$ *and all variables* x *(universalisation).*

We want to capture these formation rules by a single function F such that

a) $(0)\overset{F}{\mapsto}\bot$;

b) $(1,t_0,t_1)\overset{F}{\mapsto}t_0\equiv t_1$ for $t_0,t_1\in T^S$;

c) $(2,R,t_0,\ldots,t_{n-1})\overset{F}{\mapsto}Rt_0\ldots t_{n-1}$ for $R\in S$ an n-ary relation symbol;

d) $(3,\varphi)\overset{F}{\mapsto}\neg\varphi$;

e) $(4,\varphi,\psi)\overset{F}{\mapsto}(\varphi\to\psi)$;

f) $(5,x,\varphi)\overset{F}{\mapsto}\forall x\varphi$ for $x\in\mathrm{Var}$.

The first element of the arguments signifies which formation rule is to be used; the further arguments consist of "previously formed" formulas and other material. We now prove that the smallest class closed under F can be expressed in ST.

Theorem 127. *Let* $F\colon D\to V$ *be given. A term* B *is called* F*-closed if whenever* $(z,s)\in D$ *and* $s\in B^*$ *then* $F(z,s)\in B$. *Then we can canonically define a class* A_F *which is the uniquely determined* $\subseteq$*-minimal* F*-closed class: if* B *is another* F*-closed term then* $A_F\subseteq B$. *We call* A the class generated by F, *or the* smallest class *such that ... , where ... stands for the properties that define* F *as in the example above.*

An F-derivation *of* $x\in A_F$ *is a finite sequence* $f\colon n+1\to V$ *such that*

$$\forall m\leqslant n\exists(z,s)\in D.s\in(f[m])^*\wedge f(m)=F(z,s)\wedge f(n)=x.$$

We view F as a generating function: if $F(x,s)=t$ and $s=(s_0,\ldots,s_{n-1})$ is a finite sequence then we can view t as being built from $s_0,\ldots,s_{n-1}$, possibly using the extra material x. An F-derivation is a finite sequences in which every step is generated from earlier steps using F.

Proof. Set

$$A = \{x \mid \exists n \in \mathbb{N} \exists f : n+1 \rightarrow V . f \text{ is an } F\text{-derivation of } x\}.$$

(1) A is F-closed.
Proof. Let $(y,t) \in D$ with $t \in A^*$. We have to show that $x = F(y,t) \in A$. Take $k \in \mathbb{N}$ such that $t = (t_0, ..., t_{k-1}) \in A^k$. For $l < k$ take an F-derivation $f_l : n_l + 1 \rightarrow V$ of t_l.

Then

$$f = f_0 {}^\frown f_1 {}^\frown ... {}^\frown f_{k-1} {}^\frown (x) : 1 + \sum_{l<k} (n_l + 1) \rightarrow V$$

is an F-derivation of x . $qed(1)$
(2) Let B be F-closed. Then $A \subseteq B$.
Proof. Let $f : n+1 \rightarrow V$ be an F-derivation. It suffices to prove by induction on $n' \leqslant n$ that $\forall m \leqslant n' . f(m) \in B$. For $n' = 0$ and $m = 0$ there is some $(z,s) \in D$ such that $s \in (f[0])^* = \{\emptyset\}$ and $f(0) = F(z, \emptyset)$. $\emptyset \in B^0$ and by the assumptions on B

$$f(0) = F(z, \emptyset) \in B.$$

For the induction step consider $n' + 1 \leqslant n$ and assume that $\forall m \leqslant n' . f(m) \in B$. It suffices to see that $f(n'+1) \in B$. Set $m = n' + 1$. Take some $(z,s) \in D$ such that $s \in (f[m])^*$ and $f(m) = F(z,s)$. By the inductive assumption $s \in B^*$ and by the assumptions on B

$$f(m) = F(z,s) \in B.$$

□

The class A generated by $\vec{F}$ can be seen as a least *fixed point* with respect to a certain closure under F. In the above construction, A is reached via derivations "from below". If there is some F-closed *set* z_0 then the fixed point can also be reached by intersections "from above":

$$A = \bigcap \{y \mid y \text{ is } F\text{-closed}\}.$$

The intersection on the right-hand side is non-trivial since z_0 is a factor of the intersection.

26.2 Gödelization of finite sequences

We want to Gödelize the syntactic notions of first-order logic. In the meta-theory these are defined using finite sequences. Therefore we have to Gödelize finite sequences.

Definition 128. *Let $w = w_0 ... w_{n-1}$ be a meta-theoretic finite sequence of objects w_i which possess a Gödelization $\ulcorner w_i \urcorner$. Then the Gödelization of w is defined as*

$$\ulcorner w \urcorner = (\ulcorner w_0 \urcorner, ..., \ulcorner w_{n-1} \urcorner).$$

One can show some correctness properties by meta-theoretic induction.

Proposition 129.

a) $\ulcorner w_0 ... w_{n-1} u_0 ... u_{m-1} \urcorner = \ulcorner w_0 ... w_{n-1} \urcorner {}^\frown \ulcorner u_0 ... u_{m-1} \urcorner$

b) ...

26.3 Finite and infinite sets

Finite sequences can also be used to formalize the notions of *finite* and *infinite*.

Definition 130. *A set x is* finite *if there is a finite sequence $f\in V^*$ such that $x=\operatorname{ran}(f)$. For a finite set x define its* cardinality *by*

$$\operatorname{card}(x)=\bar{\bar{x}}=\min\{n\,|\,\exists f\colon n\to x.x=\operatorname{ran}(f)\}$$

A set x is infinite *if it is not finite.*

Lemma 131.

a) $\emptyset$ is finite with $\operatorname{card}(\emptyset)=0$.

b) $\{x\}$ and $\{x,y\}$ are finite with $\operatorname{card}(\{x\})=1$, *and* $\operatorname{card}(\{x,y\})=2$ *iff $x\neq y$.*

c) If x is finite and every element of x is finite then $\bigcup x$ is finite.

d) If x is finite and $y\subseteq x$ then y is finite with $\operatorname{card}(y)\leqslant\operatorname{card}(x)$.

e) If x and y are finite then $x\cup y$ is finite with $\operatorname{card}(x\cup y)=\operatorname{card}(x)+\operatorname{card}(y)-\operatorname{card}(x\cap y)$.

f) If x is finite then $\mathcal{P}(x)$ is finite with $\operatorname{card}(\mathcal{P}(x))=2^{\operatorname{card}(x)}$.

g) If F is a function and x is finite then $F[x]$ is finite with $\operatorname{card}(F[x])\leqslant\operatorname{card}(x)$.

h) every $n\in\mathbb{N}$ is finite with $\operatorname{card}(n)=n$.

i) If $\mathbb{N}\in V$ then $\mathbb{N}$ is infinite.

Exercise 15. Prove the Lemma.

27 Formalizing syntax within ST

We now carry out the syntactic definitions of first-order logic within ST.

Definition 132. *Set*

a) $\ulcorner\equiv\urcorner=8801$ *(equality),*

b) $\ulcorner\neg\urcorner=172$ *(negation),*

c) $\ulcorner\rightarrow\urcorner=8594$ *(implication),*

d) $\ulcorner\bot\urcorner=8869$ *(false),*

e) $\ulcorner\forall\urcorner=8704$ *(universal quantifier),*

f) $\ulcorner(\urcorner=40$ *(left bracket),*

g) $\ulcorner)\urcorner=41$ *(right bracket),*

h) $v_n=(0,n,0)$ *for $n\in\mathbb{N}$ (the n-th variable),*

i) an n-ary relation symbol is a triple of the form $R=(1,x,n)$ with $x\in V$ and $n\in\mathbb{N}$,

j) an n-ary function symbol is a triple of the form $f=(2,x,n)$ with $x\in V$ and $n\in\mathbb{N}$.

k) Let $\operatorname{Var}=\{v_n\,|\,n\in\mathbb{N}\}$ *be the class of variables.*

l) Let $S_0=\{\ulcorner\equiv\urcorner,\ulcorner\neg\urcorner,\ulcorner\rightarrow\urcorner,\ulcorner\bot\urcorner,\ulcorner\forall\urcorner,\ulcorner(\urcorner,\ulcorner)\urcorner\}\cup\operatorname{Var}$ *be the class of basic symbols.*

m) A term S is a language if every $s\in S$ is a relation symbol or a function symbol.

Note that the basic symbols, the relation symbols, and the function symbols are all pairwise distinct. We also have an unlimited supply of relation and function symbols. The (set-theoretic) numbers that we have chosen for the symbols $\equiv, \neg, \ldots$ are their numbers in the Unicode font. This choice is in principle rather arbitrary. It is often convenient to use the metatheoretic symbols $\equiv, \neg, \ldots$ as abbreviations for their Gödelizations. So $(v_3 \equiv v_4)$ is a formula in the language of set theory, and at the same time it denotes a unique class term which is the finite sequence

$$\begin{aligned} (v_3 \equiv v_4) &= ((, v_3, \equiv, v_4,)) \\ &= (40, (0,3,0), 8801, (0,4,0), 41) \end{aligned}$$

We have encoutered such "overloading" of notation already with other set-theoretic formalizations: 5 can denote the meta-theoretic number "five" or the class term 5. Whether $(v_3 \equiv v_4)$ is to be read as an $\in$-formula or as a settheoretic term should be derivable from the context. Simple meta-theoretic properties should, however, "reflect" from the metatheory into ST:

$$\text{if } v_n \in \mathrm{Var} \text{ then } \mathrm{ST} \vdash v_n \in \mathrm{Var}\,.$$

"$v_n \in \mathrm{Var}$" on the left-hand side is a meta-theoretical syntactical property. "$v_n \in \mathrm{Var}$" on the right-hand side is an $\in$-formula which can be proved in ST, due to the simple definition of the term Var.

Definition 133. *(In* ST*) Let S be a language. A word over S is a finite sequence*

$$w: n \to S_0 \cup S.$$

for some number $n \in \mathbb{N}$ which is the length of w. Let S^* be the class of all words over S.

Consider a meta-theoretic language S and let $\ulcorner S \urcorner$ be a "Gödelization" of S such that for every symbol s in S:

$$\ulcorner s \urcorner \in \ulcorner S \urcorner.$$

Then for every meta-theoretic word w over S we have

$$\mathrm{ST} \vdash \ulcorner w \urcorner \text{ is a word over} \ulcorner S \urcorner.$$

Definition 134. *The class T^S of all S-*terms *is the smallest subclass of S^* such that*

a) $x \in T^S$ for all variables $x \in \mathrm{Var}$;

b) $f t_0 \ldots t_{n-1} \in T^S$ for all $n \in \mathbb{N}$, all n-ary function symbols $f = (2, y, n) \in S$, and all $t_0, \ldots, t_{n-1} \in T^S$.

Let $\ulcorner S \urcorner$ be a "Gödelization" of S such that for every n-ary function symbol s in S:

$$\ulcorner s \urcorner \in \ulcorner S \urcorner \text{ is of the form } \ulcorner s \urcorner = (2, x, \ulcorner n \urcorner).$$

In the above situation of the Gödelization of a language we have for every meta-theoretic S-term t that

$$\mathrm{ST} \vdash \ulcorner t \urcorner \text{ is an } \ulcorner S \urcorner\text{-term}.$$

This can be proved by a meta-theoretic induction on the length of derivations of t. The Gödelization of a derivation is again a derivation in ST, and every member of the derivation is an $\ulcorner S \urcorner$-term.

In ST one can prove the unique readability of terms.

Lemma 135. *For every term $t \in T^S$ exactly one of the following holds:*

a) $t \in \mathrm{Var}$*;*

b) there is a uniquely defined function symbol $f = (2, y, n) \in S$ and a uniquely defined sequence $t_0, ..., t_{n-1} \in T^S$ of terms such that $t = f t_0 ... t_{n-1}$.

Definition 136. *The class L^S of all S-*formulas *is the smallest class such that*

a) $\bot \in L^S$*;*

b) $t_0 \equiv t_1 \in L^S$ *for all S-terms $t_0, t_1 \in T^S$;*

c) $R t_0 ... t_{n-1} \in L^S$ *for all n-ary relation symbols $R = (1, y, n) \in S$ and all S-terms $t_0, ..., t_{n-1} \in T^S$;*

d) $\neg\varphi \in L^S$ *for all $\varphi \in L^S$;*

e) $(\varphi \to \psi) \in L^S$ *for all $\varphi, \psi \in L^S$;*

f) $\forall v_n \varphi \in L^S$ *for all $\varphi \in L^S$ and all $v_n = (0, n, 0) \in \mathrm{Var}$.*

Assuming a “correct” Gödelization of the language S we have for every meta-theoretic S-formula φ that

$$\mathrm{ST} \vdash \ulcorner \varphi \urcorner \text{ is an } \ulcorner S \urcorner\text{-formula.}$$

And in ST one prove the unique readability of formulas. Then simple syntactic recursions can be formalized in ST.

Definition 137. *For $t \in T^S$ define* $\mathrm{var}(t) \subseteq \{v_n | n \in \mathbb{N}\}$ *by recursion on (the lengths of) terms:*

- $\mathrm{var}(x) = \{x\}$*;*
- $\mathrm{var}(c) = \emptyset$*;*
- $\mathrm{var}(f t_0 ... t_{n-1}) = \bigcup_{i<n} \mathrm{var}(t_i)$.

Definition 138. *Für $\varphi \in L^S$ define the set of* free variables $\mathrm{free}(\varphi) \subseteq \{v_n | n \in \mathbb{N}\}$ *by recursion on (the lengths of) formulas:*

- $\mathrm{free}(t_0 \equiv t_1) = \mathrm{var}(t_0) \cup \mathrm{var}(t_1)$*;*
- $\mathrm{free}(R t_0 ... t_{n-1}) = \mathrm{var}(t_0) \cup ... \cup \mathrm{var}(t_{n-1})$*;*
- $\mathrm{free}(\neg\varphi) = \mathrm{free}(\varphi)$*;*
- $\mathrm{free}(\varphi \to \psi) = \mathrm{free}(\varphi) \cup \mathrm{free}(\psi)$.
- $\mathrm{free}(\forall x \varphi) = \mathrm{free}(\varphi) \setminus \{x\}$.

For $\Phi \subseteq L^S$ define the class $\mathrm{free}(\Phi)$ *of* free variables *as*

$$\mathrm{free}(\Phi) = \bigcup_{\varphi \in \Phi} \mathrm{free}(\varphi) .$$

Definition 139. *For a term $s \in T^S$, pairwise distinct variables $x_0, ..., x_{r-1}$ and terms $t_0, ..., t_{r-1} \in T^S$ define the* (simultaneous) substitution

$$s \frac{t_0 .... t_{r-1}}{x_0 ... x_{r-1}}$$

of $t_0, \ldots, t_{r-1}$ *for* $x_0, \ldots, x_{r-1}$ *by recursion:*

a) $x\,\frac{t_0\ldots t_{r-1}}{x_0\ldots x_{r-1}} = \begin{cases} x, \text{ if } x \neq x_0, \ldots, x \neq x_{r-1} \\ t_i\,, \text{ if } x = x_i \end{cases}$ *for all variables* x;

b) $(f s_0 \ldots s_{n-1})\,\frac{t_0\ldots t_{r-1}}{x_0\ldots x_{r-1}} = f s_0\,\frac{t_0\ldots t_{r-1}}{x_0\ldots x_{r-1}} \ldots s_{n-1}\,\frac{t_0\ldots t_{r-1}}{x_0\ldots x_{r-1}}$ *for all* n*-ary function symbols* $f \in S$.

One can check again, that Gödelizations commute with forming the set of free variables and with substitution.

Definition 140. *A finite sequence* $(\varphi_0, \ldots, \varphi_{n-1}, \varphi_n)$ *of* S*-formulas is called a* sequent*. The initial segment* $\Gamma = (\varphi_0, \ldots, \varphi_{n-1})$ *is the* antecedent *and* φ_n *is the* succedent *of the sequent. We usually write* $\varphi_0 \ldots \varphi_{n-1}\, \varphi_n$ *or* $\Gamma\, \varphi_n$ *instead of* $(\varphi_0, \ldots, \varphi_{n-1}, \varphi_n)$*. To emphasize the last element of the antecedent we may also denote the sequent by* $\Gamma'\, \varphi_{n-1}\, \varphi_n$ *with* $\Gamma' = (\varphi_0, \ldots, \varphi_{n-2})$.

Again, if $\Gamma\,\varphi$ is a meta-theoretic sequence then

$$\mathrm{ST} \vdash \ulcorner \Gamma\,\varphi \urcorner \text{ is a sequence.}$$

Definition 141. *The* sequent calculus *consists of the following (sequent-)rules:*

- *monotonicity* (MR) $\dfrac{\Gamma \quad\ \ \varphi}{\Gamma \ \ \psi \ \ \varphi}$
- *assumption* (AR) $\dfrac{}{\Gamma \ \ \varphi \ \ \varphi}$
- $\rightarrow$*-introduction* ($\rightarrow I$) $\dfrac{\Gamma \ \ \varphi \ \ \psi}{\Gamma \quad \varphi \rightarrow \psi}$
- $\rightarrow$*-elimination* ($\rightarrow E$) $\dfrac{\begin{array}{l}\Gamma \ \ \varphi \\ \Gamma \ \ \varphi \rightarrow \psi\end{array}}{\Gamma \ \ \psi}$
- $\perp$*-introduction* ($\perp I$) $\dfrac{\begin{array}{l}\Gamma \ \ \varphi \\ \Gamma \ \ \neg\varphi\end{array}}{\Gamma \ \ \perp}$
- $\perp$*-elimination* ($\perp E$) $\dfrac{\Gamma \ \ \neg\varphi \ \ \perp}{\Gamma \quad\ \ \varphi}$
- $\forall$*-introduction* ($\forall I$) $\dfrac{\Gamma \ \ \varphi\frac{y}{x}}{\Gamma \ \ \forall x\varphi}$, *if* $y \notin \mathrm{free}(\Gamma \cup \{\forall x\varphi\})$
- $\forall$*-elimination* ($\forall E$) $\dfrac{\Gamma \ \ \forall x\varphi}{\Gamma \ \ \varphi\frac{t}{x}}$, *if* $t \in T^S$
- $\equiv$*-introduction* ($\equiv I$) $\dfrac{}{\Gamma \ \ t \equiv t}$, *if* $t \in T^S$
- $\equiv$*-elimination* ($\equiv E$) $\dfrac{\begin{array}{l}\Gamma \ \ \varphi\frac{t}{x} \\ \Gamma \ \ t \equiv t'\end{array}}{\Gamma \ \ \varphi\frac{t'}{x}}$

The provability relation *is the smallest subclass* pr $\subseteq$Seq(S) *which is closed under these rules. For* A *an arbitrary class of formulas and* φ *a formula define the binary relation "*φ *is provable from* A*"*

$$\mathrm{pv}(A,\varphi)=\exists n\in\mathbb{N}\exists(\varphi_i)_{i<n}\,(\forall i<n\,\varphi_i\in A\wedge(\varphi_0,\dots,\varphi_{n-1},\varphi)\in\mathrm{pr})\,.$$

A derivation, or a formal proof, in the sequent calculus is a finite sequence of sequents according to the above derivation rules. If such a derivation D *ends in a sequent* $(\varphi_0,\dots,\varphi_{n-1},\varphi)$ *with* $\forall i<n\varphi_i\in A$ *then we write*

$$\mathrm{pf}(D,A,\varphi).$$

Finally, we formalize consistency *of first-order theories by*

$$\mathrm{Con}(A)=\neg\mathrm{pv}(A,\ulcorner\bot\urcorner)$$

Proposition 142. *Suppose that* A *is a meta-theoretical class of formulas with a Gödelization* $\ulcorner A\urcorner$ *such that for any meta-theoretical formula* $\varphi\in A$ *we have* $\mathrm{ST}\vdash\ulcorner\varphi\urcorner\in\ulcorner A\urcorner$ *. Then*

a) If D *is, meta-theoretically, a formal proof of* φ *within* A*, then* $\mathrm{ST}\vdash\mathrm{pf}(\ulcorner D\urcorner,\ulcorner A\urcorner,\ulcorner\varphi\urcorner)$.

b) If $A\vdash\varphi$*, meta-theoretically, then* $\mathrm{ST}\vdash\mathrm{pv}(\ulcorner A\urcorner,\ulcorner\varphi\urcorner)$.

Proof. For *b*) observe that if $A\vdash\varphi$ in the metatheory then there must be a meta-theoretical formal derivation of $\varphi_0\dots\varphi_{n-1}\varphi$ where $\varphi_0,\dots,\varphi_{n-1}\in A$. □

Note that from the meta-theoretical consistency of A we can in general *not* infer that $\mathrm{ST}\vdash\mathrm{Con}(\ulcorner A\urcorner)$; consistency is usually not expressed by some *finite* witness (like a derivation) but by a model which may not be transferable into ST.

28 ST in ST

Definition 143. *Let* $\ulcorner\in\urcorner=(1,8712,2)$ *be the Gödelization of the standard* $\in$*-symbol of set theory. Let* $\ulcorner L^{\in}\urcorner=\{\ulcorner\in\urcorner\}$ *be the Gödelized language of set theory.*

Definition 144. *Gödelize the axioms of* ST *as follows:*

a) Let $\ulcorner \mathit{Ext}\urcorner=\ulcorner\forall x\forall y(\forall z(z\in x\leftrightarrow z\in y)\rightarrow x\equiv y)\urcorner$

b) Let $\ulcorner \mathit{Pair}\urcorner=\ulcorner\forall x\forall y\exists z\forall w\,(w\in z\leftrightarrow w\equiv x\vee w\equiv y)\urcorner$

c) Let $\ulcorner \mathit{Union}\urcorner=\ulcorner\forall x\exists y\forall z(z\in y\leftrightarrow\exists w(w\in x\wedge z\in w))\urcorner$

d) Let $\ulcorner \mathit{Pot}\urcorner=\ulcorner\forall x\exists y\forall z(z\in y\leftrightarrow\forall w(w\in z\rightarrow w\in x))\urcorner$

e) Let

$$\ulcorner\mathrm{Sep}\urcorner=\{\forall x_1\dots\forall x_n\forall x\exists y\forall z\,(z\in y\leftrightarrow z\in x\wedge\varphi(z,x_1,\dots,x_n))\mid\varphi(z,x_1,\dots,x_n)\in\ulcorner L^{\in}\urcorner\}.$$

Here $\ulcorner\mathrm{Sep}\urcorner$ *is the image of* $\ulcorner L^{\in}\urcorner$ *under the recursively defined syntactic function*

$$\varphi(z,x_1,\dots,x_n)\mapsto\forall x_1\dots\forall x_n\forall x\exists y\forall z\,(z\in y\leftrightarrow z\in x\wedge\varphi(z,x_1,\dots,x_n));$$

this function identifies the variables z, x_1, ..., x_n *in* φ*, forms the concatenation* $\forall x\exists y\forall z\,(z\in y\leftrightarrow z\in x\wedge\varphi(z,x_1,\dots,x_n))$ *and prefixes it with the quantifiers* $\forall x_1\dots\forall x_n$ *.*

f) *Let*

$$\ulcorner\mathrm{Rep}\urcorner = \{\forall x_1 ... \forall x_n (\forall x \forall y \forall y'((\varphi(x, y, x_1, ..., x_n) \wedge \varphi(x, y', x_1, ..., x_n)) \rightarrow y \equiv y') \rightarrow \forall u \exists v \forall y\, (y \in v \leftrightarrow \exists x(x \in u \wedge \varphi(x, y, x_1, ..., x_n))))|\, \varphi(x, y, x_1, ..., x_n) \in \ulcorner L^{\in}\urcorner\}.$$

g) *Let*

$$\ulcorner\mathrm{Found}\urcorner = \{\forall x_1 ... \forall x_n (\exists x \varphi(x, x_1, ..., x_n) \rightarrow \exists x(\varphi(x, x_1, ..., x_n) \wedge \forall x'(x' \in x \rightarrow \neg\varphi(x', x_1, ..., x_n))))|\, \varphi(x, x_1, ..., x_n) \in \ulcorner L^{\in}\urcorner\}.$$

h) *Let*

$$\ulcorner\mathrm{ST}\urcorner = \{\ulcorner\mathrm{Ext}\urcorner, \ulcorner\mathrm{Pair}\urcorner, \ulcorner\mathrm{Union}\urcorner, \ulcorner\mathrm{Pot}\urcorner\} \cup \ulcorner\mathrm{Sep}\urcorner \cup \ulcorner\mathrm{Rep}\urcorner \cup \ulcorner\mathrm{Found}\urcorner.$$

Since provability from (Gödelized) set theory will be our main concern, we define

Definition 145. *Write* $\mathrm{pv}_{\mathrm{ST}}(\varphi)$ *for* $\mathrm{pv}(\ulcorner\mathrm{ST}\urcorner, \varphi)$ *and* $\mathrm{pf}_{\mathrm{ST}}(D, \varphi)$ *for* $\mathrm{pf}(D, \ulcorner\mathrm{ST}\urcorner, \varphi)$.

These Gödelizations satisfy the usual correctness properties:

Proposition 146.

a) *If* φ *is a meta-theoretical* $\in$*-formula then* $\mathrm{ST} \vdash \ulcorner\varphi\urcorner \in \ulcorner L^{\in}\urcorner$.

b) *If* φ *is a meta-theoretical axiom of* ST *then* $\mathrm{ST} \vdash \ulcorner\varphi\urcorner \in \ulcorner\mathrm{ST}\urcorner$.

c) *If D is, meta-theoretically, a formal proof of* φ *within* ST, *then* $\mathrm{ST} \vdash \mathrm{pf}_{\mathrm{ST}}(\ulcorner D\urcorner, \ulcorner\varphi\urcorner)$.

d) *If* $\mathrm{ST} \vdash \varphi$, *meta-theoretically, then* $\mathrm{ST} \vdash \mathrm{pv}_{\mathrm{ST}}(\ulcorner\varphi\urcorner)$.

We shall later deal with the question whether $\mathrm{ST} \vdash \mathrm{Con}(\ulcorner\mathrm{ST}\urcorner)$, i.e., whether set theory can prove its own consistency.

29 The undefinability of truth

The proof of the following fix point theorem is based on the trick

$$\varphi(v_0(v_0))(\varphi(v_0(v_0))) \leftrightarrow \varphi(\varphi(v_0(v_0))(\varphi(v_0(v_0))))$$

which can be adapted to our domain.

Theorem 147. *Let* $\varphi(v_0)$ *be an* $\in$*-formula. Then there is an* $\in$*-sentence* θ *without free variables which is a* fix point *of* φ, *i.e.,*

$$\mathrm{ST} \vdash \theta \leftrightarrow \varphi(\ulcorner\theta\urcorner).$$

The sentence θ can be viewed as expressing "this sentence has the property φ" or "I have the property φ". The sentence is "reflexive" in the sense that it talks about "itself". Such reflexivity is the base for paradoxa and incompleteness.

Proof. For an $\in$-formula $\chi(v_0)$ and another $\in$-formula χ', $\chi(\ulcorner\chi'\urcorner)$ is obtained by inserting the abstraction term $\ulcorner\chi'\urcorner$ in χ . By the recursive elimination rules for abstraction terms, $\chi(\ulcorner\chi'\urcorner)$ is an $\in$-formula. The operation $\chi, \chi' \mapsto \chi(\ulcorner\chi'\urcorner)$ is a straightforward but tedious syntactic manipulation of symbol sequences which can be defined ("Gödelized") in ST. So there is a canonical abstraction term $\mathrm{Sub}(v_0, v_1)$ such that

$$\mathrm{ST} \vdash \mathrm{Sub}: \ulcorner L^{\in}\urcorner \times \ulcorner L^{\in}\urcorner \rightarrow \ulcorner L^{\in}\urcorner$$

where for all metatheoretic $\in$-formulas χ, χ':

$$\mathrm{ST} \vdash \mathrm{Sub}(\ulcorner\chi\urcorner, \ulcorner\chi'\urcorner) = \ulcorner\chi(\ulcorner\chi'\urcorner)\urcorner.$$

Then let

$$\psi(v_0) = \varphi(\mathrm{Sub}(v_0, v_0)).$$

In ST, we have the equivalences

$$\psi(\ulcorner\psi\urcorner) \leftrightarrow \varphi(\mathrm{Sub}(\ulcorner\psi\urcorner, \ulcorner\psi\urcorner)) \leftrightarrow \varphi(\ulcorner\psi(\ulcorner\psi\urcorner)\urcorner).$$

Then $\theta = \psi(\ulcorner\psi\urcorner)$ is a fix point of φ. □

The operation $v_0(v_0)$ of "selfapplication" is more common in computing, where a program is also considered as data and a program can take itself as input. Another theory where such constructions can be carried out is the λ-calculus.

Definition 148. *An $\in$-formula $\psi(v_0)$ is a* definition of truth *in* ST *if for all $\in$-sentences θ:*

$$\mathrm{ST} \vdash (\theta \leftrightarrow \psi(\ulcorner\theta\urcorner)).$$

Using the fix point theorem we can easily show Tarski's theorem on the *undefinability of truth*:

Theorem 149. *If* ST *is consistent, there is no definition of truth.*

Proof. Assume that $\psi(v_0)$ were a definition of truth. By the fix point theorem there is an $\in$-sentence θ such that

$$\mathrm{ST} \vdash (\theta \leftrightarrow \neg\psi(\ulcorner\theta\urcorner)).$$

Since $\psi(v_0)$ is a definition of truth we also have

$$\mathrm{ST} \vdash (\theta \leftrightarrow \psi(\ulcorner\theta\urcorner)).$$

Together

$$\mathrm{ST} \vdash (\neg\psi(\ulcorner\theta\urcorner) \leftrightarrow \psi(\ulcorner\theta\urcorner)),$$

contradiction. □

30 Gödel's incompleteness theorems

Naively, one might identify mathematical truth with provability: a theorem is "true" if it has a proof. Then the formula $\mathrm{pv}_{\mathrm{ST}}(v_0)$ looks like a good candidate for a definition of truth. The contradiction in the proof of the indefinability of truth was obtained by a fixed point for the negated property. This would be a sentence θ such that

$$\mathrm{ST} \vdash \theta \leftrightarrow \neg\mathrm{pv}_{\mathrm{ST}}(\ulcorner\theta\urcorner),$$

expressing "this sentence is not provable". One would expect that ST is not able to prove θ nor $\neg\theta$, unless ST is inconsistent so that it would prove everything.

Suppose that $\mathrm{ST} \vdash \theta$. Then $\mathrm{ST} \vdash \mathrm{pv}_{\mathrm{ST}}(\ulcorner\theta\urcorner)$. Hebce $\mathrm{ST} \vdash \neg\theta$, and ST would be inconsistent. Suppose that $\mathrm{ST} \vdash \neg\theta$. Then $\mathrm{ST} \vdash \mathrm{pv}_{\mathrm{ST}}(\ulcorner\neg\theta\urcorner)$. If this would imply $\mathrm{ST} \vdash \neg\mathrm{pv}_{\mathrm{ST}}(\ulcorner\theta\urcorner)$, we would obtain the contradiction $\mathrm{ST} \vdash \theta$.

We shall make this idea exact by modifying the formula $\mathrm{pv}_{\mathrm{ST}}(v_0)$ a bit.

Theorem 150. *If* ST *is consistent then* ST *is incomplete, i.e., there is an* $\in$*-sentence* φ *such that* $\mathrm{ST} \nvdash \varphi$ *and* $\mathrm{ST} \nvdash \neg\varphi$.

Proof. A formal proof from $\ulcorner \mathrm{ST} \urcorner$ consists of a combination of natural numbers: symbols are (triples of) natural numbers, formula are finite sequences of symbols, sequents are finite sequences of formulas, and formal proofs are finite sequences of sequents. We can rank proofs by the largest natural number involved in it.

Define the rank of a formula φ as $\mathrm{rk}(\varphi) = \max\{n \in \mathbb{N} \mid v_n \text{ occurs in } \varphi\} + \mathrm{length}(\varphi)$.

Define the rank of a sequent $\varphi_0 ... \varphi_{l-1}$ as $\mathrm{rk}(\varphi_0 ... \varphi_{l-1}) = \max_{i<l} \mathrm{rk}(\varphi_i) + l$.

Define the rank of a formal proof $s_0 ... s_{k-1}$ as $\mathrm{rk}(s_0 ... s_{l-1}) = \max_{i<l} \mathrm{rk}(s_i) + l$.

The rank function can be defined metatheoretically as well as in ST with usual reflection properties. Metatheoretically, for every natural number n there are only finitely many formulas, sequents, and formal proofs of rank $\leqslant n$. This fact reflects down to ST: there is a metatheoretical list $D_0, ..., D_{m-1}$ of formal proofs of rank $\leqslant n$ such that

$$\mathrm{ST} \vdash \mathrm{pf}(D) \wedge \mathrm{rk}(D) \leqslant \ulcorner n \urcorner \rightarrow D = \ulcorner D_0 \urcorner \vee D = \ulcorner D_1 \urcorner \vee ... \vee D = \ulcorner D_{m-1} \urcorner.$$

Let us now assume for a contradiction that ST is complete and consistent. We show that the formula

$$\psi(v_0) = \exists D(\mathrm{pf}(D, v_0) \wedge \forall D'(\mathrm{rk}(D') \leqslant \mathrm{rk}(D) \rightarrow \neg \mathrm{pf}(D', \neg v_0))),$$

expressing that v_0 is proved "before" $\neg v_0$ is proved, is a definition of truth.

Let θ be an $\in$-sentence.

Case 1. $\mathrm{ST} \vdash \theta$. Then there is a metatheoretic formal proof D for θ in ST. Let $n = \mathrm{rk}(D)$ and let $D_0, ..., D_{m-1}$ the list of metatheoretic formal proofs of rank $\leqslant n$ as above. Since ST is assumed to be consistent, D_i with $i < m$ is not a formal proof of $\neg\theta$. This is a simple syntactic property which reflects to ST: $\neg \mathrm{pf}(\ulcorner D_i \urcorner, \ulcorner \neg\theta \urcorner)$. By reflection and the properties of the list $D_0, ..., D_{m-1}$

$$\mathrm{pf}(\ulcorner D \urcorner, \ulcorner \theta \urcorner) \wedge \forall D'(\mathrm{rk}(D') < \mathrm{rk}(D) \rightarrow \neg \mathrm{pf}(D', \ulcorner \neg\theta \urcorner)).$$

Hence $\mathrm{ST} \vdash \psi(\ulcorner \theta \urcorner)$ and so $\mathrm{ST} \vdash \theta \leftrightarrow \psi(\ulcorner \theta \urcorner)$.

Case 2. $\mathrm{ST} \nvdash \theta$. Since ST is assumed to be complete, $\mathrm{ST} \vdash \neg\theta$. Then there is a metatheoretic formal proof D' for $\neg\theta$ in ST. Let $n = \mathrm{rk}(D')$ and let $D_0, ..., D_{m-1}$ the list of metatheoretic formal proofs of rank $\leqslant n$ as above. Since ST is assumed to be consistent, D_i with $i < m$ is not a formal proof of θ.

Work in ST. Assume for a contradiction that $\psi(\ulcorner \theta \urcorner)$. Take a derivation D such that $\mathrm{pf}(D, \ulcorner \theta \urcorner)$. Then $D \neq \ulcorner D_i \urcorner$ for $i < m$ and so $\mathrm{rk}(D) > \ulcorner n \urcorner$. $\mathrm{rk}(\ulcorner D' \urcorner) < \mathrm{rk}(\ulcorner D \urcorner)$ and $\mathrm{pf}(\ulcorner D' \urcorner, \ulcorner \theta \urcorner)$. This implies $\neg\psi(\ulcorner \theta \urcorner)$.

Hence $\mathrm{ST} \vdash \neg\psi(\ulcorner \theta \urcorner)$ and so $\mathrm{ST} \vdash \theta \leftrightarrow \psi(\ulcorner \theta \urcorner)$.

But this contradicts the undefinability of truth. □

Analysing the proof, we can concretely exhibit an "undecided" sentence θ.

Theorem 151. *If* ST *is consistent then one can construct an* $\in$*-sentence* θ *such that* $\mathrm{ST} \nvdash \theta$ *and* $\mathrm{ST} \nvdash \neg\theta$.

Proof. As in the proof of the previous theorem let

$$\psi(v_0) = \exists D(\mathrm{pf}(D, v_0) \wedge \forall D'(\mathrm{rk}(D') \leqslant \mathrm{rk}(D) \rightarrow \neg \mathrm{pf}(D', \neg v_0))),$$

be the formula "v_0 is proved, before $\neg v_0$".

By the proof of the fix point theorem one can concretely construct a sentence θ from the formula ψ such that

$$\mathrm{ST} \vdash \theta \leftrightarrow \neg\psi(\ulcorner\theta\urcorner).$$

This sentence expresses "I cannot be proved before my negation". To show that θ is not decided by ST we reuse the arguments from the proof of the incompleteness theorem.
(1) $\mathrm{ST} \nvdash \theta$.
Proof. Assume $\mathrm{ST} \vdash \theta$. Let D be a metatheoretic formal proof for θ in ST. Let $n = \mathrm{rk}(D)$ and let $D_0, \ldots, D_{m-1}$ the list of metatheoretic formal proofs of rank $\leqslant n$ as above. Since ST is assumed to be consistent, D_i with $i < m$ is not a formal proof of $\neg\theta$. As above,

$$\mathrm{pf}(\ulcorner D\urcorner, \ulcorner\theta\urcorner) \wedge \forall D'(\mathrm{rk}(D') < \mathrm{rk}(D) \rightarrow \neg\mathrm{pf}(D', \ulcorner\neg\theta\urcorner)).$$

Hence $\mathrm{ST} \vdash \psi(\ulcorner\theta\urcorner)$. By the fix point property, $\mathrm{ST} \vdash \neg\theta$ and so ST is inconsistent. *qed*(1)
(2). $\mathrm{ST} \nvdash \neg\theta$.
Proof. Assume $\mathrm{ST} \vdash \neg\theta$. Let D' be a metatheoretic formal proof for $\neg\theta$ in ST. Let $n = \mathrm{rk}(D')$ and let $D_0, \ldots, D_{m-1}$ the list of metatheoretic formal proofs of rank $\leqslant n$ as above. Since ST is assumed to be consistent, D_i with $i < m$ is not a formal proof of θ.

Work in ST. Assume for a contradiction that $\psi(\ulcorner\theta\urcorner)$. Take a derivation D such that $\mathrm{pf}(D, \ulcorner\theta\urcorner)$. Then $D \neq \ulcorner D_i\urcorner$ for $i < m$ and so $\mathrm{rk}(D) > \ulcorner n\urcorner$. $\mathrm{rk}(\ulcorner D'\urcorner) < \mathrm{rk}(\ulcorner D\urcorner)$ and $\mathrm{pf}(\ulcorner D'\urcorner, \ulcorner\theta\urcorner)$. This implies $\neg\psi(\ulcorner\theta\urcorner)$.

Hence $\mathrm{ST} \vdash \neg\psi(\ulcorner\theta\urcorner)$. By the fix point property, $\mathrm{ST} \vdash \theta$ and so ST is inconsistent. □

$\mathrm{Con}(\ulcorner\mathrm{ST}\urcorner)$ formalizes that the system ST is consistent. Gödel's second incompleteness theorem states that ST cannot prove its own consistency, i.e., that "by finitary means" mathematics cannot prove the consistency of ST and therefore not the consistency of mathematics.

Theorem 152. *If* ST *is consistent then* $\mathrm{ST} \nvdash \mathrm{Con}(\ulcorner\mathrm{ST}\urcorner)$.

Proof. By the fix point theorem there is an $\in$-sentence θ such that

$$\mathrm{ST} \vdash (\theta \leftrightarrow \neg\mathrm{pv}_{\mathrm{ST}}(\ulcorner\theta\urcorner)).$$

The sentence θ formalizes "this sentence is not provable".
(1) If ST is consistent then $\mathrm{ST} \nvdash \theta$.
Proof. Assume $\mathrm{ST} \vdash \theta$. Then $\mathrm{ST} \vdash \mathrm{pv}_{\mathrm{ST}}(\ulcorner\theta\urcorner)$. By the fix point property, $\mathrm{ST} \vdash \neg\theta$. Hence ST is inconsistent. *qed*(1)
(2) $\mathrm{ST} \vdash \mathrm{Con}(\ulcorner\mathrm{ST}\urcorner) \rightarrow \neg\mathrm{pv}_{\mathrm{ST}}(\ulcorner\theta\urcorner)$.
Proof. This is the reflection of (1) into ST. Work in ST and assume $\mathrm{pv}_{\mathrm{ST}}(\ulcorner\theta\urcorner)$. Take a derivation of $\ulcorner\theta\urcorner$. This can be Gödelized within the theory ST so that

$$\mathrm{pv}_{\mathrm{ST}}(\ulcorner\mathrm{pv}_{\mathrm{ST}}(\ulcorner\theta\urcorner)\urcorner).$$

Reflecting the fix point property yields

$$\mathrm{pv}_{\mathrm{ST}}(\ulcorner(\theta \leftrightarrow \neg\mathrm{pv}_{\mathrm{ST}}(\ulcorner\theta\urcorner))\urcorner).$$

By the properties of the $\mathrm{pv}_{\mathrm{ST}}$-predicate we also have

$$\mathrm{pv}_{\mathrm{ST}}(\ulcorner\neg\theta\urcorner),$$

so that $\mathrm{Con}(\ulcorner\mathrm{ST}\urcorner)$ is false. *qed*(2)

Assume now that $\mathrm{ST} \vdash \mathrm{Con}(\ulcorner\mathrm{ST}\urcorner)$. By (2), $\mathrm{ST} \vdash \neg\mathrm{pv}_{\mathrm{ST}}(\ulcorner\theta\urcorner)$. By the fix point property, $\mathrm{ST} \vdash \theta$. So by (1), ST is inconsistent. □

Gödel's incompleteness theorems have a number of mathematical consequences and extensions. They relate to other limiting results in logic and computability theory and they also influence general discussions about the limits of knowledge.

Corollary 153. *There is an* $\in$*-sentence* θ *such that if* ST *is consistent, then both* $\mathrm{ST}\cup\{\theta\}$ *and* $\mathrm{ST}\cup\{\neg\theta\}$ *are also consistent.*

Instead of $\mathrm{ST}\cup\{\theta\}$ let us also write $\mathrm{ST}+\theta$

Corollary 154. *If* ST *is consistent then* $\mathrm{ST}+\neg\mathrm{Con}(\ulcorner\mathrm{ST}\urcorner)$ *is consistent.*

If ST is consistent, then a (metatheoretic) standard model of ST satisfies also $\mathrm{Con}(\ulcorner\mathrm{ST}\urcorner)$. A model of $\mathrm{ST}+\neg\mathrm{Con}(\ulcorner\mathrm{ST}\urcorner)$ is thus a nonstandard model in which there is a nonstandard object which is a formal proof inconsistency proof of $\ulcorner\mathrm{ST}\urcorner$. From outside the nonstandard model this proof appears to be a fascinating infinite object that locally respects the rules of the sequent calculus but which ends in $\bot$. Although much more complex it may be pictured like a nonstandard natural number ∞ which is reached by a process which locally is just the successor operation $n\mapsto n+1$ but ends in ∞.

31 Zermelo-Fraenkel set theory

"in der Mathematik giebt es kein Ignorabimus!". David Hilbert (1900)

"Wir müssen wissen, wir werden wissen". David Hilbert (1930)

A naive reaction to the incompleteness theorems might be: if θ is not decided by ST, maybe one can adjoin θ or $\mathrm{Con}(\ulcorner\mathrm{ST}\urcorner)$ to obtain completeness. However:

Theorem 155. *Let* ST' *be a consistent extension of* ST *by a finite list* $\theta_0,\ldots,\theta_{n-1}\in L^{\in}$ *of axioms. Let* $\ulcorner\mathrm{ST}'\urcorner$ *be the Gödelization of* ST'*. Then*

a) *Truth is undefinable in* ST'*.*

b) ST' *is incomplete.*

c) $\mathrm{ST}'\nvdash\mathrm{Con}(\ulcorner\mathrm{ST}'\urcorner)$.

Proof. Redo the above proofs with $\mathrm{pf}\Big(D,\ulcorner\mathrm{ST}'\urcorner,\varphi\Big)$ and $\mathrm{pv}\Big(\ulcorner\mathrm{ST}'\urcorner,\varphi\Big)$ instead of $\mathrm{pf}_{\mathrm{ST}}(D,\varphi)$ and $\mathrm{pv}_{\mathrm{ST}}(\varphi)$. □

The same holds if ST' is an extension of ST by some Gödelizable schema of axioms.

Definition 156. *Let S and T be theories in the language of set theory which extend* ST*. We say that S has greater* consistency strength *than T if the consistency of S implies the consistency of T.*

The ordering of theories by their consistency strengths is a major topic in axiomatic set theory.

A particularly interesting extension of ST is Zermelo-Fraenkel set theory

$$\mathrm{ZF}=\mathrm{ST}+\mathrm{Inf}.$$

ZF implies that $\mathbb{N}$ is a set. The usual number systems $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ can be obtained set-theoretically from $\mathbb{N}$ (this will be discussed in the lecture course "Set Theory"). Together with the abstract notions of relation and functions which are already available in ST this indicates that ZF can be used as a foundation for "mathematics". Usually one also adds the *axiom of choice* to ZF in the form of Zorn's lemma:

$$\mathrm{ZFC} = \mathrm{ZF} + \text{Zorn's lemma}.$$

By the above theorems, ZFC is an incomplete theory (if it is consistent), but there are no further generally accepted axioms that one could add. So one can say

$$\text{Mathematics} = \mathrm{ZFC}.$$

Theorem 157. *Assume that the theory* ZFC *is consistent. Then*

a) $\mathrm{ZFC} \vdash \mathrm{Con}(\ulcorner \mathrm{PA} \urcorner)$.

b) $\mathrm{ZFC} \nvdash \mathrm{Con}(\ulcorner \mathrm{ZFC} \urcorner)$.

Proof. We have seen before that $\mathbb{N}, +, \cdot, 0, 1$ satisfy the Peano axioms. Since $\mathbb{N}$ is a set in ZFC we can reflect those arguments into ZFC to show that

$$(\mathbb{N}, +, \cdot, 0, 1) \vDash \ulcorner \mathrm{PA} \urcorner.$$

Then PA is consistent within ZFC. □

The incompleteness theorems for ZFC can be read as incompleteness theorems for all of mathematics: we assume that ZFC is consistent. Then

1. mathematics (=ZFC) is incomplete;
2. mathematics cannot prove the consistency of mathematics ($\mathrm{ZFC} \nvdash \mathrm{Con}(\ulcorner \mathrm{ZFC} \urcorner)$).

As emphasized by 2., we can only assume (believe?) the consistency of mathematics. And then there are mathematical statements that cannot be decided by the axioms. Gödel's arguments produce artificial statements, often called *Gödel sentences*, that are left undecided by the axioms. Indeed one can show in the theory of models of set theory, that there are natural set theoretical properties which are left open by $\mathrm{ZF}(C)$:

$$\text{if ZF is consistent then } \mathrm{ZF} + \mathrm{ZL} \text{ and } \mathrm{ZF} + \neg\mathrm{ZL}$$

are consistent, where ZL denotes Zorn's Lemma.

The incompleteness theorems of ZFC, rather than the original Gödel theorems, present a strong philosophical and epistomological barrier, which has given rise to deep philosophical theories.

32 Finite set theory and Peano arithmetic

Another extension of ST is *finite set theory*

$$\mathrm{FST} = \mathrm{ST} + \neg\mathrm{Inf}.$$

A standard model for this theory is the class of all finite sets, whose elements and elements of elements etc. are all finite. Such sets are called *hereditarily finite*. This class can be defined metatheoretically as follows: define a function $(V_n)_{n \in \mathbb{N}}$ recursively by:

$$\begin{aligned} V_0 &= \emptyset \\ V_{n+1} &= \mathcal{P}(V_n) \end{aligned}$$

Then let

$$V_\omega = \bigcup_{n\in\mathbb{N}} V_n\,.$$

One can show that V_ω satisfies the axioms of ST, and since every element of V_ω is finite, it satisfies $\neg\mathrm{Inf}$. As before we have

Theorem 158. *If the theory* FST *is consistent then* FST *is incomplete and*

$$\mathrm{FST}\nvdash \mathrm{Con}(\ulcorner\mathrm{FST}\urcorner).$$

FST is incomplete although it appears to adequately describe the hereditarily finite sets which seem to be definite and concrete mathematical objects.

We shall see that the model V_ω is in a certain way equivalent to the standard model $(\mathbb{N},+,\cdot,0,1)$ of Peano arithmetic.

Let us work in the theory FST. FST proves that every set is finite, i.e., the range of a finite sequence.

Lemma 159. *Assume* FST*. Define the V_n-hierarchy within* FST *by:*

$$\begin{aligned} V_0 &= \emptyset \\ V_{n+1} &= \mathcal{P}(V_n) \end{aligned}$$

Then let

$$V_\omega = \bigcup_{n\in\mathbb{N}} V_n\,.$$

Then

a) if $m\leqslant n\in\mathbb{N}$ then $V_m\subseteq V_n\subseteq V_\omega$;

b) every V_n is finite for $n\in\mathbb{N}$;

c) $V=V_\omega$, i.e., every set x is an element of V_n for some $n\in\mathbb{N}$;

d) every set is finite.

Proof. *a*) We prove the claim by complete induction on $n\in\mathbb{N}$. The claim is obvious for $n=0$. Now assume the claim holds for n and let $m\leqslant n+1$. Obviously $V_0\subseteq V_{n+1}$. So consider $m=k+1>0$. By the inductive assumption, $V_k\subseteq V_n$ and then

$$V_m=\mathcal{P}(V_k)\subseteq\mathcal{P}(V_n)=V_{n+1}\,.$$

b) is proved by induction.
c) Assume for a contradiction that $x\notin V_\omega$. By the foundation schema we may assume that x is $\in$-minimal with this property. So $\forall y\in x. y\in V_\omega$ and hence $x\subseteq V_\omega$. Define a function $f\colon x\to\mathbb{N}$ by

$$f(y)=\min\{n\in\mathbb{N}\mid y\in V_n\}.$$

(1) There is $n_0\in\mathbb{N}$ such that $f[x]\subseteq n_0$.
Proof. Assume not. Then for every $m\in\mathbb{N}$ there is $y\in x$ such that $f(y)\geqslant m$. This means that $f[x]$ is unbounded in $\mathbb{N}$ and

$$\mathbb{N}=\bigcup f[x].$$

By replacement and the union axiom, the right-hand-side is a set, which contradicts $\neg\mathrm{Inf}$.
qed(1)
(2) $x\subseteq V_{n_0}$.
Proof. If $y\in x$ then $f(y)<n_0$ and so $y\in V_{f(y)}\subseteq V_{n_0}$. *qed*(2)

But then $x \in \mathcal{P}(V_{n_0}) = V_{n_0+1} \subseteq V_\omega$ which contradicts the assumption $x \notin V_\omega$.
d) Consider a set x. By *c*), let $n \in \mathbb{N}$ such that $x \in V_n$. Then $n = m + 1$ is a successor and $x \in \mathcal{P}(V_m)$ satisfies $x \subseteq V_m$. By *b*), x is finite. □

We can now define a bijection between V and $\mathbb{N}$.

Lemma 160. *Define maps $f_n: V_n \to \mathbb{N}$ recursively:*

$$\begin{aligned} f_0 &= \emptyset, \\ f_{n+1} &= f_n \cup \left\{ \left(x, \sum_{y \in x} 2^{f_n(y)} \right) \mid x \in V_{n-1} \setminus V_n \right\}. \end{aligned}$$

Then the map

$$F = \bigcup_{n \in \mathbb{N}} f_n : V = \bigcup_{n \in \mathbb{N}} V_n \longrightarrow \mathbb{N}.$$

is a bijection.

Proof. Observe that

$$\forall n \in \mathbb{N}. f_n = F \upharpoonright V_n.$$

So

$$\forall x. F(x) = \sum_{y \in x} 2^{F(y)}.$$

We prove inductively that $F \upharpoonright V_n$ is injective. So assume that $F \upharpoonright V_n$ is injective. Let x, $x' \in V_{n+1}$ with $x \neq x'$. By extensionality take y such that $y \in x \leftrightarrow y \notin x'$. Wlog. we assume that $y \in x$ and $y \notin x'$. Then $2^{F(y)}$ is a summand in the binary expansion of $F(x)$, whereas it is not a summand in the binary expansion of $F(x')$. Therefore $F(x) \neq F(x')$.

For the surjectivity we prove by induction on $n \in \mathbb{N}$ that every $m \leqslant n$ is in ran(F). For $n = 0$ observe that

$$F(\emptyset) = \sum_{y \in \emptyset} 2^{F(y)} = 0$$

since the sum is empty. Now assume the claim for n. Let

$$n + 1 = \sum_{i < l} 2^{m_i}$$

with $m_0 < m_1 < \ldots < m_{l-1}$ be the binary expansion of $n+1$. Then $m_0 < m_1 < \ldots < m_{l-1} \leqslant n$ since $2^{n+1} > n + 1$. By the inductive assumption take (unique) sets $x_0, \ldots, x_{l-1}$ such that

$$F(x_0) = m_0, \ldots, F(x_{l-1}) = m_{l-1}.$$

Then

$$F(\{x_0, \ldots, x_{l-1}\}) = \sum_{i<l} 2^{F(x_i)} = \sum_{i<l} 2^{m_i} = n + 1.$$

□

Observe also that

$$y \in x \text{ iff } 2^{F(y)} \text{ is a summand of the binary expansion of } F(x).$$

If we define a binary relation ε on $\mathbb{N}$ by

$$m \varepsilon n \text{ iff } 2^m \text{ is a summand of the binary expansion of } n$$

then F is an isomorphism $F: (V, \in) \leftrightarrow (\mathbb{N}, \varepsilon)$.

The relation ε can be defined in Peano arithmetic.

$$m \varepsilon n \text{ iff } \exists p, q. p < 2^m \wedge 2^{m+1} | q \wedge n = p + 2^m + q.$$

So this reduces to the question whether exponentiation can be defined in Peano arithmetic - which has only addition and multiplication as basic operations, but not exponentiation. To define the term a^b in PA, we use the technique of the recursion theorem and represent the iterative computation sequence $a^0, a^1, ..., a^b$ by a pair of natural numbers.

Gödel used the *Chinese remainder theorem* for this purpose. The chinese mathematician Sun Tzu asked around the year 400 (in modern terminology):

> what is the smallest number n that when divided by 3 leaves a remainder of 2, when divided by 5 leaves a remainder of 3, and when divided by 7 leaves a remainder of 2?

(the answer is 23) The following theorem can be proved in PA.

Theorem 161. *Let $n_0, ..., n_{k-1}$ be positive integers that are pairwise relatively prime. Let $a_0, ..., a_{k-1}$ be an arbitrary sequence of integers. Then there is an integer x that solves the following system of congruences:*

$$\begin{array}{rcl} x & \equiv & a_0 \ (\operatorname{mod} n_0) \\ & \vdots & \\ x & \equiv & a_{k-1} \ (\operatorname{mod} n_{k-1}) \end{array}$$

If $a_0, ..., a_{k-1} < \min (n_0, ..., n_{k-1})$ then the sequence $a_0, ..., a_{k-1}$ is uniquely determined by x.

In the definition of a^b we shall apply the theorem with a relatively prime sequence

$$y \cdot 1 + 1, y \cdot 2 + 1, ..., y \cdot (b + 1) + 1$$

with an appropriately chosen y (e.g., where y is divisible by $b!$).

Now we can define:

$$\begin{array}{rcl} a \leqslant b & = & \exists n. a + n \equiv b \\ a < b & = & a + 1 \leqslant b \\ \operatorname{mod} (a, b, c) & = & \exists n. a = b \cdot n + c \wedge c < b \\ \operatorname{seq}(a, b, k, x) & = & \operatorname{mod} (a, b \cdot (k + 1) + 1, x) \\ a^b \equiv c & = & \exists x \exists y. \operatorname{seq}(x, y, 0, 1) \wedge \operatorname{seq}(x, y, b, c) \wedge \\ & & \wedge \forall k \forall z. k < c \wedge \operatorname{seq}(x, y, k, z) \rightarrow \operatorname{seq}(x, y, k + 1, z \cdot a) \end{array}$$

This proves:

Lemma 162. *The relation ε is definable by an arithmetic formula $\varepsilon(m, n)$.*

Let us consider the translations involved: For an $\in$-formula α define an arithmetical formula α^{ε} by replacing every subformula $u \in v$ by $\varepsilon(u, v)$. An arithmetic formula β can be interpreted in the structure $\mathbb{N}$ by restricting every quantifier in β to $\mathbb{N}$ and then replacing the symbols $+, \cdot, 0, 1$ by the homonymous abstraction terms. The result of these transformations is an $\in$-formula denoted by $(\beta)^{\mathbb{N}}$.

FST proves that $F: (V, \in) \leftrightarrow (\mathbb{N}, \varepsilon)$. So for every $\in$-formula $\chi(v_0, ..., v_{n-1})$:

$$\mathrm{FST} \vdash \chi(v_0, ..., v_{n-1}) \leftrightarrow (\chi^{\epsilon})^{\mathbb{N}}(F(v_0), ..., f(v_{n-1})).$$

We have proved earlier that $\mathbb{N}$ defined in ST is a model of PA . So for every arithmetic formula β:

$$\text{if } \mathrm{PA} \vdash \beta \text{ then } \mathrm{ST} \vdash \beta^{\mathbb{N}},$$

i.e., ST *interprets* the theory PA.

Conversely, PA interprets FST:

Theorem 163. *For every $\in$-formula α :*

$$\textit{if } \mathrm{FST} \vdash \alpha \textit{ then } \mathrm{PA} \vdash \alpha^{\varepsilon}.$$

This theorem can be proved by constructing the various sets required by the set existence axioms of FST in PA. All the sets required are intuitively finite and can be pieced together from their elements. The details of this are involved but straightforward.

Since the theories FST and PA interpret one another, they have the "same strength"; for many proof theoretic purposes they can be identified.

Theorem 164. PA *is consistent iff* ST *is consistent iff* FST *is consistent.*

Proof. Assume that PA is inconsistent so that $\mathrm{PA} \vdash \bot$. By the above, $\mathrm{ST} \vdash \bot^{\mathbb{N}} = \bot$, so that ST is inconsistent. Obviously, if ST is inconsistent then the stronger theory FST is inconsistent as well. Finally assume that FST is inconsistent and $\mathrm{FST} \vdash \bot$. By the previous theorem, $\mathrm{PA} \vdash \bot^{\varepsilon} = \bot$, so that PA is inconsistent. □

We can also work semantically with models:

Theorem 165. *From every model of* PA *one can construct a model of* FST*, and from every model of* ST *one can construct a model of* PA*.*

Proof. If $\mathcal{M} = (M, +^M, \cdot^M, 0^M, 1^M) \vDash \mathrm{PA}$ then $(M, \varepsilon^M) \vDash \mathrm{FST}$ where ε^M is defined in $\mathcal{M}$ by the formula $\varepsilon(x, y)$. Conversely, if $\mathcal{L} = (L, \in^L) \vDash \mathrm{ST}$ then $(\mathbb{N}^L, +^L, \cdot^L, 0^L, 1^L) \vDash \mathrm{PA}$ where $\mathbb{N}^L, +^L, \cdot^L, 0^L, 1^L$ are the interpretations of the abstraction terms $\mathbb{N}, +, \cdot, 0, 1$ in $\mathcal{L}$. □

Note that we have shown in both theorems that the addition of ¬Inf to ST does not raise the "danger" of inconsistency. Also the syntactic equiconsistency proof can also be Gödelized in ST.

Lemma 166. $\mathrm{ST} \vdash \mathrm{Con}(\ulcorner \mathrm{PA} \urcorner) \rightarrow \mathrm{Con}(\ulcorner \mathrm{FST} \urcorner)$ *and* $\mathrm{ST} \vdash \mathrm{Con}(\ulcorner \mathrm{ST} \urcorner) \rightarrow \mathrm{Con}(\ulcorner \mathrm{PA} \urcorner)$.

The original incompleteness theorems of Gödel were based on PA instead of ST, and we can now reprove them with the ε-interpretation.

Theorem 167. *If* PA *is consistent, then* PA *is incomplete.*

Proof. Assume that PA were complete. We show that FST is also complete, contradicting our first inompleteness theorem . Let α be an $\in$-sentence. Then α^{ε} is an arithmetic sentence.

Case 1. $\mathrm{PA} \vdash \alpha^{\varepsilon}$. Then

$$\mathrm{FST} \vdash (\alpha^{\varepsilon})^{\mathbb{N}} \wedge (\alpha \leftrightarrow (\alpha^{\varepsilon})^{\mathbb{N}}).$$

Therefore $\mathrm{FST} \vdash \alpha$.

Case 2. $\mathrm{PA} \nvdash \alpha^{\varepsilon}$. By the assumed completeness of PA we have $\mathrm{PA} \vdash \neg\alpha^{\varepsilon}$. Then

$$\mathrm{FST} \vdash (\neg\alpha^{\varepsilon})^{\mathbb{N}} \wedge (\alpha \leftrightarrow (\alpha^{\varepsilon})^{\mathbb{N}}).$$

Therefore $\mathrm{FST} \vdash \neg\alpha$. □

We have formalized consistency in ST. Since PA interprets FST we can translate consistency statements like $\mathrm{Con}(\ulcorner\mathrm{PA}\urcorner)$ into arithmetic statements like $(\mathrm{Con}(\ulcorner\mathrm{PA}\urcorner))^{\varepsilon}$ and evaluate them in PA. $(\mathrm{Con}(\ulcorner\mathrm{PA}\urcorner))^{\varepsilon}$ is an adequate arithmetic Gödelization of the consistency of PA. Then the second incompleteness theorem reads:

Theorem 168. *If* PA *is consistent then* $\mathrm{PA} \nvdash (\mathrm{Con}(\ulcorner\mathrm{PA}\urcorner))^{\varepsilon}$.

Proof. Assume for a contradiction that $\mathrm{PA} \vdash (\mathrm{Con}(\ulcorner\mathrm{PA}\urcorner))^{\varepsilon}$. Then $\mathrm{FST} \vdash ((\mathrm{Con}(\ulcorner\mathrm{PA}\urcorner))^{\varepsilon})^{\mathbb{N}}$. Since

$$\mathrm{FST} \vdash \mathrm{Con}(\ulcorner\mathrm{PA}\urcorner) \leftrightarrow ((\mathrm{Con}(\ulcorner\mathrm{PA}\urcorner))^{\varepsilon})^{\mathbb{N}},$$

we also have $\mathrm{FST} \vdash \mathrm{Con}(\ulcorner\mathrm{PA}\urcorner)$. Since $\mathrm{ST} \vdash \mathrm{Con}(\ulcorner\mathrm{PA}\urcorner) \rightarrow \mathrm{Con}(\ulcorner\mathrm{FST}\urcorner)$,

$$\mathrm{FST} \vdash \mathrm{Con}(\ulcorner\mathrm{FST}\urcorner).$$

By the 2nd incompleteness theorem for FST this implies that FST is inconsistent. But then PA is inconsistent. □

33 Fields

We formalize *fields* in the language of arithmetic together with a unary $-$-function

$$S_{\mathrm{Fd}} = \{+, \cdot, -, 0, 1\}$$

with the usual conventions for infix notation and bracket notation. The axiom system Φ_{Fd} of *field theory* consists of the following axioms:

- $\forall x \forall y \forall z\, (x+y)+z \equiv x+(y+z)$
- $\forall x \forall y \forall z\, (x \cdot y) \cdot z \equiv x \cdot (y \cdot z)$
- $\forall x \forall y x + y \equiv y + x$
- $\forall x \forall y x \cdot y \equiv y \cdot x$
- $\forall x x + 0 \equiv x$
- $\forall x x \cdot 1 \equiv x$
- $\forall x x + (-x) \equiv 0$
- $\forall x (\neg x \equiv 0 \rightarrow \exists y x \cdot y \equiv 1)$
- $0 \not\equiv 1$
- $\forall x \forall y \forall z\, x \cdot (y+z) \equiv (x \cdot y) + (x \cdot z)$

A *field* is an S_{Fd}-model satisfying Φ_{Fd}. Standard models of Φ_{Fd} are the fields $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ with addition and multiplication. Every field $\mathcal{F}$ contains at least the two elements $0^{\mathcal{F}}$ and $1^{\mathcal{F}}$. Therefore the minimal field $\mathbb{F}_2$ is given by $|\mathbb{F}_2| = \{0, 1\}$ and the operation tables

+	0	1
0	0	1
1	1	0

,

·	0	1
0	0	0
1	0	1

, and

−	
0	1
1	0

.

Note that this structure is similar to the boolean algebra of truth values.