Mathematical Logic

Helmut Schwichtenberg

Mathematisches Institut der Universität München
Wintersemester 2009/2010, Sommersemester 2010

Contents

CHAPTER 1

Logic

The main subject of Mathematical Logic is mathematical proof. In this introductory chapter we deal with the basics of formalizing such proofs and, via normalization, analysing their structure. The system we pick for the representation of proofs is Gentzen's natural deduction from (1935). Our reasons for this choice are twofold. First, as the name says this is a *natural* notion of formal proof, which means that the way proofs are represented corresponds very much to the way a careful mathematician writing out all details of an argument would go anyway. Second, formal proofs in natural deduction are closely related (via the so-called Curry-Howard correspondence) to terms in typed lambda calculus. This provides us not only with a compact notation for logical derivations (which otherwise tend to become somewhat unmanagable tree-like structures), but also opens up a route to applying (in part 3) the computational techniques which underpin lambda calculus.

Apart from classical logic we will also deal with more constructive logics: minimal and intuitionistic logic. This will reveal some interesting aspects of proofs, e.g., that it is possible and useful to distinguish beween existential proofs that actually construct witnessing objects, and others that don't.

An essential point for Mathematical Logic is to fix a formal language to be used. We take implication $\to$ and the universal quantifier $\forall$ as basic. Then the logic rules correspond precisely to lambda calculus. The additional connectives: the existential quantifier $\exists$, disjunction $\vee$ and conjunction $\wedge$, can then be added either as rules or axiom schemes. It is "natural" to treat them as rules, and that is what we do here.

An underlying theme of this chapter is to bring out the constructive content of logic, particularly in regard to the relationship between minimal and classical logic. For us the latter is most appropriately viewed as a subsystem of the former.

1.1. Natural Deduction

Rules come in pairs: we have an introduction and an elimination rule for each of the logical connectives. The resulting system is called *minimal logic*;

it was introduced by Kolmogorov (1932), Gentzen (1935) and Johansson (1937). Notice that no negation is yet present. If we go on and require *ex-falso-quodlibet* for the nullary propositional symbol $\bot$ ("falsum") we can embed *intuitionistic logic* with negation as $A \to \bot$. To embed classical logic, we need to go further and add as an axiom schema the principle of *indirect proof*, also called *stability* ($\forall_{\vec{x}}(\neg\neg R\vec{x} \to R\vec{x}$) for relation symbols R), but then it is appropriate to restrict to the language based on $\to$, $\forall$, $\bot$ and $\wedge$. The reason for this restriction is that we can neither prove $\neg\neg\exists_x A \to \exists_x A$ nor $\neg\neg(A \vee B) \to A \vee B$, for there are countermodels to both (the former is Markov's scheme). However, we can prove them for the classical existential quantifier and disjunction defined by $\neg\forall_x\neg A$ and $\neg A \to \neg B \to \bot$. Thus we need to make a distinction between two kinds of "exists" and two kinds of "or": the classical ones are "weak" and the non-classical ones "strong" since they have constructive content. In situations where both kinds occur together we must mark the distinction, and we shall do this by writing a tilde above the weak disjunction and existence symbols thus $\tilde{\vee}, \tilde{\exists}$. Of course, in a classical context this distinction does not arise and the tilde is not necessary.

1.1.1. Terms and formulas. Let a countably infinite set $\{ v_i \mid i \in \mathbb{N} \}$ of *variables* be given; they will be denoted by x, y, z. A first order language $\mathcal{L}$ then is determined by its *signature*, which is to mean the following.

(i) For every natural number $n \geq 0$ a (possible empty) set of n-ary *relation symbols* (or *predicate symbols*). 0-ary relation symbols are called *propositional symbols*. $\bot$ (read "falsum") is required as a fixed propositional symbol. The language will *not*, unless stated otherwise, contain $=$ as a primitive. Binary relation symbols can be marked as *infix*.

(ii) For every natural number $n \geq 0$ a (possible empty) set of n-ary *function symbols*. 0-ary function symbols are called *constants*. Binary function sysmbols can also be marked as infix.

We assume that all these sets of variables, relation and function symbols are disjoint. $\mathcal{L}$ is kept fixed and will only be mentioned when necessary.

Terms are inductively defined as follows.

(i) Every variable is a term.
(ii) Every constant is a term.
(iii) If $t_1, \ldots, t_n$ are terms and f is an n-ary function symbol with $n \geq 1$, then $f(t_1, \ldots, t_n)$ is a term. (If r, s are terms and $\circ$ is a binary function symbol, then $(r \circ s)$ is a term.)

From terms one constructs *prime formulas*, also called *atomic formulas* or just *atoms*: If $t_1, \ldots, t_n$ are terms and R is an n-ary relation symbol, then $R(t_1, \ldots, t_n)$ is a prime formula. (If r, s are terms and $\sim$ is a binary relation symbol, then $(r \sim s)$ is a prime formula.)

Formulas are inductively defined from prime formulas by

(i) Every prime formula is a formula.
(ii) If A and B are formulas, then so are $(A \to B)$ ("if A then B"), $(A \wedge B)$ ("A and B") and $(A \vee B)$ ("A or B").
(iii) If A is a formula and x is a variable, then $\forall_x A$ ("A holds for all x") and $\exists_x A$ ("there is an x such that A") are formulas.

Negation is defined by

$$\neg A := (A \to \bot).$$

NOTATION. In writing formulas we save on parentheses by assuming that $\forall, \exists, \neg$ bind more strongly than $\wedge, \vee$, and that in turn $\wedge, \vee$ bind more strongly than $\to, \leftrightarrow$ (where $A \leftrightarrow B$ abbreviates $(A \to B) \wedge (B \to A)$). Outermost parentheses can be dropped. Thus $A \wedge \neg B \to C$ is read as $((A \wedge (\neg B)) \to C)$. In the case of iterated implications we use the short notation

$$A_1 \to A_2 \to \cdots \to A_{n-1} \to A_n \text{ for } A_1 \to (A_2 \to \cdots \to (A_{n-1} \to A_n) \dots).$$

We also occasionally save on parentheses by writing for instance $Rxyz$, $Rt_0t_1t_2$ instead of $R(x, y, z)$, $R(t_0, t_1, t_2)$, where R is some predicate symbol. Similarly for a unary function symbol with a (typographically) simple argument, so fx for $f(x)$, etc. In this case no confusion will arise. But readability requires that we write in full $R(fx, gy, hz)$, instead of $Rfxgyhz$.

We shall often need to do induction on the height, denoted $|A|$, of formulas A. This is defined as follows: $|P| = 0$ for atoms P, $|A \circ B| = \max(|A|, |B|) + 1$ for binary operators $\circ$ (i.e., $\to, \wedge, \vee$) and $|\circ A| = |A| + 1$ for unary operators $\circ$ (i.e., $\forall_x$, $\exists_x$).

1.1.2. Substitution, free and bound variables. Expressions $\mathcal{E}, \mathcal{E}'$ which differ only in the names of bound variables will be regarded as identical. This is sometimes expressed by saying that $\mathcal{E}$ and $\mathcal{E}'$ are α-equivalent. In other words, we are only interested in expressions "modulo renaming of bound variables". There are methods of finding unique representatives for such expressions, for example the name-free terms of de Bruijn (1972). For the human reader such representations are less convenient, so we shall stick to the use of bound variables.

In the definition of "substitution of expression $\mathcal{E}'$ for variable x in expression $\mathcal{E}$", either one requires that *no* variable free in $\mathcal{E}'$ becomes bound by a variable-binding operator in $\mathcal{E}$, when the free occurrences of x are replaced by $\mathcal{E}'$ (also expressed by saying that there must be no "clashes of variables"), "$\mathcal{E}'$ *is free for* x *in* $\mathcal{E}$", or the substitution operation is taken to involve a systematic renaming operation for the bound variables, avoiding clashes. Having stated that we are only interested in expressions modulo

renaming bound variables, we can without loss of generality assume that substitution is always possible.

Also, it is never a real restriction to assume that distinct quantifier occurrences are followed by distinct variables, and that the sets of bound and free variables of a formula are disjoint.

NOTATION. "FV" is used for the (set of) free variables of an expression; so $\mathrm{FV}(r)$ is the set of variables free in the term r, $\mathrm{FV}(A)$ the set of variables free in formula A etc. A formula A is said to be *closed* if $\mathrm{FV}(A) = \emptyset$.

$\mathcal{E}[x := r]$ denotes the result of substituting the term r for the variable x in the expression $\mathcal{E}$. Similarly, $\mathcal{E}[\vec{x} := \vec{r}]$ is the result of *simultaneously* substituting the terms $\vec{r} = r_1, \ldots, r_n$ for the variables $\vec{x} = x_1, \ldots, x_n$, respectively.

In a given context we shall adopt the following convention. Once a formula has been introduced as $A(x)$, i.e., A with a designated variable x, we write $A(r)$ for $A[x := r]$, and similarly with more variables. □

1.1.3. Subformulas. Unless stated otherwise, the notion of *subformula* will be that defined by Gentzen.

DEFINITION. (Gentzen) subformulas of A are defined by

(a) A is a subformula of A;
(b) if $B \circ C$ is a subformula of A then so are B, C, for $\circ = \rightarrow, \wedge, \vee$;
(c) if $\forall_x B(x)$ or $\exists_x B(x)$ is a subformula of A, then so is $B(r)$.

DEFINITION. The notions of *positive, negative, strictly positive* subformula are defined in a similar style:

(a) A is a positive and a strictly positive subformula of itself;
(b) if $B \wedge C$ or $B \vee C$ is a positive (negative, strictly positive) subformula of A, then so are B, C;
(c) if $\forall_x B(x)$ or $\exists_x B(x)$ is a positive (negative, strictly positive) subformula of A, then so is $B(r)$;
(d) if $B \rightarrow C$ is a positive (negative) subformula of A, then B is a negative (positive) subformula of A, and C is a positive (negative) subformula of A;
(e) if $B \rightarrow C$ is a strictly positive subformula of A, then so is C.

A strictly positive subformula of A is also called a *strictly positive part (s.p.p.)* of A. Note that the set of subformulas of A is the union of the positive and negative subformulas of A.

EXAMPLE. $(P \rightarrow Q) \rightarrow R \wedge \forall_x S(x)$ has as s.p.p.'s the whole formula, $R \wedge \forall_x S(x)$, R, $\forall_x S(x)$, $S(r)$. The positive subformulas are the s.p.p.'s and in addition P; the negative subformulas are $P \rightarrow Q$, Q.

1.1.4. Examples of derivations. To motivate the rules for natural deduction, let us start with informal proofs of some simple logical facts.

$$(A \to B \to C) \to (A \to B) \to A \to C.$$

Informal proof. Assume $A \to B \to C$. To show: $(A \to B) \to A \to C$. So assume $A \to B$. To show: $A \to C$. So finally assume A. To show: C. Using the third assumption twice we have $B \to C$ by the first assumption, and B by the second assumption. From $B \to C$ and B we then obtain C. Then $A \to C$, cancelling the assumption on A; $(A \to B) \to A \to C$ cancelling the second assumption; and the result follows by cancelling the first assumption. □

$$\forall_x(A \to B) \to A \to \forall_x B, \quad \text{if } x \notin \mathrm{FV}(A).$$

Informal proof. Assume $\forall_x(A \to B)$. To show: $A \to \forall_x B$. So assume A. To show: $\forall_x B$. Let x be arbitrary; note that we have not made any assumptions on x. To show: B. We have $A \to B$ by the first assumption. Hence also B by the second assumption. Hence $\forall_x B$. Hence $A \to \forall_x B$, cancelling the second assumption. Hence the result, cancelling the first assumption. □

A characteristic feature of these proofs is that assumptions are introduced and eliminated again. At any point in time during the proof the free or "open" assumptions are known, but as the proof progresses, free assumptions may become cancelled or "closed" because of the implies-introduction rule.

We reserve the word *proof* for the informal level; a formal representation of a proof will be called a *derivation*.

An intuitive way to communicate derivations is to view them as labelled trees each node of which denotes a rule application. The labels of the inner nodes are the formulas derived as conclusions at those points, and the labels of the leaves are formulas or terms. The labels of the nodes immediately above a node k are the *premises* of the rule application. At the root of the tree we have the conclusion (or end formula) of the whole derivation. In natural deduction systems one works with *assumptions* at leaves of the tree; they can be either *open* or *closed* (cancelled). Any of these assumptions carries a *marker*. As markers we use *assumption variables* denoted $u, v, w, u_0, u_1, \ldots$. The variables of the language previously introduced will now often be called *object variables*, to distinguish them from assumption variables. If at a node below an assumption the dependency on this assumption is removed (it becomes closed) we record this by writing down the assumption variable. Since the same assumption may be used more than once (this was the case in the first example above), the assumption marked with u (written $u\colon A$) may appear many times. Of course we insist that distinct assumption formulas must have distinct markers. An inner node of

the tree is understood as the result of passing from premises to the conclusion of a given rule. The label of the node then contains, in addition to the conclusion, also the name of the rule. In some cases the rule binds or closes or cancels an assumption variable u (and hence removes the dependency of all assumptions $u\colon A$ thus marked). An application of the $\forall$-introduction rule similarly binds an object variable x (and hence removes the dependency on x). In both cases the bound assumption or object variable is added to the label of the node.

DEFINITION. A formula A is called *derivable* (in *minimal logic*), written $\vdash A$, if there is a derivation of A (without free assumptions) using the natural deduction rules. A formula B is called derivable from assumptions $A_1, \dots, A_n$, if there is a derivation of B with free assumptions among $A_1, \dots, A_n$. Let Γ be a (finite or infinite) set of formulas. We write $\Gamma \vdash B$ if the formula B is derivable from finitely many assumptions $A_1, \dots, A_n \in \Gamma$.

We now formulate the rules of natural deduction.

1.1.5. Introduction and elimination rules for $\to$ and $\forall$. First we have an assumption rule, allowing to write down an arbitrary formula A together with a marker u:

$$u\colon A \qquad \text{assumption.}$$

The other rules of natural deduction split into introduction rules (I-rules for short) and elimination rules (E-rules) for the logical connectives which, for the time being, are just $\to$ and $\forall$. For implication $\to$ there is an introduction rule $\to^+$ and an elimination rule $\to^-$ also called *modus ponens*. The left premise $A \to B$ in $\to^-$ is called the *major* (or *main*) premise, and the right premise A the *minor* (or *side*) premise. Note that with an application of the $\to^+$-rule *all* assumptions above it marked with $u\colon A$ are cancelled (which is denoted by putting square brackets around these assumptions), and the u then gets written alongside. There may of course be other uncancelled assumptions $v\colon A$ of the same formula A, which may get cancelled at a later stage.

$$\frac{\begin{array}{c}[u\colon A]\\ \vdots\ M\\ B\end{array}}{A \to B}\to^+ u \qquad \frac{\begin{array}{cc}\vdots\ M & \vdots\ N\\ A \to B & A\end{array}}{B}\to^-$$

For the universal quantifier $\forall$ there is an introduction rule $\forall^+$ (again marked, but now with the bound variable x) and an elimination rule $\forall^-$ whose right premise is the term r to be substituted. The rule $\forall^+ x$ with conclusion $\forall_x A$ is subject to the following *(Eigen-)variable condition*: the derivation M of

the premise A should not contain any open assumption having x as a free variable.

$$\frac{\begin{array}{c}|\,M\\ A\end{array}}{\forall_x A}\,\forall^+x \qquad \frac{\begin{array}{cc}|\,M & \\ \forall_x A(x) & r\end{array}}{A(r)}\,\forall^-$$

We now give derivations of the two example formulas treated informally above. Since in many cases the rule used is determined by the conclusion, we suppress in such cases the name of the rule.

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{u\colon A\to B\to C \qquad w\colon A}{B\to C} \qquad \dfrac{v\colon A\to B \qquad w\colon A}{B}}{C}\,{\to^+}w}{A\to C}\,{\to^+}v}{(A\to B)\to A\to C}\,{\to^+}u}{(A\to B\to C)\to(A\to B)\to A\to C}$$

$$\dfrac{\dfrac{\dfrac{\dfrac{u\colon \forall_x(A\to B) \qquad x}{A\to B} \qquad v\colon A}{B}\,\forall^+x}{\forall_x B}\,{\to^+}v}{A\to\forall_x B}\,{\to^+}u}{\forall_x(A\to B)\to A\to\forall_x B}$$

Note that the variable condition is satisfied: x is not free in A (and also not free in $\forall_x(A\to B)$).

1.1.6. Properties of negation. Recall that negation is defined by $\neg A := (A\to\bot)$. The following can easily be derived.

$$A\to\neg\neg A,$$
$$\neg\neg\neg A\to\neg A.$$

However, $\neg\neg A\to A$ is in general *not* derivable (without stability – we will come back to this later on).

Lemma. *The following are derivable.*

$$(A\to B)\to\neg B\to\neg A,$$
$$\neg(A\to B)\to\neg B,$$

$$\neg\neg(A\to B)\to\neg\neg A\to\neg\neg B,$$
$$(\bot\to B)\to(\neg\neg A\to\neg\neg B)\to\neg\neg(A\to B),$$

$$\neg\neg\forall_x A\to\forall_x\neg\neg A.$$

Derivations are left as an exercise.

1.1.7. Introduction and elimination rules for disjunction $\vee$, conjunction $\wedge$ and existence $\exists$. For disjunction the introduction and elimination rules are

$$\frac{\begin{array}{c}|\,M\\ A\end{array}}{A\vee B}\vee_0^+ \qquad \frac{\begin{array}{c}|\,M\\ B\end{array}}{A\vee B}\vee_1^+ \qquad \frac{\begin{array}{ccc}&[u\colon A]&[v\colon B]\\ |\,M&|\,N&|\,K\\ A\vee B&C&C\end{array}}{C}\vee^- u,v$$

For conjunction we have

$$\frac{\begin{array}{cc}|\,M&|\,N\\ A&B\end{array}}{A\wedge B}\wedge^+ \qquad \frac{\begin{array}{ccc}&[u\colon A]&[v\colon B]\\ |\,M&&|\,N\\ A\wedge B&&C\end{array}}{C}\wedge^- u,v$$

and for the existential quantifier

$$\frac{\begin{array}{cc}&|\,M\\ r&A(r)\end{array}}{\exists_x A(x)}\exists^+ \qquad \frac{\begin{array}{ccc}&[u\colon A]\\ |\,M&|\,N\\ \exists_x A&B\end{array}}{B}\exists^- x,u\ \text{(var.cond.)}$$

Similar to $\forall^+ x$ the rule $\exists^- x,u$ is subject to an *(Eigen-)variable condition*: in the derivation N the variable x (i) should not occur free in the formula of any open assumption other than $u\colon A$, and (ii) should not occur free in B.

Again, in each of the elimination rules $\vee^-$, $\wedge^-$ and $\exists^-$ the left premise is called *major* (or *main*) premise, and the right premise is called the *minor* (or *side*) premise.

It is easy to see that for each of the connectives $\vee$, $\wedge$, $\exists$ the rules and the following axioms are equivalent over minimal logic; this is left as an exercise. For disjunction the introduction and elimination axioms are

$$\begin{aligned}&\vee_0^+\colon A\to A\vee B,\\ &\vee_1^+\colon B\to A\vee B,\\ &\vee^-\colon A\vee B\to(A\to C)\to(B\to C)\to C.\end{aligned}$$

For conjunction we have

$$\wedge^+\colon A\to B\to A\wedge B,\qquad \wedge^-\colon A\wedge B\to(A\to B\to C)\to C$$

and for the existential quantifier

$$\exists^+\colon A\to\exists_x A,\qquad \exists^-\colon\exists_x A\to\forall_x(A\to B)\to B\quad(x\notin \mathrm{FV}(B)).$$

REMARK. All these axioms can be seen as special cases of a general schema, that of an *inductively defined predicate*, which is defined by some introduction rules and one elimination rule. Later we will study this kind of definition in full generality.

We collect some easy facts about derivability; $B \leftarrow A$ means $A \to B$.

LEMMA. *The following are derivable.*

$$(A \wedge B \to C) \leftrightarrow (A \to B \to C),$$
$$(A \to B \wedge C) \leftrightarrow (A \to B) \wedge (A \to C),$$

$$(A \vee B \to C) \leftrightarrow (A \to C) \wedge (B \to C),$$
$$(A \to B \vee C) \leftarrow (A \to B) \vee (A \to C),$$

$$(\forall_x A \to B) \leftarrow \exists_x (A \to B) \quad \textit{if } x \notin \mathrm{FV}(B),$$
$$(A \to \forall_x B) \leftrightarrow \forall_x (A \to B) \quad \textit{if } x \notin \mathrm{FV}(A),$$

$$(\exists_x A \to B) \leftrightarrow \forall_x (A \to B) \quad \textit{if } x \notin \mathrm{FV}(B),$$
$$(A \to \exists_x B) \leftarrow \exists_x (A \to B) \quad \textit{if } x \notin \mathrm{FV}(A).$$

PROOF. A derivation of the final formula is

$$\dfrac{\dfrac{u\colon \exists_x(A \to B) \qquad \dfrac{\dfrac{x \qquad \dfrac{w\colon A \to B \qquad v\colon A}{B}}{\exists_x B}}{\exists_x B}\,\exists^- x, w}{A \to \exists_x B}\,{\to}^+ v}{\exists_x(A \to B) \to A \to \exists_x B}\,{\to}^+ u$$

The variable condition for $\exists^-$ is satisfied since the variable x (i) is not free in the formula A of the open assumption $v\colon A$, and (ii) is not free in $\exists_x B$. The rest of the proof is left as an exercise. □

As already mentioned, we distinguish between two kinds of "exists" and two kinds of "or": the "weak" or classical ones and the "strong" or non-classical ones, with constructive content. In the present context both kinds occur together and hence we must mark the distinction; we shall do this by writing a tilde above the weak disjunction and existence symbols thus

$$A \mathbin{\tilde{\vee}} B := \neg A \to \neg B \to \bot, \qquad \tilde{\exists}_x A := \neg \forall_x \neg A.$$

One can show easily that these weak variants of disjunction and the existential quantifier are no stronger than the proper ones (in fact, they are weaker):

$$A \vee B \to A \mathbin{\tilde{\vee}} B, \qquad \exists_x A \to \tilde{\exists}_x A.$$

This can be seen easily by putting $C := \bot$ in $\vee^-$ and $B := \bot$ in $\exists^-$.

REMARK. Since $\tilde{\exists}_x \tilde{\exists}_y A$ unfolds into a rather awkward formula we extend the $\tilde{\exists}$-terminology to lists of variables:

$$\tilde{\exists}_{x_1,\dots,x_n} A := \forall_{x_1,\dots,x_n}(A \to \bot) \to \bot.$$

Moreover let

$$\tilde{\exists}_{x_1,\dots,x_n}(A_1 \mathbin{\tilde{\wedge}} \dots \mathbin{\tilde{\wedge}} A_m) := \forall_{x_1,\dots,x_n}(A_1 \to \dots \to A_m \to \bot) \to \bot.$$

This allows to stay in the $\to, \forall$ part of the language. Notice that $\tilde{\wedge}$ only makes sense in this context, i.e., in connection with $\tilde{\exists}$.

1.1.8. Intuitionistic and classical derivability. In the definition of derivability in 1.1.4 falsity $\bot$ plays no role. We may change this and require *ex-falso-quodlibet* axioms, of the form

$$\forall_{\vec{x}}(\bot \to R\vec{x}\,)$$

with R a relation symbol distinct from $\bot$. Let Efq denote the set of all such axioms. A formula A is called *intuitionistically derivable*, written $\vdash_i A$, if $\mathrm{Efq} \vdash A$. We write $\Gamma \vdash_i B$ for $\Gamma \cup \mathrm{Efq} \vdash B$.

We may even go further and require *stability* axioms, of the form

$$\forall_{\vec{x}}(\neg\neg R\vec{x} \to R\vec{x}\,)$$

with R again a relation symbol distinct from $\bot$. Let Stab denote the set of all these axioms. A formula A is called *classically derivable*, written $\vdash_c A$, if $\mathrm{Stab} \vdash A$. We write $\Gamma \vdash_c B$ for $\Gamma \cup \mathrm{Stab} \vdash B$.

It is easy to see that intuitionistically (i.e., from Efq) we can derive $\bot \to A$ for an *arbitrary* formula A, using the introduction rules for the connectives. A similar generalization of the stability axioms is only possible for formulas in the language not involving $\vee, \exists$. However, it is still possible to use the substitutes $\tilde{\vee}$ and $\tilde{\exists}$.

THEOREM (Stability, or principle of indirect proof).

(a) $\vdash (\neg\neg A \to A) \to (\neg\neg B \to B) \to \neg\neg(A \wedge B) \to A \wedge B$.
(b) $\vdash (\neg\neg B \to B) \to \neg\neg(A \to B) \to A \to B$.
(c) $\vdash (\neg\neg A \to A) \to \neg\neg\forall_x A \to A$.
(d) $\vdash_c \neg\neg A \to A$ *for every formula* A *without* $\vee, \exists$.

PROOF. (a) is left as an exercise. (b). For simplicity, in the derivation to be constructed we leave out applications of $\to^+$ at the end.

$$\dfrac{u\colon \neg\neg B \to B \qquad \dfrac{\dfrac{v\colon \neg\neg(A\to B) \qquad \dfrac{\dfrac{u_2\colon A\to B \qquad w\colon A}{B}}{\bot}\to^+ u_2}{\neg(A\to B)}}{\bot}\to^+ u_1}{\neg\neg B}}{B}$$

(c).

$$\dfrac{u\colon \neg\neg A \to A \qquad \dfrac{\dfrac{v\colon \neg\neg\forall_x A \qquad \dfrac{\dfrac{u_2\colon \forall_x A \qquad x}{A}}{\bot}\to^+ u_2}{\neg\forall_x A}}{\bot}\to^+ u_1}{\neg\neg A}}{A}$$

(d). Induction on A. The case $R\vec{t}$ with R distinct from $\bot$ is given by Stab. In the case $\bot$ the desired derivation is

$$\dfrac{v\colon (\bot\to\bot)\to\bot \qquad \dfrac{u\colon \bot}{\bot\to\bot}\to^+ u}{\bot}$$

In the cases $A\wedge B$, $A\to B$ and $\forall_x A$ use (a), (b) and (c), respectively. □

Using stability we can prove some well-known facts about the interaction of weak disjunction and the weak existential quantifier with implication. We first prove a more refined claim, stating to what extent we need to go beyond minimal logic.

LEMMA. *The following are derivable.*

$$(\tilde\exists_x A\to B)\to\forall_x(A\to B) \quad \text{if } x\notin \mathrm{FV}(B), \tag{1.1}$$

$$(\neg\neg B\to B)\to \quad \forall_x(A\to B)\to\tilde\exists_x A\to B \quad \text{if } x\notin \mathrm{FV}(B), \tag{1.2}$$

$$(\bot\to B[x{:=}c])\to(A\to\tilde\exists_x B)\to\tilde\exists_x(A\to B) \quad \text{if } x\notin \mathrm{FV}(A), \tag{1.3}$$

$$\tilde\exists_x(A\to B)\to A\to\tilde\exists_x B \quad \text{if } x\notin \mathrm{FV}(A). \tag{1.4}$$

The last two items can also be seen as simplifying a weakly existentially quantified implication whose premise does not contain the quantified variable. In case the conclusion does not contain the quantified variable we have

$$(\neg\neg B\to B)\to \quad \tilde\exists_x(A\to B)\to\forall_x A\to B \quad \text{if } x\notin \mathrm{FV}(B), \tag{1.5}$$

(1.6) $\quad \forall_x(\neg\neg A \to A) \to (\forall_x A \to B) \to \tilde{\exists}_x(A \to B) \quad \textit{if } x \notin \mathrm{FV}(B).$

Proof. (1.1)

$$\dfrac{\tilde{\exists}_x A \to B \qquad \dfrac{\dfrac{u_1 \colon \forall_x \neg A \qquad x}{\neg A} \qquad A}{\bot}}{B} \to^+ u_1$$

(1.2)

$$\dfrac{\neg\neg B \to B \qquad \dfrac{\dfrac{\neg\forall_x \neg A \qquad \dfrac{\dfrac{\dfrac{\forall_x(A \to B) \qquad x}{A \to B} \qquad u_1 \colon A}{B}}{\bot}}{\forall_x \neg A}}{\bot}}{\neg\neg B} \to^+ u_1 \to^+ u_2$$

(1.3) Writing B_0 for $B[x{:=}c]$ we have

$$\dfrac{\dfrac{\forall_x \neg(A \to B) \quad c}{\neg(A \to B_0)} \qquad \dfrac{\dfrac{\bot \to B_0 \qquad \dfrac{\dfrac{A \to \tilde{\exists}_x B \quad u_2 \colon A}{\tilde{\exists}_x B} \qquad \dfrac{\dfrac{\dfrac{\forall_x \neg(A \to B) \quad x}{\neg(A \to B)} \qquad \dfrac{u_1 \colon B}{A \to B}}{\bot}}{\neg B}}{\forall_x \neg B}}{\bot}}{B_0}}{A \to B_0} \to^+ u_2}{\bot}$$

(1.4)

$$\dfrac{\tilde{\exists}_x(A \to B) \qquad \dfrac{\dfrac{\dfrac{\forall_x \neg B \qquad x}{\neg B} \qquad \dfrac{u_1 \colon A \to B \qquad A}{B}}{\bot}}{\neg(A \to B)}}{\forall_x \neg(A \to B)} \to^+ u_1$$

(1.5)

$$\dfrac{\neg\neg B \to B \qquad \dfrac{\dfrac{\tilde{\exists}_x(A \to B) \qquad \dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\forall_x A \qquad x}{A} \qquad u_1 \colon A \to B}{B}}{\bot}\to^+ u_1}{\neg(A \to B)}}{\forall_x \neg(A \to B)}}{\bot}\to^+ u_2}{\neg\neg B}}{B}$$

(1.6) We derive $\forall_x(\bot \to A) \to (\forall_x A \to B) \to \forall_x \neg(A \to B) \to \neg\neg A$. Writing Ax, Ay for $A(x), A(y)$ we have

$$\dfrac{\dfrac{\forall_x \neg(Ax \to B) \qquad x}{\neg(Ax \to B)} \qquad \dfrac{\forall_x Ax \to B \qquad \dfrac{\dfrac{\dfrac{\forall_y(\bot \to Ay) \qquad y}{\bot \to Ay} \qquad \dfrac{u_1 \colon \neg Ax \qquad u_2 \colon Ax}{\bot}}{Ay}}{\forall_y Ay}}{\dfrac{B}{Ax \to B}}\to^+ u_2}{\bot}\to^+ u_1}{\neg\neg Ax}$$

Using this derivation M we obtain

$$\dfrac{\dfrac{\forall_x \neg(Ax \to B) \qquad x}{\neg(Ax \to B)} \qquad \dfrac{\forall_x Ax \to B \qquad \dfrac{\dfrac{\dfrac{\forall_x(\neg\neg Ax \to Ax) \qquad x}{\neg\neg Ax \to Ax} \qquad \begin{array}{l}\mid M\\ \neg\neg Ax\end{array}}{Ax}}{\forall_x Ax}}{\dfrac{B}{Ax \to B}}}{\bot}$$

Since clearly $\vdash (\neg\neg A \to A) \to \bot \to A$ the claim follows. □

Remark. An immediate consequence of (1.6) is the classical derivability of the "drinker formula" $\tilde{\exists}_x(Px \to \forall_x Px)$, to be read "in every non-empty bar there is a person such that, if this person drinks, then everybody drinks". To see this let $A := Px$ and $B := \forall_x Px$ in (1.6).

Corollary.

$\vdash_c (\tilde{\exists}_x A \to B) \leftrightarrow \forall_x(A \to B)$ *if* $x \notin \mathrm{FV}(B)$ *and* B *without* $\vee, \exists$,

$\vdash_i (A \to \tilde{\exists}_x B) \leftrightarrow \tilde{\exists}_x(A \to B)$ *if* $x \notin \mathrm{FV}(A)$,

$\vdash_c \tilde{\exists}_x(A \to B) \leftrightarrow (\forall_x A \to B)$ *if* $x \notin \mathrm{FV}(B)$ *and* A, B *without* $\vee, \exists$.

There is a similar lemma on weak disjunction:

LEMMA. *The following are derivable.*

$$\begin{aligned}
&(A \mathbin{\tilde\vee} B \to C) \to (A \to C) \wedge (B \to C),\\
(\neg\neg C \to C) \to\ &(A \to C) \to (B \to C) \to A \mathbin{\tilde\vee} B \to C,\\
(\bot \to B) \to\ &(A \to B \mathbin{\tilde\vee} C) \to (A \to B) \mathbin{\tilde\vee} (A \to C),\\
&(A \to B) \mathbin{\tilde\vee} (A \to C) \to A \to B \mathbin{\tilde\vee} C,\\
(\neg\neg C \to C) \to\ &(A \to C) \mathbin{\tilde\vee} (B \to C) \to A \to B \to C,\\
(\bot \to C) \to\ &(A \to B \to C) \to (A \to C) \mathbin{\tilde\vee} (B \to C).
\end{aligned}$$

PROOF. The derivation of the final formula is

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{A \to B \to C \quad u_1 : A}{B \to C} \qquad u_2 : B}{C}}{A \to C}\to^+ u_1 \qquad \neg(A \to C)}{\bot}}{\bot \to C} \qquad \bot}{C}}{B \to C}\to^+ u_2 \qquad \neg(B \to C)}{\bot}$$

The other derivations are similar to the ones above, if one views $\tilde\exists$ as an infinitary version of $\tilde\vee$. □

COROLLARY.

$$\begin{aligned}
&\vdash_c (A \mathbin{\tilde\vee} B \to C) \leftrightarrow (A \to C) \wedge (B \to C) \quad \textit{for } C \textit{ without } \vee, \exists,\\
&\vdash_i (A \to B \mathbin{\tilde\vee} C) \leftrightarrow (A \to B) \mathbin{\tilde\vee} (A \to C),\\
&\vdash_c (A \to C) \mathbin{\tilde\vee} (B \to C) \leftrightarrow (A \to B \to C) \quad \textit{for } C \textit{ without } \vee, \exists.
\end{aligned}$$

REMARK. It is easy to see that weak disjunction and the weak existential quantifier satisfy the same axioms as the strong variants, if one restricts the conclusion of the elimination axioms to formulas without $\vee, \exists$. In fact, we have

$$\begin{aligned}
&\vdash A \to A \mathbin{\tilde\vee} B, \quad \vdash B \to A \mathbin{\tilde\vee} B,\\
&\vdash_c A \mathbin{\tilde\vee} B \to (A \to C) \to (B \to C) \to C \quad (C \text{ without } \vee, \exists),\\
&\vdash A \to \tilde\exists_x A,\\
&\vdash_c \tilde\exists_x A \to \forall_x (A \to B) \to B \quad (x \notin \mathrm{FV}(B),\ B \text{ without } \vee, \exists).
\end{aligned}$$

The derivations are left as exercises.

1.1.9. Gödel-Gentzen translation. Classical derivability $\Gamma \vdash_c B$ was defined in 1.1.8 by $\Gamma \cup \mathrm{Stab} \vdash B$. This embedding of classical logic into minimal logic can be expressed in a somewhat different and very explicit form, namely as a syntactic translation $A \mapsto A^g$ of formulas such that A is derivable in classical logic if and only if its translation A^g is derivable in minimal logic.

Definition (Gödel-Gentzen translation A^g).

$$\begin{aligned}
(R\vec{t})^g &:= \neg\neg R\vec{t} \quad \text{for } R \text{ distinct from } \bot,\\
\bot^g &:= \bot,\\
(A \vee B)^g &:= A^g \mathbin{\tilde{\vee}} B^g,\\
(\exists_x A)^g &:= \tilde{\exists}_x A^g,\\
(A \circ B)^g &:= A^g \circ B^g \quad \text{for } \circ = \to, \wedge,\\
(\forall_x A)^g &:= \forall_x A^g.
\end{aligned}$$

Lemma. $\vdash \neg\neg A^g \to A^g$.

Proof. Induction on A.

Case $R\vec{t}$ with R distinct from $\bot$. We must show $\neg\neg\neg\neg R\vec{t} \to \neg\neg R\vec{t}$, which is a special case of $\vdash \neg\neg\neg B \to \neg B$.

Case $\bot$. Use $\vdash \neg\neg\bot \to \bot$.

Case $A \vee B$. We must show $\vdash \neg\neg(A^g \mathbin{\tilde{\vee}} B^g) \to A^g \mathbin{\tilde{\vee}} B^g$, which is a special case of $\vdash \neg\neg(\neg C \to \neg D \to \bot) \to \neg C \to \neg D \to \bot$:

$$\dfrac{\neg\neg(\neg C \to \neg D \to \bot) \qquad \dfrac{\dfrac{\dfrac{u_1 \colon \neg C \to \neg D \to \bot \qquad \neg C}{\neg D \to \bot} \qquad \neg D}{\bot}}{\neg(\neg C \to \neg D \to \bot)} \to^+ u_1}{\bot}$$

Case $\exists_x A$. In this case we must show $\vdash \neg\neg\tilde{\exists}_x A^g \to \tilde{\exists}_x A^g$, but this is a special case of $\vdash \neg\neg\neg B \to \neg B$, because $\tilde{\exists}_x A^g$ is the negation $\neg\forall_x \neg A^g$.

Case $A \wedge B$. We must show $\vdash \neg\neg(A^g \wedge B^g) \to A^g \wedge B^g$. By induction hypothesis $\vdash \neg\neg A^g \to A^g$ and $\vdash \neg\neg B^g \to B^g$. Now use part (a) of the stability theorem in 1.1.8.

The cases $A \to B$ and $\forall_x A$ are similar, using parts (b) and (c) of the stability theorem instead. □

Theorem. (a) $\Gamma \vdash_c A$ *implies* $\Gamma^g \vdash A^g$.
(b) $\Gamma^g \vdash A^g$ *implies* $\Gamma \vdash_c A$ *for* Γ, A *without* $\vee, \exists$.

Proof. (a). We use induction on $\Gamma \vdash_c A$. For a stability axiom $\forall_{\vec{x}}(\neg\neg R\vec{x} \to R\vec{x})$ we must derive $\forall_{\vec{x}}(\neg\neg\neg\neg R\vec{x} \to \neg\neg R\vec{x})$, which is easy

(as above). For the rules $\to^+$, $\to^-$, $\forall^+$, $\forall^-$, $\wedge^+$ and $\wedge^-$ the claim follows immediately from the induction hypothesis, using the same rule again. This works because the Gödel-Gentzen translation acts as a homomorphism for these connectives. For the rules $\vee_i^+$, $\vee^-$, $\exists^+$ and $\exists^-$ the claim follows from the induction hypothesis and the remark at the end of 1.1.8. For example, in case $\exists^-$ the induction hypothesis gives

$$\begin{array}{c} \\ \mid M \\ \tilde{\exists}_x A^g \end{array} \quad \text{and} \quad \begin{array}{c} u\colon A^g \\ \mid N \\ B^g \end{array}$$

with $x \notin \mathrm{FV}(B^g)$. Now use $\vdash (\neg\neg B^g \to B^g) \to \tilde{\exists}_x A^g \to \forall_x(A^g \to B^g) \to B^g$. Its premise $\neg\neg B^g \to B^g$ is derivable by the lemma above.

(b). First note that $\vdash_c (B \leftrightarrow B^g)$ if B is without $\vee, \exists$. Now assume that Γ, A are without $\vee, \exists$. From $\Gamma^g \vdash A^g$ we obtain $\Gamma \vdash_c A$ as follows. We argue informally. Assume Γ. Then Γ^g by the note, hence A^g because of $\Gamma^g \vdash A^g$, hence A again by the note. □

1.2. Normalization

A derivation in normal form does not make "detours", or more precisely, it cannot occur that an elimination rule immediately follows an introduction rule. We will use "conversions" to remove such "local maxima" of complexity, thus reducing any given derivation to normal form. However, there is a difficulty when we consider an elimination rule for $\vee$, $\wedge$ or $\exists$. An introduced formula may be used as a minor premise of an application of $\vee^-$, $\wedge^-$ or $\exists^-$, then stay the same throughout a sequence of applications of these rules, being eliminated at the end. This also constitutes a local maximum, which we should like to eliminate; *permutative conversions* are designed for exactly this situation. In a permutative conversion we permute an E-rule upwards over the minor premises of $\vee^-$, $\wedge^-$ or $\exists^-$.

We analyse the shape of derivations in normal form, and then prove the (crucial) subformula property, which says that every formula in a normal derivation is a subformula of the end-formula or else of an assumption.

It will be convenient to represent derivations as typed "derivation terms", where the derived formula is seen as the "type" of the term (and displayed as a superscript). This representation is known under the name *Curry-Howard correspondence*. We give an inductive definition of such derivation terms for the $\to, \forall$-rules in table 1 where for clarity we have written the corresponding derivations to the left. In table 2 this is extended to the rules for $\vee$, $\wedge$ and $\exists$.

1.2.1. Conversions. A conversion eliminates a detour in a derivation, i.e., an elimination immediately following an introduction. We now spell

Derivation	Term
$u\colon A$	u^A
$\dfrac{\begin{array}{c}[u\colon A]\\ \mid M\\ B\end{array}}{A \to B} \to^+ u$	$(\lambda_{u^A} M^B)^{A\to B}$
$\dfrac{\begin{array}{cc}\mid M & \mid N\\ A \to B & A\end{array}}{B} \to^-$	$(M^{A\to B} N^A)^B$
$\dfrac{\begin{array}{c}\mid M\\ A\end{array}}{\forall_x A} \forall^+ x$ (with var.cond.)	$(\lambda_x M^A)^{\forall_x A}$ (with var.cond.)
$\dfrac{\begin{array}{cc}\mid M & \\ \forall_x A(x) & r\end{array}}{A(r)} \forall^-$	$(M^{\forall_x A(x)} r)^{A(r)}$

TABLE 1. Derivation terms for $\to$ and $\forall$

out in detail which conversions we shall allow. This is done for derivations written in tree notation and also as derivation terms.

$\to$-conversion.

$$\dfrac{\dfrac{\begin{array}{c}[u\colon A]\\ \mid M\\ B\end{array}}{A \to B} \to^+ u \qquad \begin{array}{c}\mid N\\ A\end{array}}{B} \to^- \quad \mapsto \quad \begin{array}{c}\mid N\\ A\\ \mid M\\ B\end{array}$$

or written as derivation terms $(\lambda_u M(u^A)^B)^{A\to B} N^A \mapsto M(N^A)^B$. The reader familiar with λ-calculus should note that this is nothing other than β-conversion.

Derivation	Term
$\dfrac{\begin{array}{c}\mid M\\ A\end{array}}{A \vee B}\vee_0^+ \qquad \dfrac{\begin{array}{c}\mid M\\ B\end{array}}{A \vee B}\vee_1^+$	$(\vee_{0,B}^+ M^A)^{A\vee B} \quad (\vee_{1,A}^+ M^B)^{A\vee B}$
$\dfrac{\begin{array}{ccc} & [u\colon A] & [v\colon B]\\ \mid M & \mid N & \mid K\\ A \vee B & C & C\end{array}}{C}\vee^- u, v$	$(M^{A\vee B}(u^A.N^C, v^B.K^C))^C$
$\dfrac{\begin{array}{cc}\mid M & \mid N\\ A & B\end{array}}{A \wedge B}\wedge^+$	$\langle M^A, N^B\rangle^{A\wedge B}$
$\dfrac{\begin{array}{cc} & [u\colon A] \quad [v\colon B]\\ \mid M & \mid N\\ A \wedge B & C\end{array}}{C}\wedge^- u, v$	$(M^{A\wedge B}(u^A, v^B.N^C))^C$
$\dfrac{\begin{array}{cc} & \mid M\\ r & A(r)\end{array}}{\exists_x A(x)}\exists^+$	$(\exists_{x,A}^+ r M^{A(r)})^{\exists_x A(x)}$
$\dfrac{\begin{array}{cc} & [u\colon A]\\ \mid M & \mid N\\ \exists_x A & B\end{array}}{B}\exists^- x, u$ (var.cond.)	$(M^{\exists_x A}(u^A.N^B))^B$ (var.cond.)

TABLE 2. Derivation terms for $\vee$, $\wedge$ and $\exists$

$\forall$-*conversion.*

$$\dfrac{\dfrac{\begin{array}{c}\mid M\\ A(x)\end{array}}{\forall_x A(x)}\,\forall^+x \qquad r}{A(r)}\,\forall^- \quad\mapsto\quad \begin{array}{c}\mid M'\\ A(r)\end{array}$$

or written as derivation terms $(\lambda_x M(x)^{A(x)})^{\forall_x A(x)} r \mapsto M(r)$.

$\vee$-*conversion.*

$$\dfrac{\dfrac{\begin{array}{c}\mid M\\ A\end{array}}{A\vee B}\,\vee_0^+ \qquad \begin{array}{c}[u\colon A]\\ \mid N\\ C\end{array} \qquad \begin{array}{c}[v\colon B]\\ \mid K\\ C\end{array}}{C}\,\vee^- u, v \quad\mapsto\quad \begin{array}{c}\mid M\\ A\\ \mid N\\ C\end{array}$$

or as derivation terms $(\vee_{0,B}^+ M^A)^{A\vee B}(u^A.N(u)^C, v^B.K(v)^C) \mapsto N(M^A)^C$, and similarly for $\vee_1^+$ with K instead of N.

$\wedge$-*conversion.*

$$\dfrac{\dfrac{\begin{array}{c}\mid M\\ A\end{array} \qquad \begin{array}{c}\mid N\\ B\end{array}}{A\wedge B}\,\wedge^+ \qquad \begin{array}{c}[u\colon A]\\ [v\colon B]\\ \mid K\\ C\end{array}}{C}\,\wedge^- u, v \quad\mapsto\quad \begin{array}{c}\mid M\\ A\\ \mid N\\ B\\ \mid K\\ C\end{array}$$

or $\langle M^A, N^B\rangle^{A\wedge B}(u^A, v^B.K(u,v)^C) \mapsto K(M^A, N^B)^C$.

$\exists$-*conversion.*

$$\dfrac{\dfrac{r \qquad \begin{array}{c}\mid M\\ A(r)\end{array}}{\exists_x A(x)}\,\exists^+ \qquad \begin{array}{c}[u\colon A(x)]\\ \mid N\\ B\end{array}}{B}\,\exists^- x, u \quad\mapsto\quad \begin{array}{c}\mid M\\ A(r)\\ \mid N'\\ B\end{array}$$

or $(\exists_{x,A}^+ r M^{A(r)})^{\exists_x A(x)}(u^{A(x)}.N(x,u)^B) \mapsto N(r, M^{A(r)})^B$.

1.2.2. Permutative conversions.

$\vee$-*permutative conversion.*

$$\dfrac{\begin{array}{c}\mid M\\ A\vee B\end{array} \qquad \begin{array}{c}\mid N\\ C\end{array} \qquad \begin{array}{c}\mid K\\ C\end{array}}{\dfrac{C}{D}\,\text{E-rule} \qquad \begin{array}{c}\mid L\\ C'\end{array}}\,\text{E-rule} \quad\mapsto$$

$$\dfrac{\begin{array}{c}\mid M\\ A\vee B\end{array} \qquad \dfrac{\begin{array}{c}\mid N\\ C\end{array} \qquad \begin{array}{c}\mid L\\ C'\end{array}}{D}\,\text{E-rule} \qquad \dfrac{\begin{array}{c}\mid K\\ C\end{array} \qquad \begin{array}{c}\mid L\\ C'\end{array}}{D}\,\text{E-rule}}{D}$$

or with for instance $\to^-$ as E-rule $(M^{A\vee B}(u^A.N^{C\to D}, v^B.K^{C\to D}))^{C\to D}L^C \mapsto (M^{A\vee B}(u^A.(N^{C\to D}L^C)^D, v^B.(K^{C\to D}L^C)^D))^D$.

$\wedge$-*permutative conversion.*

$$\dfrac{\dfrac{\begin{array}{c}|\,M\\ A\wedge B\end{array}\qquad\begin{array}{c}|\,N\\ C\end{array}}{C}\qquad\begin{array}{c}|\,K\\ C'\end{array}}{D}\text{ E-rule}\quad\mapsto$$

$$\dfrac{\begin{array}{c}|\,M\\ A\wedge B\end{array}\qquad\dfrac{\begin{array}{c}|\,N\\ C\end{array}\qquad\begin{array}{c}|\,K\\ C'\end{array}}{D}\text{ E-rule}}{D}$$

or $(M^{A\wedge B}(u^A, v^B.N^{C\to D}))^{C\to D}K^C \mapsto (M^{A\wedge B}(u^A, v^B.(N^{C\to D}K^C)^D))^D$.

$\exists$-*permutative conversion.*

$$\dfrac{\dfrac{\begin{array}{c}|\,M\\ \exists_x A\end{array}\qquad\begin{array}{c}|\,N\\ B\end{array}}{B}\qquad\begin{array}{c}|\,K\\ C\end{array}}{D}\text{ E-rule}\quad\mapsto$$

$$\dfrac{\begin{array}{c}|\,M\\ \exists_x A\end{array}\qquad\dfrac{\begin{array}{c}|\,N\\ B\end{array}\qquad\begin{array}{c}|\,K\\ C\end{array}}{D}\text{ E-rule}}{D}$$

or $(M^{\exists_x A}(u^A.N^{C\to D}))^{C\to D}K^C \mapsto (M^{\exists_x A}(u^A.(N^{C\to D}K^C)^D))^D$.

1.2.3. Simplification conversions. These are somewhat trivial conversions, which remove unnecessary applications of the elimination rules for $\vee$, $\wedge$ and $\exists$. For $\vee$ we have

$$\dfrac{\begin{array}{c}|\,M\\ A\vee B\end{array}\qquad\begin{array}{c}[u\colon A]\\ |\,N\\ C\end{array}\qquad\begin{array}{c}[v\colon B]\\ |\,K\\ C\end{array}}{C}\vee^- u, v\quad\mapsto\quad\begin{array}{c}|\,N\\ C\end{array}$$

if $u\colon A$ is not free in N, or $(M^{A\vee B}(u^A.N^C, v^B.K^C))^C \mapsto N^C$; similar for the second component. For $\wedge$ there is the conversion

$$\dfrac{\begin{array}{c}|\,M\\ A\wedge B\end{array}\qquad\begin{array}{c}[u\colon A]\\ [v\colon B]\\ |\,N\\ C\end{array}}{C}\wedge^- u, v\quad\mapsto\quad\begin{array}{c}|\,N\\ C\end{array}$$

if neither $u\colon A$ nor $v\colon B$ is free in N, or $(M^{A\wedge B}(u^A, v^B.N^C))^C \mapsto N^C$. For $\exists$ the simplification conversion is

$$\dfrac{\begin{array}{cc} & [u\colon A] \\ \mid M & \mid N \\ \exists_x A & B \end{array}}{B}\,\exists^- x,u \quad \mapsto \quad \begin{array}{c} \mid N \\ B \end{array}$$

if again $u\colon A$ is not free in N, or $(M^{\exists_x A}(u^A.N^B))^B \mapsto N^B$.

1.2.4. Strong normalization. We now show that no matter in which order we apply the conversion rules, they will always terminate and produce a derivation in "normal form", where no further conversions can be applied.

We shall write derivation terms without formula super- or subscripts. For instance, we write $\exists^+$ instead of $\exists^+_{x,A}$. Hence we consider derivation terms M, N, K now of the forms

$$\begin{aligned} &u \mid \lambda_v M \mid \lambda_y M \mid \vee_0^+ M \mid \vee_1^+ M \mid \langle M, N\rangle \mid \exists^+ rM \mid \\ &\quad MN \mid Mr \mid M(v_0.N_0, v_1.N_1) \mid M(v, w.N) \mid M(v.N) \end{aligned}$$

where, in these expressions, the variables v, y, v_0, v_1, w are bound.

To simplify the technicalities, we restrict our treatment to the rules for $\to$ and $\exists$. The argument easily extends to the full set of rules. Hence we consider

$$u \mid \lambda_v M \mid \exists^+ rM \mid MN \mid M(v.N).$$

The strategy for strong normalization is set out below, but a word about notation is crucial here. Whenever we write an applicative term as $M\vec{N} := MN_1 \dots N_k$ the convention is that bracketing to the left operates. That is, $M\vec{N} = (\dots(MN_1)\dots N_k)$.

We reserve the letters E, F, G for *eliminations*, i.e., expressions of the form $(v.N)$, and R, S, T for both terms and eliminations. Using this notation we obtain a second (and clearly equivalent) inductive definition of terms:

$$\begin{aligned} &u\vec{M} \mid u\vec{M}E \mid \lambda_v M \mid \exists^+ rM \mid \\ &(\lambda_v M)N\vec{R} \mid \exists^+ rM(v.N)\vec{R} \mid u\vec{M}ER\vec{S}. \end{aligned}$$

Here only the final three forms are not normal: $(\lambda_v M)N\vec{R}$ and $\exists^+ rM(v.N)\vec{R}$ both are *β-redexes*, and $u\vec{M}ER\vec{S}$ is a *permutative redex*. The conversion rules for them are

$$\begin{array}{lll} (\lambda_v M(v))N & \mapsto_\beta M(N) & \beta_\to\text{-conversion}, \\ \exists^+_{x,A} rM(v.N(x,v)) & \mapsto_\beta N(r, M) & \beta_\exists\text{-conversion}, \\ M(v.N)R & \mapsto_\pi M(v.NR) & \text{permutative conversion}. \end{array}$$

In addition we also allow

$$M(v.N) \mapsto_\sigma N \quad \text{if } v\colon A \text{ is not free in } N.$$

The latter is called a *simplification conversion*, and $M(v.N)$ a *simplification redex*.

The *closure* of these conversions is defined by

(a) If $M \mapsto_\xi M'$ for $\xi = \beta, \pi, \sigma$, then $M \to M'$.
(b) If $M \to M'$, then $MR \to M'R$, $NM \to NM'$, $N(v.M) \to N(v.M')$, $\lambda_v M \to \lambda_v M'$, $\exists^+ rM \to \exists^+ rM'$ (*inner reductions*).

So $M \to N$ means that M *reduces in one step to* N, i.e., N is obtained from M by replacement of (an occurrence of) a redex M' of M by a conversum M'' of M', i.e., by a single conversion. The relation $\to^+$ ("*properly reduces to*") is the transitive closure of $\to$, and $\to^*$ ("*reduces to*") is the reflexive transitive closure of $\to$. A term M is *in normal form* (or simply *normal*) if M does not contain a redex. M *has a normal form* if there is a normal N such that $M \to^* N$. A *reduction sequence* is a (finite or infinite) sequence $M_0, M_1, M_2 \ldots$ such that $M_i \to M_{i+1}$, for all i.

We inductively define a set SN of derivation terms. In doing so we take care that for a given M there is exactly one rule applicable to generate $M \in \mathrm{SN}$. This will be crucial to make the later proofs work.

Definition (SN).

$$\frac{\vec{M} \in \mathrm{SN}}{u\vec{M} \in \mathrm{SN}}\,(\mathrm{Var}_0) \qquad \frac{M \in \mathrm{SN}}{\lambda_v M \in \mathrm{SN}}\,(\lambda) \qquad \frac{M \in \mathrm{SN}}{\exists^+ rM \in \mathrm{SN}}\,(\exists)$$

$$\frac{\vec{M}, N \in \mathrm{SN}}{u\vec{M}(v.N) \in \mathrm{SN}}\,(\mathrm{Var}) \qquad \frac{u\vec{M}(v.NR)\vec{S} \in \mathrm{SN}}{u\vec{M}(v.N)R\vec{S} \in \mathrm{SN}}\,(\mathrm{Var}_\pi)$$

$$\frac{M(N)\vec{R} \in \mathrm{SN} \qquad N \in \mathrm{SN}}{(\lambda_v M(v))N\vec{R} \in \mathrm{SN}}\,(\beta_\to)$$

$$\frac{N(r, M)\vec{R} \in \mathrm{SN} \qquad M \in \mathrm{SN}}{\exists^+_{x,A} rM(v.N(x, v))\vec{R} \in \mathrm{SN}}\,(\beta_\exists)$$

In (Var_π) we require that x (from $\exists_x A$) and v are not free in R.

It is easy to see that SN is closed under substitution for object variables: if $M(x) \in \mathrm{SN}$, then $M(r) \in \mathrm{SN}$. The proof is by induction on $M \in \mathrm{SN}$, applying the induction hypothesis first to the premise(es) and then reapplying the same rule.

We write $M{\downarrow}$ to mean that M is strongly normalizing, i.e., that every reduction sequence starting from M terminates. By analysing the possible reduction steps we now show that the set $\{\, M \mid M{\downarrow} \,\}$ has the closure properties of the definition of SN above, and hence $\mathrm{SN} \subseteq \{\, M \mid M{\downarrow} \,\}$.

LEMMA. *Every term in* SN *is strongly normalizing.*

PROOF. We distinguish cases according to the generation rule of SN applied last. The following rules deserve special attention.

Case (Var_π). We prove, as an auxiliary lemma, that

$$u\vec{M}(v.NR)\vec{S}{\downarrow} \text{ implies } u\vec{M}(v.N)R\vec{S}{\downarrow}.$$

As a typical case consider

$$u\vec{M}(v.N(v'.N'))TS{\downarrow} \text{ implies } u\vec{M}(v.N)(v'.N')TS{\downarrow}.$$

However, it is easy to see that any infinite reduction sequence of the latter would give rise to an infinite reduction sequence of the former.

Case ($\beta_\to$). We show that $M(N)\vec{R}{\downarrow}$ and $N{\downarrow}$ imply $(\lambda_v M(v))N\vec{R}{\downarrow}$. This is done by induction on $N{\downarrow}$, with a side induction on $M(N)\vec{R}{\downarrow}$. We need to consider all possible reducts of $(\lambda_v M(v))N\vec{R}$. In case of an outer β-reduction use the assumption. If N is reduced, use the induction hypothesis. Reductions in M and in $\vec{R}$ as well as permutative reductions within $\vec{R}$ are taken care of by the side induction hypothesis.

Case ($\beta_\exists$). We show that

$$N(r, M)\vec{R}{\downarrow} \text{ and } M{\downarrow} \text{ together imply } \exists^+ rM(v.N(x, v))\vec{R}{\downarrow}.$$

This is done by a threefold induction: first on $M{\downarrow}$, second on $N(r, M)\vec{R}{\downarrow}$ and third on the length of $\vec{R}$. We need to consider all possible reducts of $\exists^+ rM(v.N(x, v))\vec{R}$. In case of an outer β-reduction it must reduce to $N(r, M)\vec{R}$, hence the result by assumption. If M is reduced, use the first induction hypothesis. Reductions in $N(x, v)$ and in $\vec{R}$ as well as permutative reductions within $\vec{R}$ are taken care of by the second induction hypothesis. The only remaining case is when $\vec{R} = S\vec{S}$ and $(v.N(x, v))$ is permuted with S, to yield $\exists^+ rM(v.N(x, v)S)\vec{S}$, in which case the third induction hypothesis applies. □

For later use we prove a slightly generalized form of the rule (Var_π):

PROPOSITION. *If* $M(v.NR)\vec{S} \in \mathrm{SN}$, *then* $M(v.N)R\vec{S} \in \mathrm{SN}$.

PROOF. Induction on the generation of $M(v.NR)\vec{S} \in \mathrm{SN}$. We distinguish cases according to the form of M.

Case $u\vec{T}(v.NR)\vec{S} \in \mathrm{SN}$. If $\vec{T} = \vec{M}$ (i.e., $\vec{T}$ consists of derivation terms only), use (Var_π). Else we have $u\vec{M}(v'.N')\vec{R}(v.NR)\vec{S} \in \mathrm{SN}$. This must be

generated by repeated applications of (Var_π) from $u\vec{M}(v'.N'\vec{R}(v.NR)\vec{S}) \in \mathrm{SN}$, and finally by (Var) from $\vec{M} \in \mathrm{SN}$ and $N'\vec{R}(v.NR)\vec{S} \in \mathrm{SN}$. The induction hypothesis for the latter fact yields $N'\vec{R}(v.N)R\vec{S} \in \mathrm{SN}$, hence $u\vec{M}(v'.N'\vec{R}(v.N)R\vec{S}) \in \mathrm{SN}$ by (Var) and finally $u\vec{M}(v'.N')\vec{R}(v.N)R\vec{S} \in \mathrm{SN}$ by (Var_π).

Case $\exists^+ rM\vec{T}(v.N(x,v)R)\vec{S} \in \mathrm{SN}$. Similar, with $(\beta_\exists)$ instead of (Var_π). In detail: If $\vec{T}$ is empty, by $(\beta_\exists)$ this came from $N(r,M)R\vec{S} \in \mathrm{SN}$ and $M \in \mathrm{SN}$, hence $\exists^+ rM(v.N(x,v))R\vec{S} \in \mathrm{SN}$ again by $(\beta_\exists)$. Otherwise we have $\exists^+ rM(v'.N'(x',v'))\vec{T}(v.NR)\vec{S} \in \mathrm{SN}$. This must be generated by $(\beta_\exists)$ from $N'(r,M)\vec{T}(v.NR)\vec{S} \in \mathrm{SN}$. The induction hypothesis yields $N'(r,M)\vec{T}(v.N)R\vec{S} \in \mathrm{SN}$, hence $\exists^+ rM(v'.N'(x,v'))\vec{T}(v.N)R\vec{S} \in \mathrm{SN}$ by $(\beta_\exists)$.

Case $(\lambda_v M(v))N'\vec{R}(w.NR)\vec{S} \in \mathrm{SN}$. By $(\beta_\to)$ this came from $N' \in \mathrm{SN}$ and $M(N')\vec{R}(w.NR)\vec{S} \in \mathrm{SN}$. But the induction hypothesis yields $M(N')\vec{R}(w.N)R\vec{S} \in \mathrm{SN}$, hence $(\lambda_v M(v))N'\vec{R}(w.N)R\vec{S} \in \mathrm{SN}$ by $(\beta_\to)$. □

We show, finally, that *every* term is in SN and hence is strongly normalizing. Given the definition of SN we only have to show that SN is closed under $\to^-$ and $\exists^-$. But in order to prove this we must prove simultaneously the closure of SN under substitution.

THEOREM (Properties of SN). *For all formulas A,*

(a) *for all $M \in \mathrm{SN}$, if M proves $A = A_0 \to A_1$ and $N \in \mathrm{SN}$, then $MN \in \mathrm{SN}$,*
(b) *for all $M \in \mathrm{SN}$, if M proves $A = \exists_x B$ and $N \in \mathrm{SN}$, then $M(v.N) \in \mathrm{SN}$,*
(c) *for all $M(v) \in \mathrm{SN}$, if $N^A \in \mathrm{SN}$, then $M(N) \in \mathrm{SN}$.*

PROOF. Induction on $|A|$. We prove (a) and (b) before (c), and hence have (a) and (b) available for the proof of (c). More formally, by induction on A we simultaneously prove that (a) holds, that (b) holds and that (a), (b) together imply (c).

(a). By side induction on $M \in \mathrm{SN}$. Let $M \in \mathrm{SN}$ and assume that M proves $A = A_0 \to A_1$ and $N \in \mathrm{SN}$. We distinguish cases according to how $M \in \mathrm{SN}$ was generated. For (Var_0), (Var_π), $(\beta_\to)$ and $(\beta_\exists)$ use the same rule again.

Case $u\vec{M}(v.N') \in \mathrm{SN}$ by (Var) from $\vec{M}, N' \in \mathrm{SN}$. Then $N'N \in \mathrm{SN}$ by side induction hypothesis for N', hence $u\vec{M}(v.N'N) \in \mathrm{SN}$ by (Var), hence $u\vec{M}(v.N')N \in \mathrm{SN}$ by (Var_π).

Case $(\lambda_v M(v))^{A_0 \to A_1} \in \mathrm{SN}$ by (λ) from $M(v) \in \mathrm{SN}$. Use $(\beta_\to)$; for this we need to know $M(N) \in \mathrm{SN}$. But this follows from induction hypothesis (c) for $M(v)$, since N derives A_0.

(b). By side induction on $M \in \mathrm{SN}$. Let $M \in \mathrm{SN}$ and assume that M proves $A = \exists_x B$ and $N \in \mathrm{SN}$. The goal is $M(v.N) \in \mathrm{SN}$. We distinguish

cases according to how $M \in \mathrm{SN}$ was generated. For (Var_π), $(\beta_\to)$ and $(\beta_\exists)$ use the same rule again.

Case $u\vec{M} \in \mathrm{SN}$ by (Var_0) from $\vec{M} \in \mathrm{SN}$. Use (Var).

Case $(\exists^+ rM)^{\exists_x A} \in \mathrm{SN}$ by $(\exists)$ from $M \in \mathrm{SN}$. We must show that $\exists^+ rM(v.N(x,v)) \in \mathrm{SN}$. Use $(\beta_\exists)$; for this we need to know $N(r,M) \in \mathrm{SN}$. But this follows from induction hypothesis (c) for $N(r,v)$ (which is in SN by the remark above), since M derives $A(r)$.

Case $u\vec{M}(v'.N') \in \mathrm{SN}$ by (Var) from $\vec{M}, N' \in \mathrm{SN}$. Then $N'(v.N) \in \mathrm{SN}$ by side induction hypothesis for N', hence $u\vec{M}(v.N'(v.N)) \in \mathrm{SN}$ by (Var) and therefore $u\vec{M}(v.N')(v.N) \in \mathrm{SN}$ by (Var_π).

(c). By side induction on $M(v) \in \mathrm{SN}$. Let $N^A \in \mathrm{SN}$; the goal is $M(N) \in \mathrm{SN}$. We distinguish cases according to how $M(v) \in \mathrm{SN}$ was generated. For (λ), $(\exists)$, $(\beta_\to)$ and $(\beta_\exists)$ use the same rule again, after applying the induction hypothesis to the premise(es).

Case $u\vec{M}(v) \in \mathrm{SN}$ by (Var_0) from $\vec{M}(v) \in \mathrm{SN}$. Then $\vec{M}(N) \in \mathrm{SN}$ by side induction hypothesis (c). If $u \neq v$, use (Var_0) again. If $u = v$, we must show $N\vec{M}(N) \in \mathrm{SN}$. Note that N proves A; hence the claim follows from $\vec{M}(N) \in \mathrm{SN}$ by (a) with $M = N$.

Case $u\vec{M}(v)(v'.N'(v)) \in \mathrm{SN}$ by (Var) from $\vec{M}(v), N'(v) \in \mathrm{SN}$. If $u \neq v$, use (Var) again. If $u = v$, we must show $N\vec{M}(N)(v'.N'(N)) \in \mathrm{SN}$. Note that N proves A; hence in case $\vec{M}(v)$ is empty the claim follows from (b) with $M = N$, and otherwise from (a), (b) and the induction hypothesis.

Case $u\vec{M}(v)(v'.N'(v))R(v)\vec{S}(v) \in \mathrm{SN}$ has been obtained by (Var_π) from $u\vec{M}(v)(v'.N'(v)R(v))\vec{S}(v) \in \mathrm{SN}$. If $u \neq v$, use (Var_π) again. If $u = v$, from the side induction hypothesis we obtain $N\vec{M}(N)(v'.N'(N)R(N))\vec{S}(N) \in \mathrm{SN}$. Now use the proposition above with $M := N\vec{M}(N)$. □

Corollary. *Every derivation term is in* SN *and therefore strongly normalizing.*

Proof. Induction on the (first) inductive definition of derivation terms. In cases u, $\lambda_v M$ and $\exists^+ rM$ the claim follows from the definition of SN, and in cases MN and $M(v.N)$ from parts (a), (b) of the previous theorem. □

1.2.5. On disjunction. Incorporating the full set of rules adds no other technical complications but merely increases the length. For the energetic reader, however, we include here the details necessary for disjunction. The conjunction case is entirely straightforward.

We have additional β-conversions

$$\vee_i^+ M(v_0.N_0, v_1.N_1) \mapsto_\beta N_i[v_i := M] \quad \beta_{\vee_i}\text{-conversion.}$$

The definition of SN needs to be extended by

$$\frac{M \in \mathrm{SN}}{\vee_i^+ M \in \mathrm{SN}}\,(\vee_i)$$

$$\frac{\vec{M}, N_0, N_1 \in \mathrm{SN}}{u\vec{M}(v_0.N_0, v_1.N_1) \in \mathrm{SN}}\,(\mathrm{Var}_\vee) \qquad \frac{u\vec{M}(v_0.N_0R, v_1.N_1R)\vec{S} \in \mathrm{SN}}{u\vec{M}(v_0.N_0, v_1.N_1)R\vec{S} \in \mathrm{SN}}\,(\mathrm{Var}_{\vee,\pi})$$

$$\frac{N_i[v_i := M]\vec{R} \in \mathrm{SN} \qquad N_{1-i}\vec{R} \in \mathrm{SN} \qquad M \in \mathrm{SN}}{\vee_i^+ M(v_0.N_0, v_1.N_1)\vec{R} \in \mathrm{SN}}\,(\beta_{\vee_i})$$

The former rules (Var), (Var_π) should then be renamed into ($\mathrm{Var}_\exists$), ($\mathrm{Var}_{\exists,\pi}$).

The lemma above stating that every term in SN is strongly normalizable needs to be extended by an additional clause:

Case $(\beta_{\vee_i})$. We show that $N_i[v_i := M]\vec{R}\!\downarrow$, $N_{1-i}\vec{R}\!\downarrow$ and $M\!\downarrow$ together imply $\vee_i^+ M(v_0.N_0, v_1.N_1)\vec{R}\!\downarrow$. This is done by a fourfold induction: first on $M\!\downarrow$, second on $N_i[v_i := M]\vec{R}\!\downarrow$, $N_{1-i}\vec{R}\!\downarrow$, third on $N_{1-i}\vec{R}\!\downarrow$ and fourth on the length of $\vec{R}$. We need to consider all possible reducts of $\vee_i^+ M(v_0.N_0, v_1.N_1)\vec{R}$. In case of an outer β-reduction use the assumption. If M is reduced, use the first induction hypothesis. Reductions in N_i and in $\vec{R}$ as well as permutative reductions within $\vec{R}$ are taken care of by the second induction hypothesis. Reductions in N_{1-i} are taken care of by the third induction hypothesis. The only remaining case is when $\vec{R} = S\vec{S}$ and $(v_0.N_0, v_1.N_1)$ is permuted with S, to yield $(v_0.N_0S, v_1.N_1S)$. Apply the fourth induction hypothesis, since $(N_iS)[v := M]\vec{S} = N_i[v := M]S\vec{S}$.

Finally the theorem above stating properties of SN needs an additional clause:

(b') for all $M \in \mathrm{SN}$, if M proves $A = A_0 \vee A_1$ and $N_0, N_1 \in \mathrm{SN}$, then $M(v_0.N_0, v_1.N_1) \in \mathrm{SN}$.

Proof. The new clause is proved by induction on $M \in \mathrm{SN}$. Let $M \in \mathrm{SN}$ and assume that M proves $A = A_0 \vee A_1$ and $N_0, N_1 \in \mathrm{SN}$. The goal is $M(v_0.N_0, v_1.N_1) \in \mathrm{SN}$. We distinguish cases according to how $M \in \mathrm{SN}$ was generated. For ($\mathrm{Var}_{\exists,\pi}$), ($\mathrm{Var}_{\vee,\pi}$), $(\beta_\to)$, $(\beta_\exists)$ and $(\beta_{\vee_i})$ use the same rule again.

Case $u\vec{M} \in \mathrm{SN}$ by (Var_0) from $\vec{M} \in \mathrm{SN}$. Use ($\mathrm{Var}_\vee$).

Case $(\vee_i^+ M)^{A_0 \vee A_1} \in \mathrm{SN}$ by $(\vee_i)$ from $M \in \mathrm{SN}$. Use $(\beta_{\vee_i})$; for this we need to know $N_i[v_i := M] \in \mathrm{SN}$ and $N_{1-i} \in \mathrm{SN}$. The latter is assumed, and the former follows from main induction hypothesis (with N_i) for the substitution clause of the theorem, since M derives A_i.

Case $u\vec{M}(v'.N') \in \mathrm{SN}$ by $(\mathrm{Var}_\exists)$ from $\vec{M}, N' \in \mathrm{SN}$. For brevity let $E := (v_0.N_0, v_1.N_1)$. Then $N'E \in \mathrm{SN}$ by side induction hypothesis for N', so $u\vec{M}(v'.N'E) \in \mathrm{SN}$ by $(\mathrm{Var}_\exists)$ and therefore $u\vec{M}(v'.N')E \in \mathrm{SN}$ by $(\mathrm{Var}_{\exists,\pi})$.

Case $u\vec{M}(v'_0.N'_0, v'_1.N'_1) \in \mathrm{SN}$ by $(\mathrm{Var}_\vee)$ from $\vec{M}, N'_0, N'_1 \in \mathrm{SN}$. Let $E := (v_0.N_0, v_1.N_1)$. Then $N'_iE \in \mathrm{SN}$ by side induction hypothesis for N'_i, so $u\vec{M}(v'_0.N'_0E, v'_1.N'_1E) \in \mathrm{SN}$ by $(\mathrm{Var}_\vee)$ and therefore $u\vec{M}(v'_0.N'_0, v'_1.N'_1)E \in \mathrm{SN}$ by $(\mathrm{Var}_{\vee,\pi})$.

Clause (c) now needs additional cases, e.g.,

Case $u\vec{M}(v_0.N_0, v_1.N_1) \in \mathrm{SN}$ by $(\mathrm{Var}_\vee)$ from $\vec{M}, N_0, N_1 \in \mathrm{SN}$. If $u \neq v$, use $(\mathrm{Var}_\vee)$. If $u = v$, we show $N\vec{M}[v := N](v_0.N_0[v := N], v_1.N_1[v := N]) \in \mathrm{SN}$. Note that N proves A; hence in case $\vec{M}$ empty the claim follows from (b), and otherwise from (a) and the induction hypothesis. □

1.2.6. The structure of normal derivations. To analyse normal derivations, it will be useful to introduce the notions of a *segment* and of a *track* in a proof tree, which make sense for non-normal derivations as well.

Definition. A *segment* (of *length* n) in a derivation M is a sequence $A_0, \ldots, A_n$ of occurrences of the same formula A such that

(a) for $0 \leq i < n$, A_i is a minor premise of an application of $\vee^-$, $\wedge^-$ or $\exists^-$, with conclusion A_{i+1};
(b) A_n is not a minor premise of $\vee^-$, $\wedge^-$ or $\exists^-$.
(c) A_0 is not the conclusion of $\vee^-$, $\wedge^-$ or $\exists^-$.

Notice that a formula occurrence (f.o.) which is neither a minor premise nor the conclusion of an application of $\vee^-$, $\wedge^-$ or $\exists^-$ always constitutes a segment of length 1. A segment is *maximal* or a *cut (segment)* if A_n is the major premise of an E-rule, and either $n > 0$, or $n = 0$ and $A_0 = A_n$ is the conclusion of an I-rule.

We use σ, σ' for segments. σ is called a *subformula* of σ' if the formula A in σ is a subformula of B in σ'.

The notion of a track is designed to retain the subformula property in case one passes through the major premise of an application of a $\vee^-, \wedge^-, \exists^-$-rule. In a track, when arriving at an A_i which is the major premise of an application of such a rule, we take for A_{i+1} a hypothesis discharged by this rule.

Definition. A *track* of a derivation M is a sequence of f.o.'s $A_0, \ldots, A_n$ such that

(a) A_0 is a top f.o. in M not discharged by an application of an $\vee^-, \wedge^-, \exists^-$-rule;

(b) A_i for $i < n$ is not the minor premise of an instance of $\to^-$, and *either*
  (i) A_i is not the major premise of an instance of a $\vee^-, \wedge^-, \exists^-$-rule and A_{i+1} is directly below A_i, *or*
  (ii) A_i is the major premise of an instance of a $\vee^-, \wedge^-, \exists^-$-rule and A_{i+1} is an assumption discharged by this instance;

(c) A_n is *either*
  (i) the minor premise of an instance of $\to^-$, *or*
  (ii) the end formula of M, *or*
  (iii) the major premise of an instance of a $\vee^-, \wedge^-, \exists^-$-rule in case there are no assumptions discharged by this instance.

LEMMA. *In a derivation each formula occurrence belongs to some track.*

PROOF. By induction on derivations. For example, suppose a derivation K ends with an $\exists^-$-application:

$$\dfrac{\begin{array}{cc} & [u\colon A] \\ \mid M & \mid N \\ \exists_x A & B \end{array}}{B}\,\exists^- x, u$$

B in N belongs to a track π (induction hypothesis); either this does not start in $u\colon A$, and then π, B is a track in K which ends in the end formula; or π starts in $u\colon A$, and then there is a track π' in M (induction hypothesis) such that π', π, B is a track in K ending in the end formula. The other cases are left to the reader. □

DEFINITION. A *track of order* 0, or *main track*, in a derivation is a track ending either in the end formula of the whole derivation or in the major premise of an application of a $\vee^-$, $\wedge^-$ or $\exists^-$-rule, provided there are no assumption variables discharged by the application. A *track of order* $n+1$ is a track ending in the minor premise of an $\to^-$-application, with major premise belonging to a track of order n.

A *main branch* of a derivation is a branch π (i.e., a linearly ordered subtree) in the proof tree such that π passes only through premises of I-rules and *major premises* of E-rules, and π begins at a top node and ends in the end formula.

Since by simplification conversions we have removed every application of an $\vee^-$, $\wedge^-$ or $\exists^-$-rule that discharges no assumption variables, each track of order 0 in a normal derivation is a track ending in the end formula of the whole derivation. Note also that if we search for a main branch going upwards from the end formula, the branch to be followed is unique as long as we do not encounter an $\wedge^+$-application. Now let us consider normal

derivations. Recall the notion of a strictly positive part of a formula, defined in 1.1.3.

PROPOSITION. *Let M be a normal derivation, and let $\pi = \sigma_0, \ldots, \sigma_n$ be a track in M. Then there is a segment σ_i in π, the* minimum *segment or* minimum *part of the track, which separates two (possibly empty) parts of π, called the E-part (*elimination part*) and the I-part (*introduction part*) of π such that*

(a) *for each σ_j in the E-part one has $j < i$, σ_j is a major premise of an E-rule, and σ_{j+1} is a strictly positive part of σ_j, and therefore each σ_j is a s.p.p. of σ_0;*
(b) *for each σ_j which is in the I-part or is the minimum segment one has $i \leq j$, and if $j \neq n$, then σ_j is a premise of an I-rule and a s.p.p. of σ_{j+1}, so each σ_j is a s.p.p. of σ_n.*

PROOF. By tracing through the definitions. □

THEOREM (Subformula property). *Let M be a normal derivation. Then each formula occurring in the derivation is a subformula of either the end formula or else an (uncancelled) assumption formula.*

PROOF. As noted above, each track of order 0 in M is a track ending in the end formula of M. Furthermore each track has an E-part above an I-part. Therefore any formula on a track of order 0 is either a subformula of the end formula or else a subformula of an (uncancelled) assumption. We can now prove the theorem for tracks of order n, by induction on n. So assume the result holds for tracks of order n. If A is any formula on a track of order $n+1$, either A lies in the E-part in which case it is a subformula of an assumption, or else it lies in the I-part and is therefore a subformula of the minor premise of an $\rightarrow^-$ whose main premise belongs to a track of order n. In this case A is a subformula of a formula on a track of order n and we can apply the induction hypothesis. □

THEOREM (Disjunction property). *If no strictly positive part of a formula in Γ is a disjunction, then $\Gamma \vdash A \vee B$ implies $\Gamma \vdash A$ or $\Gamma \vdash B$.*

PROOF. Consider a normal derivation M of $A \vee B$ from assumptions Γ not containing a disjunction as s.p.p. The end formula $A \vee B$ is the final formula of a (main) track. If the I-part of this track is empty, then the structure of main tracks ensures that $A \vee B$ would be a s.p.p. of an assumption in Γ, but this is not allowed. Hence $A \vee B$ lies in the I-part of a main track. If above $A \vee B$ this track goes through a minor premise of an $\vee^-$, then the major premise would again be a disjunctive s.p.p. of an assumption, which is not allowed. Thus $A \vee B$ belongs to a segment within the I-part of the track, above which there can only be finitely many $\exists^-$ and

$\wedge^-$ followed by an $\vee_i^+$. Its premise is either A or B, and therefore we can replace the segment of $A \vee B$'s by a segment of A's or a segment of B's, thus transforming the proof into either a proof of A or a proof of B. □

There is a similar theorem for the existential quantifier:

THEOREM (Explicit definability under hypotheses). *If no strictly positive part of a formula in* Γ *is existential, then* $\Gamma \vdash \exists_x A(x)$ *implies* $\Gamma \vdash A(r_1) \vee \cdots \vee A(r_n)$ *for some terms* $r_1, \dots, r_n$. *If in addition no s.p.p. of a formula in* Γ *is disjunctive then* $\Gamma \vdash \exists_x A(x)$ *implies there is even a single term* r *such that* $\Gamma \vdash A(r)$.

PROOF. Consider a normal derivation M of $\exists_x A(x)$ from assumptions Γ not containing an existential s.p.p. We use induction on the derivation, and distinguish cases on the last rule.

By assumption the last rule cannot be $\exists^-$, using a similar argument to the above. Again as before, the only critical case is when the last rule is $\vee^-$.

$$\dfrac{\begin{array}{ccc} & [u\colon B] & [v\colon C] \\ \mid M & \mid N_0 & \mid N_1 \\ B \vee C & \exists_x A(x) & \exists_x A(x) \end{array}}{\exists_x A(x)} \vee^- u, v$$

By assumption again neither B nor C can have an existential s.p.p. Applying the induction hypothesis to N_0 and N_1 we obtain

$$\dfrac{\begin{array}{ccc} & [u\colon B] & [v\colon C] \\ & \mid & \mid \\ \mid M & \dfrac{\bigvee_{i=1}^{n} A(r_i)}{\bigvee_{i=1}^{n+m} A(r_i)} \vee^+ & \dfrac{\bigvee_{i=n+1}^{n+m} A(r_i)}{\bigvee_{i=1}^{n+m} A(r_i)} \vee^+ \\ B \vee C & & \end{array}}{\bigvee_{i=1}^{n+m} A(r_i)} \vee^- u, v$$

The remaining cases are left to the reader.

The second part of the theorem is proved similarly; by assumption the last rule can be neither $\vee^-$ nor $\exists^-$, so it may be an $\wedge^-$. In that case there is only one minor premise and so no need to duplicate instances of $A(x)$. □

1.2.7. Orevkov formulas. We give examples due to Orevkov (1979) of formulas C_i which need derivation height superexponential in i if normal derivations are required, but have non-normal derivations of height linear in i. The non-normal derivations of C_i make use of auxiliary formulas with an i-fold nesting of implications and universal quantifiers. This sheds some light on the power of abstract notions in mathematics: their use can shorten proofs dramatically.

We work in a language with a ternary relation symbol R, a constant 0 and a unary function symbol S. The intuitive meaning of $Ryxz$ is $y+2^x = z$, and we can express this by means of two ("Horn"-) clauses

$$\begin{aligned}&\mathrm{Hyp}_1 := \forall_y R(y, 0, \mathrm{S}y),\\ &\mathrm{Hyp}_2 := \forall_{y,x,z,z_1}(Ryxz \to Rzxz_1 \to R(y, \mathrm{S}x, z_1)).\end{aligned}$$

Let

$$\begin{aligned}&D_i := \tilde{\exists}_{z_i,z_{i-1},\dots,z_0}(R00z_i \mathbin{\tilde{\wedge}} R0z_iz_{i-1} \mathbin{\tilde{\wedge}} \dots \mathbin{\tilde{\wedge}} R0z_1z_0),\\ &C_i := \mathrm{Hyp}_1 \to \mathrm{Hyp}_2 \to D_i.\end{aligned}$$

(for $\tilde{\wedge}$ cf. the remark at the end of 1.1.7). D_i intuitively means that there are numbers $z_i = 1$, $z_{i-1} = 2^{z_i} = 2$, $z_{i-2} = 2^{z_{i-1}} = 2^2$, $z_{i-3} = 2^{z_{i-2}} = 2^{2^2}$ and finally $z_0 = 2_i$ (where $2_0 := 1$, $2_{n+1} := 2^{2_n}$).

To obtain short derivations of C_i we use the following "lifting" formulas:

$$\begin{aligned}&A_0(x) := \forall_y \tilde{\exists}_z Ryxz,\\ &A_{i+1}(x) := \forall_{y\in A_i} \tilde{\exists}_{z\in A_i} Ryxz,\end{aligned}$$

where $\forall_{z\in A_i} B$ abbreviates $\forall_z(A_i(z) \to B)$.

Lemma. *There are derivations of*
(a) $\forall_x(A_i(x) \to A_i(\mathrm{S}x))$ *from* Hyp_2 *and of*
(b) $A_i(0)$ *from* Hyp_1 *and* Hyp_2,
both of constant (i.e., independent of i) height.

Proof. Unfolding $\tilde{\exists}$ gives

$$\begin{aligned}&D_i = \forall_{z_i,z_{i-1},\dots,z_0}(R00z_i \to R0z_iz_{i-1} \to \dots \to R0z_1z_0 \to \bot) \to \bot,\\ &A_0(x) = \forall_y(\forall_z(Ryxz \to \bot) \to \bot),\\ &A_{i+1}(x) = \forall_{y\in A_i}(\forall_{z\in A_i}(Ryxz \to \bot) \to \bot).\end{aligned}$$

(a). Derivations M_i of $\forall_x(A_i(x) \to A_i(\mathrm{S}x))$ from Hyp_2 with constant height are constructed as follows. We use assumption variables

$$d\colon A_i(x), \quad e_3\colon Ryxz, \quad e_5\colon Rzxz_1, \quad w_0\colon \forall_{z_1}\neg R(y, \mathrm{S}x, z_1)$$

and in case $i > 0$

$$e_1\colon A_{i-1}(y), \quad e_2\colon A_{i-1}(z), \quad e_4\colon A_{i-1}(z_1), \quad w\colon \forall_{z_1\in A_{i-1}}\neg R(y, \mathrm{S}x, z_1).$$

Take in case $i = 0$

$$M_i := \lambda_{x,d,y,w_0}(dy\lambda_{z,e_3}(dz\lambda_{z_1,e_5}(w_0z_1(\mathrm{Hyp}_2yxzz_1e_3e_5))))$$

and in case $i > 0$

$$M_i := \lambda_{x,d,y,e_1,w}(dye_1\lambda_{z,e_2,e_3}(dze_2\lambda_{z_1,e_4,e_5}(wz_1e_4(\mathrm{Hyp}_2yxzz_1e_3e_5)))).$$

Notice that d is used twice in these derivations.

(b). Clearly $A_0(0)$ can be derived from Hyp_1. For $i > 0$ the required derivation of $A_i(0)$ from $\mathrm{Hyp}_1, \mathrm{Hyp}_2$ of constant height can be constructed from $M_{i-1}\colon \forall_x(A_{i-1}(x) \to A_{i-1}(\mathrm{S}x))$ and the assumption variables

$$d\colon A_{i-1}(x), \quad e\colon \forall_{z\in A_{i-1}} \neg Rx0z.$$

Take

$$N_i := \lambda_{x,d,e}(e(\mathrm{S}x)(M_{i-1}xd)(\mathrm{Hyp}_1 x)). \qquad \square$$

PROPOSITION. *There are derivations of D_i from* Hyp_1 *and* Hyp_2 *with height linear in i.*

PROOF. Let N_i be the derivation $A_i(0)$ from Hyp_1, Hyp_2 constructed in the lemma above. Let

$$\begin{aligned} K_0 &:= w_0 z_0 v_0, \\ K_1 &:= u_1 0 \lambda_{z_0,v_0}(w_1 z_1 v_1 z_0 v_0), \\ K_i &:= u_i 0 N_{i-2} \lambda_{z_{i-1},u_{i-1},v_{i-1}} K_{i-1}[w_{i-1} := w_i z_i v_i] \quad (i \geq 2) \end{aligned}$$

with assumption variables

$$\begin{aligned} &u_i\colon A_{i-1}(z_i) \quad (i > 0), \\ &v_i\colon R0z_{i+1}z_i, \\ &w_i\colon \forall_{z_i}(R0z_{i+1}z_i \to \forall_{z_{i-1}}(R0z_iz_{i-1} \to \ldots \forall_{z_0}(R0z_1z_0 \to \bot)\ldots)). \end{aligned}$$

K_i has free object variables z_{i+1}, z_i and free assumption variables u_i, v_i, w_i (with u_i missing in case $i = 0$). Substitute z_{i+1} by 0 and z_i by S0 in K_i. The result has free assumption variables among Hyp_1, Hyp_2 and

$$\begin{aligned} &u_i'\colon A_{i-1}(\mathrm{S0}) \quad (i > 0), \\ &v_i'\colon R(0, 0, \mathrm{S0}), \\ &w_i'\colon \forall_{z_i}(R00z_i \to \forall_{z_{i-1}}(R0z_iz_{i-1} \to \ldots \forall_{z_0}(R0z_1z_0 \to \bot)\ldots)). \end{aligned}$$

Now $A_{i-1}(\mathrm{S0})$ can be derived from Hyp_1, Hyp_2 with constant height by the lemma above, and clearly $R(0, 0, \mathrm{S0})$ as well. K_i has height linear in i. Hence we have a derivation of

$$\forall_{z_i}(R00z_i \to \forall_{z_{i-1}}(R0z_iz_{i-1} \to \ldots \forall_{z_0}(R0z_1z_0 \to \bot)\ldots)) \to \bot$$

from Hyp_1, Hyp_2 of height linear in i. But this formula is up to making the premise prenex the same as D_i, and this transformation can clearly be done by a derivation of height again linear in i. $\square$

THEOREM. *Every normal derivation of D_i from* Hyp_1, Hyp_2 *has at least* 2_i *nodes.*

PROOF. Let L_i be a normal derivation of $\bot$ from Hyp_1, Hyp_2 and the assumption

$$E_i := \forall_{z_i,z_{i-1},\dots,z_0}(R0\overline{0}z_i \to R0z_iz_{i-1} \to \cdots \to R0z_1z_0 \to \bot).$$

We can assume that L_i has no free variables; otherwise replace them by 0.

The main branch of L_i starts with E_i followed by $i+1$ applications of $\forall^-$ followed by $i+1$ applications of $\to^-$. All minor premises are of the form $R0\bar{n}\bar{k}$ (where $\bar{0} := 0$, $\overline{n+1} := \mathrm{S}\bar{n}$).

Let M be an arbitrary normal derivation of $R\bar{m}\bar{n}\bar{k}$ from E_i, Hyp_1, Hyp_2. We show that M (i) contains at least 2^n occurrences of Hyp_1, and (ii) satisfies $m+2^n=k$. We prove (i) and (ii) by induction on n. The base case is obvious. For the step case we can assume that every normal derivation of $R\bar{m}\bar{n}\bar{k}$ from E_i, Hyp_1, Hyp_2 contains at least 2^n occurrences of Hyp_1, and satisfies $m+2^n=k$. Now consider an arbitrary normal derivation of $R(\bar{m},\mathrm{S}\bar{n},\bar{k})$. It must end with

$$\dfrac{\dfrac{\begin{array}{cc}\vert & \vert M_1\\ R\bar{m}\bar{n}\bar{l}\to R\bar{l}\bar{n}\bar{k}\to R(\bar{m},\mathrm{S}\bar{n},\bar{k}) & R\bar{m}\bar{n}\bar{l}\end{array}}{R\bar{l}\bar{n}\bar{k}\to R(\bar{m},\mathrm{S}\bar{n},\bar{k})} \qquad \begin{array}{c}\vert M_2\\ R\bar{l}\bar{n}\bar{k}\end{array}}{R(\bar{m},\mathrm{S}\bar{n},\bar{k})}$$

By induction hypothesis both M_1, M_2 contain at least 2^n occurrences of Hyp_1, and we have $m+2^n=l$ and $l+2^n=k$, hence $m+2^{n+1}=k$. (It is easy to see that M does not use the assumption E_i.)

We now come back to the main branch of L_i, in particular its minor premises. They derive $R00\bar{1}$, $R0\bar{1}\bar{2}$ and so on until $R(0,\overline{2_{i-1}},\overline{2_i})$. Hence altogether we have $\sum_{j\le i}2^j = 2^{i+1}-1$ occurrences of Hyp_1. □

1.3. Soundness and Completeness for Tree Models

It is an obvious question to ask whether the logical rules we have been considering suffice, i.e., whether we have forgotten some necessary rules. To answer this question we first have to fix the *meaning* of a formula, i.e., provide a semantics. This will be done by means of the tree models introduced by Beth (1956). Using this concept of a model we will prove soundness and completeness.

1.3.1. Tree models. Consider a finitely branching tree of "possible worlds". The worlds are represented as nodes in this tree. They may be thought of as possible states such that all nodes "above" a node k are the ways in which k may develop in the future. The worlds are increasing, that is, if an atomic formula $R\vec{s}$ is true in a world k, then $R\vec{s}$ is true in all future worlds k'.

More formally, each tree model is based on a finitely branching tree T. A *node* k over a set S is a finite sequence $k = \langle a_0, a_1, \dots, a_{n-1} \rangle$ of elements of S; $\mathrm{lh}(k)$ is the length of k. We write $k \preceq k'$ if k is an initial segment of k'. A *tree* on S is a set of nodes closed under initial segments. A tree T is *finitely branching* if every node in T has finitely many immediate successors. A tree T is *infinite* if for every $n \in \mathbb{N}$ there is a node $k \in T$ such that $\mathrm{lh}(k) = n$. A *branch* of a tree T is a linearly ordered subtree of T, and a *leaf* of T is a node without successors in T. A tree T is *complete* is every node in T has an immediate successor, i.e., T has no leaves.

For the proof of the completeness theorem, the complete tree over $\{0, 1\}$ (whose branches constitute Cantor space) will suffice. The nodes will be all the finite sequences of 0's and 1's, and the ordering is as above. The root is the empty sequence and $k0$ is the sequence k with the element 0 added at the end; similarly for $k1$.

For the rest of this section, fix a countable formal language $\mathcal{L}$.

Definition. Let T be a finitely branching tree. A *tree model* on T is a triple $\mathcal{T} = (D, I_0, I_1)$ such that

(a) D is a nonempty set;
(b) for every n-ary function symbol f (in the underlying language $\mathcal{L}$), I_0 assigns to f a map $I_0(f)\colon D^n \to D$;
(c) for every n-ary relation symbol R and every node $k \in T$, $I_1(R, k) \subseteq D^n$ is assigned in such a way that monotonicity is preserved:

$$k \preceq k' \to I_1(R, k) \subseteq I_1(R, k').$$

If $n = 0$, then $I_1(R, k)$ is either true or false. There is no special requirement set on $I_1(\bot, k)$. (Recall that minimal logic places no particular constraints on falsum $\bot$.) We write $R^{\mathcal{T}}(\vec{a}, k)$ for $\vec{a} \in I_1(R, k)$, and $|\mathcal{T}|$ to denote the domain D.

It is obvious from the definition that any tree T can be extended to a complete tree $\bar{T}$ (i.e., without leaves), in which for every leaf $k \in T$ all sequences $k0$, $k00$, $k000, \dots$ are added to T. For every node $k0 \dots 0$, we then add $I_1(R, k0 \dots 0) := I_1(R, k)$.

An *assignment* (or variable assignment) in D is a map η assigning to every variable $x \in \mathrm{dom}(\eta)$ a value $\eta(x) \in D$. Finite assignments will be written as $[x_1 := a_1, \dots, x_n := a_n]$ or else as $[a_1/x_1, \dots, a_n/x_n]$, with distinct $x_1, \dots, x_n$. If η is an assignment in D and $a \in D$, let η^a_x be the assignment in D mapping x to a and coinciding with η elsewhere:

$$\eta^a_x(y) := \begin{cases} \eta(y) & \text{if } y \neq x, \\ a & \text{if } y = x. \end{cases}$$

Let a tree model $\mathcal{T} = (D, I_0, I_1)$ and an assignment η in D be given. We define a homomorphic extension of η (denoted by η as well) to terms t whose variables lie in $\mathrm{dom}(\eta)$ by

$$\begin{aligned}\eta(c) &:= I_0(c),\\ \eta(f(t_1,\dots,t_n)) &:= I_0(f)(\eta(t_1),\dots,\eta(t_n)).\end{aligned}$$

Observe that the extension of η depends on $\mathcal{T}$; we often write $t^{\mathcal{T}}[\eta]$ for $\eta(t)$.

Definition. $\mathcal{T}, k \Vdash A[\eta]$ ($\mathcal{T}$ *forces* A at node k for an assignment η) is defined inductively. We write $k \Vdash A[\eta]$ when it is clear from the context what the underlying model $\mathcal{T}$ is, and $\forall_{k' \succeq_n k} A$ for $\forall_{k' \succeq k}(\mathrm{lh}(k') = \mathrm{lh}(k) + n \to A)$.

$$\begin{aligned}
k \Vdash (R\vec{s})[\eta] &:= \exists_n \forall_{k' \succeq_n k} R^{\mathcal{T}}(\vec{s}^{\mathcal{T}}[\eta], k'),\\
k \Vdash (A \vee B)[\eta] &:= \exists_n \forall_{k' \succeq_n k}(k' \Vdash A[\eta] \vee k' \Vdash B[\eta]),\\
k \Vdash (\exists_x A)[\eta] &:= \exists_n \forall_{k' \succeq_n k} \exists_{a \in |\mathcal{T}|}(k' \Vdash A[\eta^a_x]),\\
k \Vdash (A \to B)[\eta] &:= \forall_{k' \succeq k}(k' \Vdash A[\eta] \to k' \Vdash B[\eta]),\\
k \Vdash (A \wedge B)[\eta] &:= k \Vdash A[\eta] \wedge k \Vdash B[\eta],\\
k \Vdash (\forall_x A)[\eta] &:= \forall_{a \in |\mathcal{T}|}(k \Vdash A[\eta^a_x]).
\end{aligned}$$

Thus in the atomic, disjunctive and existential cases, the set of k' whose length is $\mathrm{lh}(k) + n$ acts as a "bar" in the complete tree. Note that the implicational case is treated differently, and refers to the "unbounded future".

In this definition, the logical connectives $\to, \wedge, \vee, \forall, \exists$ on the left hand side are part of the object language, whereas the same connectives on the right hand side are to be *understood* in the usual sense: they belong to the "metalanguage". It should always be clear from the context whether a formula is part of the object or the metalanguage.

1.3.2. Covering lemma. It is easily seen (using the definition and monotonicity) that from $k \Vdash A[\eta]$ and $k \preceq k'$ we can conclude $k' \Vdash A[\eta]$. The converse is true as well:

Lemma (Covering).

$$\forall_{k' \succeq_n k}(k' \Vdash A[\eta]) \to k \Vdash A[\eta].$$

Proof. Induction on A. We write $k \Vdash A$ for $k \Vdash A[\eta]$.
Case $R\vec{s}$. Assume

$$\forall_{k' \succeq_n k}(k' \Vdash R\vec{s}),$$

hence by definition

$$\forall_{k' \succeq_n k} \exists_m \forall_{k'' \succeq_m k'} R^{\mathcal{T}}(\vec{s}^{\mathcal{T}}[\eta], k'').$$

Since T is a finitely branching tree,

$$\exists_m \forall_{k' \succeq_m k} R^{\mathcal{T}}(\vec{s}^{\mathcal{T}}[\eta], k').$$

Hence $k \Vdash R\vec{s}$.

The cases $A \vee B$ and $\exists_x A$ are handled similarly.

Case $A \to B$. Let $k' \Vdash A \to B$ for all $k' \succeq k$ with $\mathrm{lh}(k') = \mathrm{lh}(k) + n$. We show

$$\forall_{l \succeq k}(l \Vdash A \to l \Vdash B).$$

Let $l \succeq k$ and $l \Vdash A$. We must show $l \Vdash B$. To this end we apply the induction hypothesis to B and $m := \max(\mathrm{lh}(k) + n, \mathrm{lh}(l))$. So assume $l' \succeq l$ and $\mathrm{lh}(l') = m$. It is sufficient to show $l' \Vdash B$. If $\mathrm{lh}(l') = \mathrm{lh}(l)$, then $l' = l$ and we are done. If $\mathrm{lh}(l') = \mathrm{lh}(k) + n > \mathrm{lh}(l)$, then l' is an extension of l as well as of k and has length $\mathrm{lh}(k) + n$, and hence $l' \Vdash A \to B$ by assumption. Moreover, $l' \Vdash A$, since $l' \succeq l$ and $l \Vdash A$. It follows that $l' \Vdash B$.

The cases $A \wedge B$ and $\forall_x A$ are easy. □

1.3.3. Soundness.

Lemma (Coincidence). *Let $\mathcal{T}$ be a tree model, t a term, A a formula and η, ξ assignments in $|\mathcal{T}|$.*

(a) *If $\eta(x) = \xi(x)$ for all $x \in \mathrm{vars}(t)$, then $\eta(t) = \xi(t)$.*
(b) *If $\eta(x) = \xi(x)$ for all $x \in \mathrm{FV}(A)$, then $\mathcal{T}, k \Vdash A[\eta]$ if and only if $\mathcal{T}, k \Vdash A[\xi]$.*

Proof. Induction on terms and formulas. □

Lemma (Substitution). *Let $\mathcal{T}$ be a tree model, $t, r(x)$ terms, $A(x)$ a formula and η an assignment in $|\mathcal{T}|$. Then*

(a) $\eta(r(t)) = \eta_x^{\eta(t)}(r(x))$.
(b) *$\mathcal{T}, k \Vdash A(t)[\eta]$ if and only if $\mathcal{T}, k \Vdash A(x)[\eta_x^{\eta(t)}]$.*

Proof. Induction on terms and formulas. □

Theorem (Soundness). *Let $\Gamma \cup \{A\}$ be a set of formulas such that $\Gamma \vdash A$. Then, if $\mathcal{T}$ is a tree model, k any node and η an assignment in $|\mathcal{T}|$, it follows that $\mathcal{T}, k \Vdash \Gamma[\eta]$ implies $\mathcal{T}, k \Vdash A[\eta]$.*

Proof. Induction on derivations.

We begin with the axiom schemes $\vee_0^+$, $\vee_1^+$, $\vee^-$, $\wedge^+$, $\wedge^-$, $\exists^+$ and $\exists^-$. $k \Vdash C[\eta]$ is abbreviated $k \Vdash C$, when η is known from the context.

Case $\vee_0^+ \colon A \to A \vee B$. We show $k \Vdash A \to A \vee B$. Assume for $k' \succeq k$ that $k' \Vdash A$. Show: $k' \Vdash A \vee B$. This follows from the definition, since $k' \Vdash A$. The case $\vee_1^+ \colon B \to A \vee B$ is symmetric.

Case $\vee^- \colon A \vee B \to (A \to C) \to (B \to C) \to C$. We show that $k \Vdash A \vee B \to (A \to C) \to (B \to C) \to C$. Assume for $k' \succeq k$ that

$k' \Vdash A \vee B$, $k' \Vdash A \to C$ and $k' \Vdash B \to C$ (we can safely assume that k' is the same for all three premises.) Show that $k' \Vdash C$. By definition, there is an n s.t. for all $k'' \succeq_n k'$, $k'' \Vdash A$ or $k'' \Vdash B$. In both cases it follows that $k'' \Vdash C$, since $k' \Vdash A \to C$ and $k' \Vdash B \to C$. By the covering lemma, $k' \Vdash C$.

The cases $\wedge^+$, $\wedge^-$ are easy.

Case $\exists^+$: $A \to \exists_x A$. We show $k \Vdash (A \to \exists_x A)[\eta]$. Assume $k' \succeq k$ and $k' \Vdash A[\eta]$. We show $k' \Vdash (\exists_x A)[\eta]$. Since $\eta = \eta_x^{\eta(x)}$ there is an $a \in |\mathcal{T}|$ (namely $a := \eta(x)$) such that $k' \Vdash A[\eta_x^a]$. Hence, $k' \Vdash (\exists_x A)[\eta]$.

Case $\exists^-$: $\exists_x A \to \forall_x(A \to B) \to B$ and $x \notin \mathrm{FV}(B)$. We show that $k \Vdash (\exists_x A \to \forall_x(A \to B) \to B)[\eta]$. Assume that $k' \succeq k$ and $k' \Vdash (\exists_x A)[\eta]$ and $k' \Vdash \forall_x(A \to B)[\eta]$. We show $k' \Vdash B[\eta]$. By definition, there is an n such that for all $k'' \succeq_n k'$ we have $a \in |\mathcal{T}|$ and $k'' \Vdash A[\eta_x^a]$. From $k' \Vdash \forall_x(A \to B)[\eta]$ it follows that $k'' \Vdash B[\eta_x^a]$, and since $x \notin \mathrm{FV}(B)$, from the coincidence lemma, $k'' \Vdash B[\eta]$. Then, finally, by the covering lemma $k' \Vdash B[\eta]$.

This concludes the treatment of the axioms. We now consider the rules. In case of the assumption rule u: A we have $A \in \Gamma$ and the claim is obvious.

Case $\to^+$. Assume $k \Vdash \Gamma$. We show $k \Vdash A \to B$. Assume $k' \succeq k$ and $k' \Vdash A$. Our goal is $k' \Vdash B$. We have $k' \Vdash \Gamma \cup \{A\}$. Thus, $k' \Vdash B$ by induction hypothesis.

Case $\to^-$. Assume $k \Vdash \Gamma$. The induction hypothesis gives us $k \Vdash A \to B$ and $k \Vdash A$. Hence $k \Vdash B$.

Case $\forall^+$. Assume $k \Vdash \Gamma[\eta]$ and $x \notin \mathrm{FV}(\Gamma)$. We show $k \Vdash (\forall_x A)[\eta]$, i.e., $k \Vdash A[\eta_x^a]$ for an arbitrary $a \in |\mathcal{T}|$. We have

$$\begin{array}{ll} k \Vdash \Gamma[\eta_x^a] & \text{by the coincidence lemma, since } x \notin \mathrm{FV}(\Gamma) \\ k \Vdash A[\eta_x^a] & \text{by induction hypothesis.} \end{array}$$

Case $\forall^-$. Let $k \Vdash \Gamma[\eta]$. We show that $k \Vdash A(t)[\eta]$. This follows from

$$\begin{array}{ll} k \Vdash (\forall_x A(x))[\eta] & \text{by induction hypothesis} \\ k \Vdash A(x)[\eta_x^{\eta(t)}] & \text{by definition} \\ k \Vdash A(t)[\eta] & \text{by the substitution lemma.} \end{array}$$

This concludes the proof. □

1.3.4. Counter models. With soundness at hand, it is easy to build counter models proving that certain formulas are underivable in minimal or intuitionistic logic. A *tree model for intuitionistic logic* is a tree model $\mathcal{T} = (D, I_0, I_1)$ in which $I_1(\bot, k)$ is false for all k. This is equivalent to saying that $\bot$ is never forced:

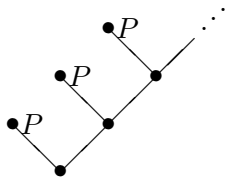
LEMMA. *Given any tree model $\mathcal{T}$, $\bot^{\mathcal{T}}(k)$ is false at all nodes k if and only if $k \Vdash \bot$ for all nodes k.*

PROOF. Clearly if $k \Vdash \bot$ then $\bot$ is false at node k. Conversely, suppose $\bot^{\mathcal{T}}(k')$ is false at all nodes k'. We must show $\forall_k(k \Vdash \bot)$. Let k be given. Then, since $\bot^{\mathcal{T}}(k')$ is false at all nodes k', is is certainly false at some $k' \succeq_n k$, for every n. This means $k \Vdash \bot$ by definition. □

Therefore by unravelling the implication clause in the forcing definition, one sees that in any tree model for intuitionistic logic,

$$\begin{aligned}(k \Vdash \neg A) &\leftrightarrow \forall_{k' \succeq k}(k' \nVdash A),\\ (k \Vdash \neg\neg A) &\leftrightarrow \forall_{k' \succeq k}(k' \nVdash \neg A)\\ &\leftrightarrow \forall_{k' \succeq k}\tilde{\exists}_{k'' \succeq k'}(k'' \Vdash A).\end{aligned}$$

As an example we show that $\nvdash_i \neg\neg P \to P$. We describe the desired tree model by means of a diagram below. Next to every node we write all propositions forced at that node.



This is a tree model because monotonicity clearly holds. Observe also that $I_1(\bot, k)$ is false at all nodes k. Hence this is an intuitionistic tree model, and moreover $\langle\rangle \nVdash P$. Using the remark above, it is easily seen that $\langle\rangle \Vdash \neg\neg P$. Thus $\langle\rangle \nVdash (\neg\neg P \to P)$ and hence $\nvdash_i (\neg\neg P \to P)$. The model also shows that the *Peirce formula* $((P \to Q) \to P) \to P$ is not derivable in intuitionistic logic.

As another example we show that the drinker formula $\tilde{\exists}_x(Px \to \forall_x Px)$ from 1.1.8 is intuitionistically underivable, using a quite different tree model. In this case the underlying tree is the full binary one, i.e., its nodes are the finite sequences $k = \langle i_0, i_1, \ldots, i_{n-1}\rangle$ of numbers 0 or 1. For the language determined by $\bot$ and a unary predicate symbol P consider $\mathcal{T} := (D, I_1)$ with $I_1(\bot, k)$ false, $D := \mathbb{N}$ and

$$I_1(P, \langle i_0, \ldots, i_{n-1}\rangle) := \{\, a \in D \mid i_0, \ldots, i_{n-1} \text{ contains at least } a \text{ zeros} \,\}.$$

Cleary $\mathcal{T}$ is an intuitionistic tree model (monotonicity is easily checked), $k \nVdash \forall_x Px$ for every k, and $\forall_{a,k}\exists_{l \succeq k}(l \Vdash Px[x := a])$. Therefore

$$\begin{aligned}&\forall_{a,k}(k \nVdash (Px \to \forall_x Px)[x := a])\\ &\langle\rangle \Vdash \forall_x\neg(Px \to \forall_x Px).\end{aligned}$$

Hence $\nvdash_i \neg\forall_x\neg(Px \to \forall_x Px)$.

1.3.5. Completeness.

THEOREM (Completeness). *Let $\Gamma \cup \{A\}$ be a set of formulas. Then the following propositions are equivalent.*

(a) $\Gamma \vdash A$.

(b) $\Gamma \Vdash A$, *i.e., for all tree models $\mathcal{T}$, nodes k and assignments η*

$$\mathcal{T}, k \Vdash \Gamma[\eta] \to \mathcal{T}, k \Vdash A[\eta].$$

PROOF. Soundness already gives "(a) implies (b)". For the other direction we employ a technique due to Harvey Friedman and construct a tree model $\mathcal{T}$ (over the set T_{01} of all finite 0-1-sequences) whose domain D is the set of all terms of the underlying language, with the property that $\Gamma \vdash B$ is equivalent to $\mathcal{T}, \langle\rangle \Vdash B[\mathrm{id}]$. We can assume here that Γ and also A are closed.

In order to define $\mathcal{T}$, we will need an enumeration $A_0, A_1, A_2, \dots$ of the underlying language $\mathcal{L}$ (assumed countable), in which every formula occurs infinitely often. We also fix an enumeration $x_0, x_1, \dots$ of distinct variables. Since Γ is countable it can we written $\Gamma = \bigcup_n \Gamma_n$ with finite sets Γ_n such that $\Gamma_n \subseteq \Gamma_{n+1}$. With every node $k \in T_{01}$, we associate a finite set Δ_k of formulas and a set V_k of variables, by induction on the length of k.

Let $\Delta_{\langle\rangle} := \emptyset$ and $V_{\langle\rangle} := \emptyset$. Take a node k such that $\mathrm{lh}(k) = n$ and suppose that Δ_k, V_k are already defined. Write $\Delta \vdash_n B$ to mean that there is a derivation of length $\leq n$ of B from Δ. We define Δ_{k0}, V_{k0} and Δ_{k1}, V_{k1} as follows:

Case 0. $\mathrm{FV}(A_n) \not\subseteq V_k$. Then let

$$\Delta_{k0} := \Delta_{k1} := \Delta_k \quad \text{and} \quad V_{k0} := V_{k1} := V_k.$$

Case 1. $\mathrm{FV}(A_n) \subseteq V_k$ and $\Gamma_n, \Delta_k \nvdash_n A_n$. Let

$$\Delta_{k0} := \Delta_k \quad \text{and} \quad \Delta_{k1} := \Delta_k \cup \{A_n\},$$
$$V_{k0} := V_{k1} := V_k.$$

Case 2. $\mathrm{FV}(A_n) \subseteq V_k$ and $\Gamma_n, \Delta_k \vdash_n A_n = A'_n \vee A''_n$. Let

$$\Delta_{k0} := \Delta_k \cup \{A_n, A'_n\} \quad \text{and} \quad \Delta_{k1} := \Delta_k \cup \{A_n, A''_n\},$$
$$V_{k0} := V_{k1} := V_k.$$

Case 3. $\mathrm{FV}(A_n) \subseteq V_k$ and $\Gamma_n, \Delta_k \vdash_n A_n = \exists_x A'_n(x)$. Let

$$\Delta_{k0} := \Delta_{k1} := \Delta_k \cup \{A_n, A'_n(x_i)\} \quad \text{and} \quad V_{k0} := V_{k1} := V_k \cup \{x_i\},$$

where x_i is the first variable $\notin V_k$.

Case 4. $\mathrm{FV}(A_n) \subseteq V_k$ and $\Gamma_n, \Delta_k \vdash_n A_n$, with A_n neither a disjunction nor an existentially quantified formula. Let

$$\Delta_{k0} := \Delta_{k1} := \Delta_k \cup \{A_n\} \quad \text{and} \quad V_{k0} := V_{k1} := V_k.$$

Obviously $\mathrm{FV}(\Delta_k) \subseteq V_k$, and $k \preceq k'$ implies that $\Delta_k \subseteq \Delta_{k'}$. Notice also that because of $\vdash \exists_x(\bot \to \bot)$ and the fact that this formula is repeated infinitely often in the given enumeration, for every variable x_i there is an m such that $x_i \in V_k$ for all k with $\mathrm{lh}(k) = m$.

We note that

$$\forall_{k' \succeq_n k}\,(\Gamma, \Delta_{k'} \vdash B) \to \Gamma, \Delta_k \vdash B, \quad \text{provided } \mathrm{FV}(B) \subseteq V_k. \tag{1.7}$$

It is sufficient to show that, for $\mathrm{FV}(B) \subseteq V_k$,

$$(\Gamma, \Delta_{k0} \vdash B) \wedge (\Gamma, \Delta_{k1} \vdash B) \to (\Gamma, \Delta_k \vdash B).$$

In cases 0, 1 and 4, this is obvious. For case 2, the claim follows immediately from the axiom schema $\vee^-$. In case 3, we have $\mathrm{FV}(A_n) \subseteq V_k$ and $\Gamma_n, \Delta_k \vdash_n A_n = \exists_x A'_n(x)$. Assume $\Gamma, \Delta_k \cup \{A_n, A'_n(x_i)\} \vdash B$ with $x_i \notin V_k$, and $\mathrm{FV}(B) \subseteq V_k$. Then $x_i \notin \mathrm{FV}(\Delta_k \cup \{A_n, B\})$, hence $\Gamma, \Delta_k \cup \{A_n\} \vdash B$ by $\exists^-$ and therefore $\Gamma, \Delta_k \vdash B$.

Next, we show

$$\Gamma, \Delta_k \vdash B \to \exists_n \forall_{k' \succeq_n k}\,(B \in \Delta_{k'}), \quad \text{provided } \mathrm{FV}(B) \subseteq V_k. \tag{1.8}$$

Choose $n \geq \mathrm{lh}(k)$ such that $B = A_n$ and $\Gamma_n, \Delta_k \vdash_n A_n$. For all $k' \succeq k$, if $\mathrm{lh}(k') = n + 1$ then $A_n \in \Delta_{k'}$ (cf. the cases 2-4).

Using the sets Δ_k we can define a tree model $\mathcal{T}$ as (Ter, I_0, I_1) where Ter denotes the set of terms of the underlying language, $I_0(f)(\vec{s}) := f\vec{s}$ and

$$R^{\mathcal{T}}(\vec{s}, k) = I_1(R, k)(\vec{s}) := (R\vec{s} \in \Delta_k).$$

Obviously, $t^{\mathcal{T}}[\mathrm{id}] = t$ for all terms t.

Now write $k \Vdash B$ for $\mathcal{T}, k \Vdash B[\mathrm{id}]$. We show:

Claim. $\Gamma, \Delta_k \vdash B \leftrightarrow k \Vdash B$ *provided* $\mathrm{FV}(B) \subseteq V_k$.

The proof is by induction on B.

Case $R\vec{s}$. Assume $\mathrm{FV}(R\vec{s}) \subseteq V_k$. The following are equivalent.

$\Gamma, \Delta_k \vdash R\vec{s}$	
$\exists_n \forall_{k' \succeq_n k}\,(R\vec{s} \in \Delta_{k'})$	by (1.8) and (1.7)
$\exists_n \forall_{k' \succeq_n k}\, R^{\mathcal{T}}(\vec{s}, k')$	by definition of $\mathcal{T}$
$k \Vdash R\vec{s}$	by definition of $\Vdash$, since $t^{\mathcal{T}}[\mathrm{id}] = t$.

Case $B \vee C$. Assume $\mathrm{FV}(B \vee C) \subseteq V_k$. For the implication $\to$ let $\Gamma, \Delta_k \vdash B \vee C$. Choose an $n \geq \mathrm{lh}(k)$ such that $\Gamma_n, \Delta_k \vdash_n A_n = B \vee C$. Then, for all $k' \succeq k$ s.t. $\mathrm{lh}(k') = n$,

$$\Delta_{k'0} = \Delta_{k'} \cup \{B \vee C, B\} \quad \text{and} \quad \Delta_{k'1} = \Delta_{k'} \cup \{B \vee C, C\},$$

and therefore by induction hypothesis

$$k'0 \Vdash B \quad \text{and} \quad k'1 \Vdash C.$$

Then by definition we have $k \Vdash B \vee C$. For the reverse implication $\leftarrow$ argue as follows.

$$
\begin{array}{ll}
k \Vdash B \vee C & \\
\exists_n \forall_{k' \succeq_n k} (k' \Vdash B \vee k' \Vdash C) & \\
\exists_n \forall_{k' \succeq_n k} ((\Gamma, \Delta_{k'} \vdash B) \vee (\Gamma, \Delta_{k'} \vdash C)) & \text{by induction hypothesis} \\
\exists_n \forall_{k' \succeq_n k} (\Gamma, \Delta_{k'} \vdash B \vee C) & \\
\Gamma, \Delta_k \vdash B \vee C & \text{by (1.7).}
\end{array}
$$

Case $B \wedge C$. This is evident.

Case $B \to C$. Assume $\mathrm{FV}(B \to C) \subseteq V_k$. For $\to$ let $\Gamma, \Delta_k \vdash B \to C$. We must show $k \Vdash B \to C$, i.e.,

$$\forall_{k' \succeq k} (k' \Vdash B \to k' \Vdash C).$$

Let $k' \succeq k$ be such that $k' \Vdash B$. By induction hypothesis, it follows that $\Gamma, \Delta_{k'} \vdash B$. Hence $\Gamma, \Delta_{k'} \vdash C$ follows by assumption. Then again by induction hypothesis $k' \Vdash C$.

For $\leftarrow$ let $k \Vdash B \to C$, i.e., $\forall_{k' \succeq k} (k' \Vdash B \to k' \Vdash C)$. We show that $\Gamma, \Delta_k \vdash B \to C$, using (1.7). Choose $n \geq \mathrm{lh}(k)$ such that $B = A_n$. For all $k' \succeq_m k$ with $m := n - \mathrm{lh}(k)$ we show that $\Gamma, \Delta_{k'} \vdash B \to C$.

If $\Gamma_n, \Delta_{k'} \vdash_n A_n$, then $k' \Vdash B$ by induction hypothesis, and $k' \Vdash C$ by assumption. Hence $\Gamma_n, \Delta_{k'} \vdash C$ again by induction hypothesis and thus $\Gamma, \Delta_{k'} \vdash B \to C$.

If $\Gamma, \Delta_{k'} \nvdash_n A_n$, then by definition $\Delta_{k'1} = \Delta_{k'} \cup \{B\}$. Hence $\Gamma, \Delta_{k'1} \vdash B$, and thus $k'1 \Vdash B$ by induction hypothesis. Now $k'1 \Vdash C$ by assumption, and finally $\Gamma, \Delta_{k'1} \vdash C$ by induction hypothesis. From $\Delta_{k'1} = \Delta_{k'} \cup \{B\}$ it follows that $\Gamma, \Delta_{k'} \vdash B \to C$.

Case $\forall_x B(x)$. Assume $\mathrm{FV}(\forall_x B(x)) \subseteq V_k$. For $\to$ let $\Gamma, \Delta_k \vdash \forall_x B(x)$. Fix a term t. Then $\Gamma, \Delta_k \vdash B(t)$. Choose n such that $\mathrm{FV}(B(t)) \subseteq V_{k'}$ for all $k' \succeq_n k$. Then $\forall_{k' \succeq_n k} (\Gamma, \Delta_{k'} \vdash B(t))$, hence $\forall_{k' \succeq_n k} (k' \Vdash B(t))$ by induction hypothesis, hence $k \Vdash B(t)$ by the covering lemma. This holds for every term t, hence $k \Vdash \forall_x B(x)$.

For $\leftarrow$ assume $k \Vdash \forall_x B(x)$. Pick $k' \succeq_n k$ such that $A_m = \exists_x (\bot \to \bot)$, for $m := \mathrm{lh}(k) + n$. Then at height m we put some x_i into the variable sets: for $k' \succeq_n k$ we have $x_i \notin V_{k'}$ but $x_i \in V_{k'j}$. Clearly $k'j \Vdash B(x_i)$, hence $\Gamma, \Delta_{k'j} \vdash B(x_i)$ by induction hypothesis, hence (since at this height we consider the trivial formula $\exists_x (\bot \to \bot)$) also $\Gamma, \Delta_{k'} \vdash B(x_i)$. Since $x_i \notin V_{k'}$ we obtain $\Gamma, \Delta_{k'} \vdash \forall_x B(x)$. This holds for all $k' \succeq_n k$, hence $\Gamma, \Delta_k \vdash \forall_x B(x)$ by (1.7).

Case $\exists_x B(x)$. Assume $\mathrm{FV}(\exists_x B(x)) \subseteq V_k$. For $\to$ let $\Gamma, \Delta_k \vdash \exists_x B(x)$. Choose an $n \geq \mathrm{lh}(k)$ such that $\Gamma_n, \Delta_k \vdash_n A_n = \exists_x B(x)$. Then, for all $k' \succeq k$

with $\mathrm{lh}(k') = n$

$$\Delta_{k'0} = \Delta_{k'1} = \Delta_{k'} \cup \{\exists_x B(x), B(x_i)\}$$

where $x_i \notin V_{k'}$. Hence by induction hypothesis for $B(x_i)$ (applicable since $\mathrm{FV}(B(x_i)) \subseteq V_{k'j}$ for $j = 0, 1$)

$$k'0 \Vdash B(x_i) \quad \text{and} \quad k'1 \Vdash B(x_i).$$

It follows by definition that $k \Vdash \exists_x B(x)$.

For $\leftarrow$ assume $k \Vdash \exists_x B(x)$. Then $\forall_{k' \succeq_n k} \exists_{t \in \mathrm{Ter}} (k' \Vdash B(x)[\mathrm{id}^t_x])$ for some n, hence $\forall_{k' \succeq_n k} \exists_{t \in \mathrm{Ter}} (k' \Vdash B(t))$. For each of the finitely many $k' \succeq_n k$ pick an m such that $\forall_{k'' \succeq_m k'} (\mathrm{FV}(B(t_{k'})) \subseteq V_{k''})$. Let m_0 be the maximum of all these m. Then

$$\forall_{k'' \succeq_{m_0+n} k} \exists_{t \in \mathrm{Ter}} ((k'' \Vdash B(t)) \wedge \mathrm{FV}(B(t)) \subseteq V_{k''}).$$

The induction hypothesis for $B(t)$ yields

$$\begin{aligned}
&\forall_{k'' \succeq_{m_0+n} k} \exists_{t \in \mathrm{Ter}} (\Gamma, \Delta_{k''} \vdash B(t)) \\
&\forall_{k'' \succeq_{m_0+n} k} (\Gamma, \Delta_{k''} \vdash \exists_x B(x)) \\
&\Gamma, \Delta_k \vdash \exists_x B(x) \qquad\qquad \text{by (1.7)}
\end{aligned}$$

and this completes the proof of the claim.

Now we can finish the proof of the completeness theorem by showing that (b) implies (a). We apply (b) to the tree model $\mathcal{T}$ constructed above from Γ, the empty node $\langle\rangle$ and the assignment $\eta = \mathrm{id}$. Then $\mathcal{T}, \langle\rangle \Vdash \Gamma[\mathrm{id}]$ by the claim (since each formula in Γ is derivable from Γ). Hence $\mathcal{T}, \langle\rangle \Vdash A[\mathrm{id}]$ by (b) and therefore $\Gamma \vdash A$ by the claim again. □

Completeness of intuitionistic logic follows as a corollary.

Corollary. *Let $\Gamma \cup \{A\}$ be a set of formulas. The following propositions are equivalent.*

(a) $\Gamma \vdash_i A$.

(b) $\Gamma, \mathrm{Efq} \Vdash A$, *i.e., for all tree models $\mathcal{T}$ for intuitionistic logic, nodes k and assignments η*

$$\mathcal{T}, k \Vdash \Gamma[\eta] \to \mathcal{T}, k \Vdash A[\eta].$$ □

1.4. Soundness and Completeness of the Classical Fragment

We give a proof of completeness of classical logic relying on the completeness proof for minimal logic above.

1.4.1. Models. We define the notion of a (classical) model (or more accurately, $\mathcal{L}$-model), and what the value of a term and the meaning of a formula in a model should be. The latter definition is by induction on formulas, where in the quantifier case we need a quantifier in the definition.

For the rest of this section, fix a countable formal language $\mathcal{L}$; we do not mention the dependence on $\mathcal{L}$ in the notation. Since we deal with classical logic, we only consider formulas built without $\vee, \exists$.

DEFINITION. A *model* is a triple $\mathcal{M} = (D, I_0, I_1)$ such that

(a) D is a nonempty set;
(b) for every n-ary function symbol f, I_0 assigns to f a map $I_0(f)\colon D^n \to D$;
(c) for every n-ary relation symbol R, I_1 assigns to R an n-ary relation on D^n. In case $n = 0$, $I_1(R)$ is either true or false. We require that $I_1(\bot)$ is false.

We write $|\mathcal{M}|$ for the carrier set D of $\mathcal{M}$ and $f^{\mathcal{M}}$, $R^{\mathcal{M}}$ for the interpretations $I_0(f)$, $I_1(R)$ of the function and relation symbols. *Assignments* η and their homomorphic extensions are defined as in 1.3.1. Again we write $t^{\mathcal{M}}[\eta]$ for $\eta(t)$.

DEFINITION (Validity). For every model $\mathcal{M}$, assignment η in $|\mathcal{M}|$ and formula A such that $\mathrm{FV}(A) \subseteq \mathrm{dom}(\eta)$ we define $\mathcal{M} \models A[\eta]$ (read: A is *valid* in $\mathcal{M}$ under the assignment η) by induction on A.

$$\begin{aligned}
\mathcal{M} \models (R\vec{s})[\eta] &:= R^{\mathcal{M}}(\vec{s}^{\mathcal{M}}[\eta]),\\
\mathcal{M} \models (A \to B)[\eta] &:= ((\mathcal{M} \models A[\eta]) \to (\mathcal{M} \models B[\eta])),\\
\mathcal{M} \models (A \wedge B)[\eta] &:= ((\mathcal{M} \models A[\eta]) \wedge (\mathcal{M} \models B[\eta])),\\
\mathcal{M} \models (\forall_x A)[\eta] &:= \forall_{a \in |\mathcal{M}|}(\mathcal{M} \models A[\eta^a_x]).
\end{aligned}$$

Since $I_1(\bot)$ is false, we have $\mathcal{M} \not\models \bot[\eta]$.

1.4.2. Soundness of classical logic.

LEMMA (Coincidence). *Let $\mathcal{M}$ be a model, t a term, A a formula and η, ξ assignments in $|\mathcal{M}|$.*

(a) *If $\eta(x) = \xi(x)$ for all $x \in \mathrm{vars}(t)$, then $\eta(t) = \xi(t)$.*
(b) *If $\eta(x) = \xi(x)$ for all $x \in \mathrm{FV}(A)$, then $\mathcal{M} \models A[\eta]$ if and only if $\mathcal{M} \models A[\xi]$.*

PROOF. Induction on terms and formulas. □

LEMMA (Substitution). *Let $\mathcal{M}$ be a model, $t, r(x)$ terms, $A(x)$ a formula and η an assignment in $|\mathcal{M}|$. Then*

(a) $\eta(r(t)) = \eta^{\eta(t)}_x(r(x))$.
(b) $\mathcal{M} \models A(t)$ *if and only if* $\mathcal{M} \models A(x)[\eta^{\eta(t)}_x]$.

PROOF. Induction on terms and formulas. □

A model $\mathcal{M}$ is called *classical* if $\neg\neg R^{\mathcal{M}}(\vec{a}) \to R^{\mathcal{M}}(\vec{a})$ for all relation symbols R and all $\vec{a} \in |\mathcal{M}|$. We prove that every formula derivable in classical logic is valid in an arbitrary classical model.

THEOREM (Soundness of classical logic). *Let $\Gamma \cup \{A\}$ be a set of formulas such that $\Gamma \vdash_c A$. Then, if $\mathcal{M}$ is a classical model and η an assignment in $|\mathcal{M}|$, it follows that $\mathcal{M} \models \Gamma[\eta]$ implies $\mathcal{M} \models A[\eta]$.*

PROOF. Induction on derivations. We begin with the axioms in Stab and the axiom schemes $\wedge^+$, $\wedge^-$. $\mathcal{M} \models C[\eta]$ is abbreviated $\mathcal{M} \models C$ when η is known from the context.

For the stability axiom $\forall_{\vec{x}}(\neg\neg R\vec{x} \to R\vec{x})$ the claim follows from our assumption that $\mathcal{M}$ is classical, i.e., $\neg\neg R^{\mathcal{M}}(\vec{a}) \to R^{\mathcal{M}}(\vec{a})$ for all $\vec{a} \in |\mathcal{M}|$. The axioms $\wedge^+$, $\wedge^-$ are clearly valid.

This concludes the treatment of the axioms. We now consider the rules. In case of the assumption rule $u\colon A$ we have $A \in \Gamma$ and the claim is obvious.

Case $\to^+$. Assume $\mathcal{M} \models \Gamma$. We show $\mathcal{M} \models (A \to B)$. So assume in addition $\mathcal{M} \models A$. We must show $\mathcal{M} \models B$. By induction hypothesis (with $\Gamma \cup \{A\}$ instead of Γ) this clearly holds.

Case $\to^-$. Assume $\mathcal{M} \models \Gamma$. We must show $\mathcal{M} \models B$. By induction hypothesis, $\mathcal{M} \models (A \to B)$ and $\mathcal{M} \models A$. The claim follows from the definition of $\models$.

Case $\forall^+$. Assume $\mathcal{M} \models \Gamma[\eta]$ and $x \notin \mathrm{FV}(\Gamma)$. We show $\mathcal{M} \models (\forall_x A)[\eta]$, i.e., $\mathcal{M} \models A[\eta^a_x]$ for an arbitrary $a \in |\mathcal{M}|$. We have

$$\begin{array}{ll} \mathcal{M} \models \Gamma[\eta^a_x] & \text{by the coincidence lemma, since } x \notin \mathrm{FV}(\Gamma) \\ \mathcal{M} \models A[\eta^a_x] & \text{by induction hypothesis.} \end{array}$$

Case $\forall^-$. Let $\mathcal{M} \models \Gamma[\eta]$. We show that $\mathcal{M} \models A(t)[\eta]$. This follows from

$$\begin{array}{ll} \mathcal{M} \models (\forall_x A(x))[\eta] & \text{by induction hypothesis} \\ \mathcal{M} \models A(x)[\eta^{\eta(t)}_x] & \text{by definition} \\ \mathcal{M} \models A(t)[\eta] & \text{by the substitution lemma.} \end{array}$$

This concludes the proof. □

1.4.3. Completeness of classical logic. We give a constructive analysis of the completeness of classical logic by using, in the metatheory below, constructively valid arguments only, mentioning explicitly any assumptions which go beyond. When dealing with the classical fragment we of course need to restrict to classical models. The only non-constructive principle

will be the use of the *axiom of dependent choice* for the weak existential quantifier

$$\tilde{\exists}_x A(0,x) \to \forall_{n,x}(A(n,x) \to \tilde{\exists}_y A(n+1,y)) \to \tilde{\exists}_f \forall_n A(n,fn).$$

Recall that we only consider formulas without $\vee, \exists$.

THEOREM (Completeness of classical logic). *Let $\Gamma \cup \{A\}$ be a set of formulas. Assume that for all classical models $\mathcal{M}$ and assignments η,*

$$\mathcal{M} \models \Gamma[\eta] \to \mathcal{M} \models A[\eta].$$

Then there must exist a derivation of A from $\Gamma \cup \mathrm{Stab}$.

PROOF. Since "there must exist a derivation" expresses the weak existential quantifier in the metalanguage, we need to prove a contradiction from the assumption $\Gamma, \mathrm{Stab} \nvdash A$.

By the completeness theorem for minimal logic, there must be a tree model $\mathcal{T} = (\mathrm{Ter}, I_0, I_1)$ on the complete binary tree T_{01} and a node l_0 such that $l_0 \Vdash \Gamma, \mathrm{Stab}$ and $l_0 \nVdash A$.

Call a node k *consistent* if $k \nVdash \bot$, and *stable* if $k \Vdash \mathrm{Stab}$. We prove

$$k \nVdash B \to \tilde{\exists}_{k' \succeq k}(k' \Vdash \neg B \wedge k' \nVdash \bot) \qquad (k \text{ stable}). \tag{1.9}$$

Let k be a stable node, and B a formula (without $\vee$, $\exists$). Then $\mathrm{Stab} \vdash \neg\neg B \to B$ by the stability theorem, and therefore $k \Vdash \neg\neg B \to B$. Hence from $k \nVdash B$ we obtain $k \nVdash \neg\neg B$. By a remark in 1.3.4 this implies that $\neg\forall_{k' \succeq k}(k' \Vdash \neg B \to k' \Vdash \bot)$, which proves (1.9).

Let α be a branch in the underlying tree T_{01}. We define

$$\begin{aligned} \alpha \Vdash A \quad &:= \tilde{\exists}_{k \in \alpha}(k \Vdash A), \\ \alpha \text{ is consistent} &:= \alpha \nVdash \bot, \\ \alpha \text{ is stable} \quad &:= \tilde{\exists}_{k \in \alpha}(k \Vdash \mathrm{Stab}). \end{aligned}$$

Note that from $\alpha \Vdash \vec{A}$ and $\vdash \vec{A} \to B$ it follows that $\alpha \Vdash B$. To see this, consider $\alpha \Vdash \vec{A}$. Then $k \Vdash \vec{A}$ for a $k \in \alpha$, since α is linearly ordered. From $\vdash \vec{A} \to B$ it follows that $k \Vdash B$, i.e., $\alpha \Vdash B$.

A branch α is *generic* (in the sense that it generates a classical model) if it is consistent and stable, if in addition for all formulas B

$$(\alpha \Vdash B) \,\tilde{\vee}\, (\alpha \Vdash \neg B), \tag{1.10}$$

and if for all formulas $\forall_{\vec{y}} B(\vec{y})$ with $B(\vec{y})$ not a universal formula,

$$\forall_{\vec{s} \in \mathrm{Ter}}(\alpha \Vdash B(\vec{s})) \to \alpha \Vdash \forall_{\vec{y}} B(\vec{y}). \tag{1.11}$$

For a branch α, we define a classical model $\mathcal{M}^\alpha = (\mathrm{Ter}, I_0, I_1^\alpha)$ as

$$I_1^\alpha(R)(\vec{s}) := \tilde{\exists}_{k \in \alpha} I_1(R,k)(\vec{s}) \qquad (R \neq \bot).$$

Since $\tilde{\exists}$ is used in this definition, $\mathcal{M}^\alpha$ is stable.

We show that for every generic branch α and formula B (without $\vee$, $\exists$)

$$\alpha \Vdash B \leftrightarrow \mathcal{M}^\alpha \models B. \tag{1.12}$$

The proof is by induction on the logical complexity of B.

Case $R\vec{s}$ with $R \neq \bot$. Then (1.12) holds for all α.

Case $\bot$. We have $\alpha \nVdash \bot$ since α is consistent.

Case $B \to C$. Let $\alpha \Vdash B \to C$ and $\mathcal{M}^\alpha \models B$. We must show that $\mathcal{M}^\alpha \models C$. Note that $\alpha \Vdash B$ by induction hypothesis, hence $\alpha \Vdash C$, hence $\mathcal{M}^\alpha \models C$ again by induction hypothesis. Conversely let $\mathcal{M}^\alpha \models B \to C$. Clearly $(\mathcal{M}^\alpha \models B) \mathbin{\tilde{\vee}} (\mathcal{M}^\alpha \not\models B)$. If $\mathcal{M}^\alpha \models B$, then $\mathcal{M}^\alpha \models C$. Hence $\alpha \Vdash C$ by induction hypothesis and therefore $\alpha \Vdash B \to C$. If $\mathcal{M}^\alpha \not\models B$ then $\alpha \nVdash B$ by induction hypothesis. Hence $\alpha \Vdash \neg B$ by (1.10) and therefore $\alpha \Vdash B \to C$, since α is stable (and $\vdash (\neg\neg C \to C) \to \bot \to C$). [Note that for this argument to be contructively valid one needs to observe that the formula $\alpha \Vdash B \to C$ is a negation, and therefore we can argue by the case distinction based on $\tilde{\vee}$. This is because, with $P_1 := \mathcal{M}^\alpha \models B$, $P_2 := \mathcal{M}^\alpha \not\models B$ and $Q := \alpha \Vdash B \to C$, the formula $(P_1 \mathbin{\tilde{\vee}} P_2) \to (P_1 \to Q) \to (P_2 \to Q) \to Q$ is derivable in minimal logic.]

Case $B \wedge C$. Easy.

Case $\forall_{\vec{y}} B(\vec{y})$ ($\vec{y}$ not empty) where $B(\vec{y})$ is not a universal formula. The following are equivalent.

$$\begin{aligned}
&\alpha \Vdash \forall_{\vec{y}} B(\vec{y}) \\
&\forall_{\vec{s} \in \mathrm{Ter}}(\alpha \Vdash B(\vec{s})) && \text{by (1.11)} \\
&\forall_{\vec{s} \in \mathrm{Ter}}(\mathcal{M}^\alpha \models B(\vec{s})) && \text{by induction hypothesis} \\
&\mathcal{M}^\alpha \models \forall_{\vec{y}} B(\vec{y}).
\end{aligned}$$

This concludes the proof of (1.12).

Next we show that for every consistent and stable node k there must be a generic branch containing k:

$$k \nVdash \bot \to k \Vdash \mathrm{Stab} \to \tilde{\exists}_\alpha(\alpha \text{ generic} \wedge k \in \alpha). \tag{1.13}$$

For the proof, let $A_0, A_1, \ldots$ enumerate all formulas. We define a sequence $k = k_0 \preceq k_1 \preceq k_2 \ldots$ of consistent stable nodes by dependent choice. Let $k_0 := k$. Assume that k_n is defined. We write A_n in the form $\forall_{\vec{y}} B(\vec{y})$ (with $\vec{y}$ possibly empty) where B is not a universal formula. In case $k_n \Vdash \forall_{\vec{y}} B(\vec{y})$ let $k_{n+1} := k_n$. Otherwise we have $k_n \nVdash B(\vec{s})$ for some $\vec{s}$, and by (1.9) there must be a consistent node $k' \succeq k_n$ such that $k' \Vdash \neg B(\vec{s})$. Let $k_{n+1} := k'$. Since $k_n \preceq k_{n+1}$, the node k_{n+1} is stable.

Let $\alpha := \{\, l \mid \exists_n(l \preceq k_n) \,\}$, hence $k \in \alpha$. We show that α is generic. Clearly α is consistent and stable. We now prove both (1.10) and (1.11).

Let $C = \forall_{\vec{y}} B(\vec{y})$ (with $\vec{y}$ possibly empty) where $B(\vec{y})$ is not a universal formula, and choose n such that $C = A_n$. In case $k_n \Vdash \forall_{\vec{y}} B(\vec{y})$ we are done. Otherwise by construction $k_{n+1} \Vdash \neg B(\vec{s})$ for some $\vec{s}$. For (1.10) we get $k_{n+1} \Vdash \neg\forall_{\vec{y}} B(\vec{y})$ since $\vdash \forall_{\vec{y}} B(\vec{y}) \to B(\vec{s})$, and (1.11) follows from the consistency of α. This concludes the proof of (1.13).

Now we can finalize the completeness proof. Recall that $l_0 \Vdash \Gamma, \mathrm{Stab}$ and $l_0 \nVdash A$. Since $l_0 \nVdash A$ and l_0 is stable, (1.9) yields a consistent node $k \succeq l_0$ such that $k \Vdash \neg A$. Evidently, k is stable as well. By (1.13) there must be a generic branch α such that $k \in \alpha$. Since $k \Vdash \neg A$ it follows that $\alpha \Vdash \neg A$, hence $\mathcal{M}^\alpha \models \neg A$ by (1.12). Moreover, $\alpha \Vdash \Gamma$, thus $\mathcal{M}^\alpha \models \Gamma$ by (1.12). This contradicts our assumption. □

1.4.4. Compactness and Löwenheim-Skolem theorems. Among the many important corollaries of the completeness theorem the compactness and Löwenheim-Skolem theorems stand out as particularly important. A set Γ of formulas is *consistent* if $\Gamma \nvdash_c \bot$, and *satisfiable* if there is (in the weak sense) a classical model $\mathcal{M}$ and an assignment η in $|\mathcal{M}|$ such that $\mathcal{M} \models \Gamma[\eta]$.

Corollary. *Let Γ be a set of formulas.*

(a) *If Γ is consistent, then Γ is satisfiable.*
(b) *(Compactness). If each finite subset of Γ is satisfiable, Γ is satisfiable.*

Proof. (a). Assume $\Gamma \nvdash_c \bot$ and that for all classical models $\mathcal{M}$ we have $\mathcal{M} \not\models \Gamma$, i.e., $\mathcal{M} \models \Gamma$ implies $\mathcal{M} \models \bot$. Then the completeness theorem yields a contradiction.

(b). Otherwise by the completeness theorem there must be a derivation of $\bot$ from $\Gamma \cup \mathrm{Stab}$, hence also from $\Gamma_0 \cup \mathrm{Stab}$ for some finite subset $\Gamma_0 \subseteq \Gamma$. This contradicts the assumption that Γ_0 is satisfiable. □

Corollary (Löwenheim and Skolem). *Let Γ be a set of formulas (we assume that $\mathcal{L}$ is countable). If Γ is satisfiable, then Γ is satisfiable in a model with a countably infinite carrier set.*

Proof. Assume that Γ is not satisfiable in a countable model. Then by the completeness theorem $\Gamma \cup \mathrm{Stab} \vdash \bot$. Therefore by the soundness theorem Γ cannot be satisfiable. □

CHAPTER 2

Model Theory

Model theory is an established branch of mathematical logic. It uses tools from logic to study questions in algebra. In model theory it is common to disregard the distinction between strong and weak existential quantifiers; we shall do the same in the present chapter. Also, the restriction to countable languages that we have maintained until now is given up. Moreover one makes free use of other concepts and axioms from set theory like the *axiom of choice* (for the weak existential quantifier), most often in the form of *Zorn's lemma*.

2.1. Ultraproducts

2.1.1. Filters and ultrafilters. Let $M \neq \emptyset$ be a set. $F \subseteq \mathcal{P}(M)$ is called *filter* on M if

(a) $M \in F$ and $\emptyset \notin F$;
(b) if $X \in F$ and $X \subseteq Y \subseteq M$, then $Y \in F$;
(c) $X, Y \in F$ entails $X \cap Y \in F$.

F is called *ultrafilter* if for all $X \in \mathcal{P}(M)$

$$X \in F \text{ or } M \setminus X \in F.$$

The intuition here is that the elements X of a filter F are considered to be "big". For instance, for M infinite the set $F = \{ X \subseteq M \mid M \setminus X \text{ finite} \}$ is a filter (called *Fréchet*-filter).

LEMMA. *Suppose F is an ultrafilter and $X \cup Y \in F$. Then $X \in F$ or $Y \in F$.*

PROOF. If both X and Y are not in F, then $M \setminus X$ and $M \setminus Y$ are in F, hence also $(M \setminus X) \cap (M \setminus Y)$, which is $M \setminus (X \cup Y)$. This contradicts the assumption $X \cup Y \in F$. □

Let $M \neq \emptyset$ be a set and $S \subseteq \mathcal{P}(M)$. S has the *finite intersection property* if $X_1 \cap \cdots \cap X_n \neq \emptyset$ for all $X_1, \ldots, X_n \in S$ and all $n \in \mathbb{N}$.

LEMMA. *If S has the finite intersection property, then there exists a filter F on M such that $F \supseteq S$.*

PROOF. $F := \{\, X \mid X \supseteq X_1 \cap \cdots \cap X_n \text{ for some } X_1, \ldots, X_n \in S \,\}$. □

THEOREM (Ultrafilter). *Let $M \neq \emptyset$ be a set and F a filter on M. Then there is an ultrafilter U on M such that $U \supseteq F$.*

PROOF. By Zorn's lemma (which will be proved from the axiom of choice later, in the chapter on set theory), there is a maximal filter U with $F \subseteq U$. We claim that U is an ultrafilter. So let $X \subseteq M$ and assume $X \notin U$ and $M \setminus X \notin U$. Since U is maximal, $U \cup \{X\}$ cannot have the finite intersection property; hence there is a $Y \in U$ such that $Y \cap X = \emptyset$. Similary we obtain $Z \in U$ such that $Z \cap (M \setminus X) = \emptyset$. But then $Y \cap Z = \emptyset$, a contradiction. □

2.1.2. Products and ultraproducts. Let $I \neq \emptyset$ be a set and $D_i \neq \emptyset$ sets for $i \in I$. Let

$$\prod_{i \in I} D_i := \{\, \alpha \mid \alpha \text{ is a function, } \mathrm{dom}(\alpha) = I \text{ and } \alpha(i) \in D_i \text{ for all } i \in I \,\}.$$

Observe that, by the *axiom of choice*, $\prod_{i\in I} D_i \neq \emptyset$. We write $\alpha \in \prod_{i\in I} D_i$ as $\langle\, \alpha(i) \mid i \in I \,\rangle$.

Now let $I \neq \emptyset$ be a set, F a filter on I and $\mathcal{M}_i$ models for $i \in I$. Then the *F-product* $\mathcal{M} = \prod_{i\in I}^F \mathcal{M}_i$ is defined by

(a) $|\mathcal{M}| := \prod_{i\in I} |\mathcal{M}_i|$ (notice that $|\mathcal{M}| \neq \emptyset$).

(b) for an n-ary relation symbol R and $\alpha_1, \ldots, \alpha_n \in |\mathcal{M}|$ let

$$R^{\mathcal{M}}(\alpha_1, \ldots, \alpha_n) := (\{\, i \in I \mid R^{\mathcal{M}_i}(\alpha_1(i), \ldots, \alpha_n(i)) \,\} \in F).$$

(c) for an n-ary function symbol f and $\alpha_1, \ldots, \alpha_n \in |\mathcal{M}|$ let

$$f^{\mathcal{M}}(\alpha_1, \ldots, \alpha_n) := \langle\, f^{\mathcal{M}_i}(\alpha_1(i), \ldots, \alpha_n(i)) \mid i \in I \,\rangle.$$

For an ultrafilter U we call $\mathcal{M} = \prod_{i\in I}^U \mathcal{M}_i$ the *U-ultraproduct* of the $\mathcal{M}_i$.

THEOREM (Fundamental theorem on ultraproducts, Łoś (1955)). *Let $\mathcal{M} = \prod_{i\in I}^U \mathcal{M}_i$ be a U-ultraproduct, A a formula and η an assignment in $|\mathcal{M}|$. Then*

$$\mathcal{M} \models A[\eta] \leftrightarrow \{\, i \in I \mid \mathcal{M}_i \models A[\eta_i] \,\} \in U,$$

where η_i is the assignment induced by $\eta_i(x) = \eta(x)(i)$ for $i \in I$.

PROOF. We first prove a similar property for terms.

$$t^{\mathcal{M}}[\eta] = \langle\, t^{\mathcal{M}_i}[\eta_i] \mid i \in I \,\rangle. \tag{2.1}$$

The proof is by induction on t. For a variable the claim follows from the definition. *Case* $f(t_1, \ldots, t_n)$. For simplicity assume $n = 1$; so we consider ft. We obtain

$$\begin{aligned}(ft)^{\mathcal{M}}[\eta] &= f^{\mathcal{M}}(t^{\mathcal{M}}[\eta]) \\ &= f^{\mathcal{M}}\langle\, t^{\mathcal{M}_i}[\eta_i] \mid i \in I \,\rangle \quad \text{by induction hypothesis}\end{aligned}$$

$$= \langle\, (ft)^{\mathcal{M}_i}[\eta_i] \mid i \in I \,\rangle.$$

Case $R(t_1, \ldots, t_n)$. For simplicity assume $n = 1$; so consider Rt. We obtain

$$\begin{aligned}
\mathcal{M} \models Rt[\eta] &\leftrightarrow R^{\mathcal{M}}(t^{\mathcal{M}}[\eta]) \\
&\leftrightarrow \{\, i \in I \mid R^{\mathcal{M}_i}(t^{\mathcal{M}}[\eta](i)) \,\} \in U \\
&\leftrightarrow \{\, i \in I \mid R^{\mathcal{M}_i}(t^{\mathcal{M}_i}[\eta_i]) \,\} \in U \qquad \text{by (2.1)} \\
&\leftrightarrow \{\, i \in I \mid \mathcal{M}_i \models Rt[\eta_i] \,\} \in U.
\end{aligned}$$

Case $A \to B$.

$\mathcal{M} \models (A \to B)[\eta]$
$\leftrightarrow$ if $\mathcal{M} \models A[\eta]$, then $\mathcal{M} \models B[\eta]$
$\leftrightarrow$ if $\{\, i \in I \mid \mathcal{M}_i \models A[\eta_i] \,\} \in U$, then $\{\, i \in I \mid \mathcal{M}_i \models B[\eta_i] \,\} \in U$
by induction hypothesis
$\leftrightarrow \{\, i \in I \mid \mathcal{M}_i \models A[\eta_i] \,\} \notin U$ or $\{\, i \in I \mid \mathcal{M}_i \models B[\eta_i] \,\} \in U$
$\leftrightarrow \{\, i \in I \mid \mathcal{M}_i \models \neg A[\eta_i] \,\} \in U$ or $\{\, i \in I \mid \mathcal{M}_i \models B[\eta_i] \,\} \in U$
for U is an ultrafilter
$\leftrightarrow \{\, i \in I \mid \mathcal{M}_i \models (A \to B)[\eta_i] \,\} \in U.$

The case $A \wedge B$ is easy.
Case $\forall_x A$.

$\mathcal{M} \models (\forall_x A)[\eta]$
$\leftrightarrow \forall_{\alpha \in |\mathcal{M}|}(\mathcal{M} \models A[\eta_x^\alpha])$
$\leftrightarrow \forall_{\alpha \in |\mathcal{M}|}(\{\, i \in I \mid \mathcal{M}_i \models A[(\eta_i)_x^{\alpha(i)}] \,\} \in U)$ by induction hypothesis
$\leftrightarrow^{(*)} \{\, i \in I \mid \forall_{a \in |\mathcal{M}_i|}(\mathcal{M}_i \models A[(\eta_i)_x^a]) \,\} \in U$ see below
$\leftrightarrow \{\, i \in I \mid \mathcal{M}_i \models (\forall_x A)[\eta_i] \,\} \in U.$

It remains to the equivalence marked $(*)$. Let

$$X := \{\, i \in I \mid \forall_{a \in |\mathcal{M}_i|}(\mathcal{M}_i \models A[(\eta_i)_x^a]) \,\}$$

and $Y_\alpha := \{\, i \in I \mid \mathcal{M}_i \models A[(\eta_i)_x^{\alpha(i)}] \,\}$ for $\alpha \in |\mathcal{M}|$.

$\leftarrow$. Let $\alpha \in |\mathcal{M}|$ and $X \in U$. Clearly $X \subseteq Y_\alpha$, hence also $Y_\alpha \in U$.

$\rightarrow$. Let $Y_\alpha \in U$ for all α. Assume $X \notin U$. Since U is an ultrafilter,

$$I \setminus X = \{\, i \in I \mid \exists_{a \in |\mathcal{M}_i|}(\mathcal{M}_i \not\models A[(\eta_i)_x^a]) \,\} \in U.$$

We choose by the axiom of choice an $\alpha_0 \in |\mathcal{M}|$ such that

$$\alpha_0(i) = \begin{cases} \text{some } a \in |\mathcal{M}_i| \text{ such that } \mathcal{M}_i \not\models A[(\eta_i)_x^a] & \text{if } i \in I \setminus X, \\ \text{an arbitrary} \in |\mathcal{M}_i| & \text{otherwise.} \end{cases}$$

Then $Y_{\alpha_0} \cap (I \setminus X) = \emptyset$, contradicting $Y_{\alpha_0}, I \setminus X \in U$. □

If we choose $\mathcal{M}_i = \mathcal{N}$ constant, then $\mathcal{M} = \prod_{i\in I}^U \mathcal{N}$ satisfies the same closed formulas as $\mathcal{N}$ (such models will be called *elementary equivalent*; the notation is $\mathcal{M} \equiv \mathcal{N}$). $\prod_{i\in I}^U \mathcal{N}$ is called an *ultrapower* of $\mathcal{N}$.

2.1.3. General compactness and completeness. Recall that the underlying language may be uncountable.

Corollary (General compactness theorem). *Let Γ be a set of formulas. If every finite subset of Γ is satisfiable, then so is Γ.*

Proof. Let $I := \{\, i \subseteq \Gamma \mid i \text{ finite} \,\}$. For $i \in I$ let $\mathcal{M}_i$ be a model of i under the assignment η_i. For $A \in \Gamma$ let $Z_A := \{\, i \in I \mid A \in i \,\} = \{\, i \subseteq \Gamma \mid i \text{ finite and } A \in i \,\}$. Then $F := \{\, Z_A \mid A \in \Gamma \,\}$ has the finite intersection property (for $\{A_1, \dots, A_n\} \in Z_{A_1} \cap \dots \cap Z_{A_n}$). By the Ultrafilter Theorem in 2.1.1 there is an ultrafilter U on I such that $F \subseteq U$. We consider the ultraproduct $\mathcal{M} := \prod_{i\in I}^U \mathcal{M}_i$ and the product assigment η defined by $\eta(x)(i) := \eta_i(x)$, and show $\mathcal{M} \models \Gamma[\eta]$. So let $A \in \Gamma$. By Łoś's theorem it suffices to show

$$X_A := \{\, i \in I \mid \mathcal{M}_i \models A[\eta_i] \,\} \in U.$$

But this follows from $Z_A \subseteq X_A$ and $Z_A \in F \subseteq U$. □

For every set Γ of formulas let $L(\Gamma)$ be the set of all function and relation symbols occurring in Γ. If $\mathcal{L}'$ is a sublanguage of $\mathcal{L}$, $\mathcal{M}'$ an $\mathcal{L}'$-model and $\mathcal{M}$ an $\mathcal{L}$-model, then $\mathcal{M}$ is called an *expansion* of $\mathcal{M}'$ (and $\mathcal{M}'$ a *reduct* of $\mathcal{M}$) if $|\mathcal{M}'| = |\mathcal{M}|$, $f^{\mathcal{M}'} = f^{\mathcal{M}}$ for all function symbols and $R^{\mathcal{M}'} = R^{\mathcal{M}}$ for all relation symbols in the language $\mathcal{L}'$. The (uniquely determined) $\mathcal{L}'$-reduct of $\mathcal{M}$ is denoted by $\mathcal{M}{\restriction}\mathcal{L}'$. If $\mathcal{M}$ is an expansion of $\mathcal{M}'$ and η an assignment in $|\mathcal{M}'|$, then clearly $t^{\mathcal{M}'}[\eta] = t^{\mathcal{M}}[\eta]$ for every $\mathcal{L}'$-term t and $\mathcal{M}' \models A[\eta]$ if and only if $\mathcal{M} \models A[\eta]$, for every $\mathcal{L}'$-formula A.

Corollary (General completeness theorem). *Let $\Gamma \cup \{A\}$ be a set of formulas. Assume that for all models $\mathcal{M}$ and assignments η,*

$$\mathcal{M} \models \Gamma[\eta] \to \mathcal{M} \models A[\eta].$$

Then $\Gamma \vdash_c A$.

Proof. By assumption $\Gamma \cup \{\neg A\}$ is not satisfiable. Hence by the general compactness theorem there is a finite subset $\Gamma' \subseteq \Gamma$ such that already $\Gamma' \cup \{\neg A\}$ is not satisfiable. Let $\mathcal{L}$ be the underlying (possibly uncountable) language, and $\mathcal{L}'$ the countable sublanguage containing only function and relation symbols from Γ'. By the remark above $\Gamma' \cup \{\neg A\}$ is not satisfiable w.r.t. $\mathcal{L}'$ as well. By the completeness theorem for countable languages we obtain $\Gamma' \vdash_c A$, hence $\Gamma \vdash_c A$. □

2.2. Complete Theories and Elementary Equivalence

We assume in this section that our underlying language $\mathcal{L}$ contains a binary relation symbol $=$.

2.2.1. Equality axioms. The set $\mathrm{Eq}_\mathcal{L}$ of $\mathcal{L}$-equality axioms consists of (the universal closures of)

$$x = x \quad \text{(reflexivity)},$$
$$x = y \to y = x \quad \text{(symmetry)},$$
$$x = y \to y = z \to x = z \quad \text{(transitivity)},$$
$$x_1 = y_1 \to \cdots \to x_n = y_n \to f(x_1, \ldots, x_n) = f(y_1, \ldots, y_n),$$
$$x_1 = y_1 \to \cdots \to x_n = y_n \to R(x_1, \ldots, x_n) \to R(y_1, \ldots, y_n),$$

for all n-ary function symbols f and relation symbols R of the language $\mathcal{L}$.

Lemma (Equality). (a) $\mathrm{Eq}_\mathcal{L} \vdash t = s \to r(t) = r(s)$.
(b) $\mathrm{Eq}_\mathcal{L} \vdash t = s \to (A(t) \leftrightarrow A(s))$.

Proof. (a). Induction on r. (b). Induction on A. □

An $\mathcal{L}$-model $\mathcal{M}$ satisfies the equality axioms if and only if $=^\mathcal{M}$ is a *congruence relation* (i.e., an equivalence relation compatible with the functions and relations of $\mathcal{M}$). In this section we assume that all $\mathcal{L}$-models $\mathcal{M}$ considered satisfy the equality axioms. The coincidence lemma then also holds with $=^\mathcal{M}$ instead of $=$:

Lemma (Coincidence). *Let η and ξ be assignments in $|\mathcal{M}|$ such that $\mathrm{dom}(\eta) = \mathrm{dom}(\xi)$ and $\eta(x) =^\mathcal{M} \xi(x)$ for all $x \in \mathrm{dom}(\eta)$. Then*
(a) *$t^\mathcal{M}[\eta] =^\mathcal{M} t^\mathcal{M}[\xi]$ if $\mathrm{vars}(t) \subseteq \mathrm{dom}(\eta)$ and*
(b) *$\mathcal{M} \models A[\eta] \leftrightarrow \mathcal{M} \models A[\xi]$ if $\mathrm{FV}(A) \subseteq \mathrm{dom}(\eta)$.*

Proof. Induction on t and A, respectively. □

2.2.2. Cardinality of models. Let $\mathcal{M}/=^\mathcal{M}$ be the *quotient model*, whose carrier set consists of congruence classes. We call a model $\mathcal{M}$ *infinite* (*countable*, *of cardinality n*) if $|\mathcal{M}/=^\mathcal{M}|$ is infinite (countable, of cardinality n). By an *axiom system* Γ we mean a set of closed formulas such that $\mathrm{Eq}_{L(\Gamma)} \subseteq \Gamma$. A *model* of an axiom system Γ is an $\mathcal{L}$-model $\mathcal{M}$ such that $L(\Gamma) \subseteq \mathcal{L}$ and $\mathcal{M} \models \Gamma$. For sets Γ of closed formulas we write

$$\mathrm{Mod}_\mathcal{L}(\Gamma) := \{\, \mathcal{M} \mid \mathcal{M} \text{ is an } \mathcal{L}\text{-model and } \mathcal{M} \models \Gamma \cup \mathrm{Eq}_\mathcal{L} \,\}.$$

Clearly Γ is satisfiable if and only if Γ has an $\mathcal{L}$-model.

Theorem. *If an axiom system has arbitrarily large finite models, then it has an infinite model.*

PROOF. Let Γ be such an axiom system. Suppose $x_0, x_1, x_2, \dots$ are distinct variables and

$$\Gamma' := \Gamma \cup \{\, x_i \neq x_j \mid i, j \in \mathbb{N} \text{ such that } i < j \,\}.$$

By assumption every finite subset of Γ' is satisfiable, hence by the general compactness theorem so is Γ'. Then we have $\mathcal{M}$ and η such that $\mathcal{M} \models \Gamma'[\eta]$ and therefore $\eta(x_i) \neq^{\mathcal{M}} \eta(x_j)$ for $i < j$. Hence $\mathcal{M}$ is infinite. □

2.2.3. Complete theories, elementary equivalence. Let $\overline{\mathcal{L}}$ be the set of all closed $\mathcal{L}$-formulas. By a *theory* T we mean an axiom system closed under $\vdash_c$, that is, $\mathrm{Eq}_{L(T)} \subseteq T$ and

$$T = \{\, A \in \overline{L(T)} \mid T \vdash_c A \,\}.$$

A theory T is called *complete* if for every formula $A \in \overline{L(T)}$, $T \vdash_c A$ or $T \vdash_c \neg A$.

For every $\mathcal{L}$-model $\mathcal{M}$ (satisfying the equality axioms) the set of all closed $\mathcal{L}$-formulas A such that $\mathcal{M} \models A$ clearly is a theory; it is called the *theory of* $\mathcal{M}$ and denoted by $\mathrm{Th}(\mathcal{M})$.

Two $\mathcal{L}$-models $\mathcal{M}$ and $\mathcal{M}'$ are called *elementarily equivalent* (written $\mathcal{M} \equiv \mathcal{M}'$) if $\mathrm{Th}(\mathcal{M}) = \mathrm{Th}(\mathcal{M}')$. Two $\mathcal{L}$-models $\mathcal{M}$ and $\mathcal{M}'$ are called *isomorphic* (written $\mathcal{M} \cong \mathcal{M}'$) if there is a map $\pi\colon |\mathcal{M}| \to |\mathcal{M}'|$ inducing a bijection between $|\mathcal{M}/{=^{\mathcal{M}}}|$ and $|\mathcal{M}'/{=^{\mathcal{M}'}}|$, that is,

$$\forall_{a,b \in |\mathcal{M}|}(a =^{\mathcal{M}} b \leftrightarrow \pi(a) =^{\mathcal{M}'} \pi(b)),$$
$$\forall_{a' \in |\mathcal{M}'|} \exists_{a \in |\mathcal{M}|}(\pi(a) =^{\mathcal{M}'} a'),$$

such that for all $a_1, \dots, a_n \in |\mathcal{M}|$

$$\pi(f^{\mathcal{M}}(a_1, \dots, a_n)) =^{\mathcal{M}'} f^{\mathcal{M}'}(\pi(a_1), \dots, \pi(a_n)),$$
$$R^{\mathcal{M}}(a_1, \dots, a_n) \leftrightarrow R^{\mathcal{M}'}(\pi(a_1), \dots, \pi(a_n))$$

for all n-ary function symbols f and relation symbols R of the language $\mathcal{L}$.

We collect some simple properties of the notions of the theory of a model $\mathcal{M}$ and of elementary equivalence.

LEMMA. (a) $\mathrm{Th}(\mathcal{M})$ *is complete.*

(b) *If* Γ *is an axiom system such that* $L(\Gamma) \subseteq \mathcal{L}$*, then*

$$\{\, A \in \overline{\mathcal{L}} \mid \Gamma \cup \mathrm{Eq}_{\mathcal{L}} \vdash_c A \,\} = \bigcap \{\, \mathrm{Th}(\mathcal{M}) \mid \mathcal{M} \in \mathrm{Mod}_{\mathcal{L}}(\Gamma) \,\}.$$

(c) $\mathcal{M} \equiv \mathcal{M}' \leftrightarrow \mathcal{M} \models \mathrm{Th}(\mathcal{M}')$.

(d) *If* $\mathcal{L}$ *is countable, then for every* $\mathcal{L}$*-model* $\mathcal{M}$ *there is a countable* $\mathcal{L}$*-model* $\mathcal{M}'$ *such that* $\mathcal{M} \equiv \mathcal{M}'$.

PROOF. (a). Let $\mathcal{M}$ be an $\mathcal{L}$-model and $A \in \overline{\mathcal{L}}$. Then $\mathcal{M} \models A$ or $\mathcal{M} \models \neg A$, hence $\mathrm{Th}(\mathcal{M}) \vdash_c A$ or $\mathrm{Th}(\mathcal{M}) \vdash_c \neg A$.

(b). For all $A \in \overline{\mathcal{L}}$ we have

$$\begin{aligned}\Gamma \cup \mathrm{Eq}_{\mathcal{L}} \vdash_c A &\leftrightarrow \text{for all } \mathcal{L}\text{-models } \mathcal{M},\ (\mathcal{M} \models \Gamma \to \mathcal{M} \models A)\\ &\leftrightarrow \text{for all } \mathcal{L}\text{-models } \mathcal{M},\ (\mathcal{M} \in \mathrm{Mod}_{\mathcal{L}}(\Gamma) \to A \in \mathrm{Th}(\mathcal{M}))\\ &\leftrightarrow A \in \bigcap\{\, \mathrm{Th}(\mathcal{M}) \mid \mathcal{M} \in \mathrm{Mod}_{\mathcal{L}}(\Gamma)\,\}.\end{aligned}$$

(c). For $\to$ assume $\mathcal{M} \equiv \mathcal{M}'$ and $A \in \mathrm{Th}(\mathcal{M}')$. Then $\mathcal{M}' \models A$, hence $\mathcal{M} \models A$. For $\leftarrow$ assume $\mathcal{M} \models \mathrm{Th}(\mathcal{M}')$. Then clearly $\mathrm{Th}(\mathcal{M}') \subseteq \mathrm{Th}(\mathcal{M})$. For the converse inclusion let $A \in \mathrm{Th}(\mathcal{M})$. If $A \notin \mathrm{Th}(\mathcal{M}')$, then $\neg A \in \mathrm{Th}(\mathcal{M}')$ by (a) and hence $\mathcal{M} \models \neg A$, contradicting $A \in \mathrm{Th}(\mathcal{M})$.

(d). Let $\mathcal{L}$ be countable and $\mathcal{M}$ an $\mathcal{L}$-model. Then $\mathrm{Th}(\mathcal{M})$ is satisfiable and therefore by the theorem of Löwenheim and Skolem possesses a satisfying $\mathcal{L}$-model $\mathcal{M}'$ with the countable carrier set $\mathrm{Ter}_{\mathcal{L}}$. By (c), $\mathcal{M} \equiv \mathcal{M}'$. □

Moreover, we can characterize complete theories as follows:

THEOREM. *Let T be a theory and $\mathcal{L} = L(T)$. Then the following are equivalent.*

(a) *T is complete.*
(b) *For every model $\mathcal{M} \in \mathrm{Mod}_{\mathcal{L}}(T)$, $\mathrm{Th}(\mathcal{M}) = T$.*
(c) *Any two models $\mathcal{M}, \mathcal{M}' \in \mathrm{Mod}_{\mathcal{L}}(T)$ are elementarily equivalent.*

PROOF. (a) $\to$ (b). Let T be complete and $\mathcal{M} \in \mathrm{Mod}_{\mathcal{L}}(T)$. Then $\mathcal{M} \models T$, hence $T \subseteq \mathrm{Th}(\mathcal{M})$. For the converse assume $A \in \mathrm{Th}(\mathcal{M})$. Then $\neg A \notin \mathrm{Th}(\mathcal{M})$, hence $\neg A \notin T$ and therefore $A \in T$.

(b) $\to$ (c) is clear.

(c) $\to$ (a). Let $A \in \overline{\mathcal{L}}$ and $T \nvdash_c A$. Then there is a model $\mathcal{M}_0$ of $T \cup \{\neg A\}$. Now let $\mathcal{M} \in \mathrm{Mod}_{\mathcal{L}}(T)$ be arbitrary. By (c) we have $\mathcal{M} \equiv \mathcal{M}_0$, hence $\mathcal{M} \models \neg A$. Therefore $T \vdash_c \neg A$. □

2.2.4. Elementary equivalence and isomorphism.

LEMMA. *Let π be an isomorphism between $\mathcal{M}$ and $\mathcal{M}'$. Then for all terms t and formulas A and for every sufficiently big assignment η in $|\mathcal{M}|$*

(a) *$\pi(t^{\mathcal{M}}[\eta]) =^{\mathcal{M}'} t^{\mathcal{M}'}[\pi \circ \eta]$ and*
(b) *$\mathcal{M} \models A[\eta] \leftrightarrow \mathcal{M}' \models A[\pi \circ \eta]$. In particular,*

$$\mathcal{M} \cong \mathcal{M}' \to \mathcal{M} \equiv \mathcal{M}'.$$

PROOF. (a). Induction on t. For simplicity we only consider the case of a unary function symbol.

$$\begin{aligned}\pi(x^{\mathcal{M}}[\eta]) &= \pi(\eta(x)) = x^{\mathcal{M}'}[\pi \circ \eta]\\ \pi((ft)^{\mathcal{M}}[\eta]) &= \pi(f^{\mathcal{M}}(t^{\mathcal{M}}[\eta]))\\ &=^{\mathcal{M}'} f^{\mathcal{M}'}(\pi(t^{\mathcal{M}}[\eta]))\\ &=^{\mathcal{M}'} f^{\mathcal{M}'}(t^{\mathcal{M}'}[\pi \circ \eta])\\ &= (ft)^{\mathcal{M}'}[\pi \circ \eta].\end{aligned}$$

(b). Induction on A. For simplicity we only consider the case of a unary relation symbol P and the case $\forall_x A$.

$$\begin{aligned}\mathcal{M} \models (Pr)[\eta] &\leftrightarrow P^{\mathcal{M}}(r^{\mathcal{M}}[\eta])\\ &\leftrightarrow P^{\mathcal{M}'}(\pi(r^{\mathcal{M}}[\eta]))\\ &\leftrightarrow P^{\mathcal{M}'}(r^{\mathcal{M}'}[\pi \circ \eta])\\ &\leftrightarrow \mathcal{M}' \models (Pr)[\pi \circ \eta],\end{aligned}$$

$$\begin{aligned}\mathcal{M} \models \forall_x A[\eta] &\leftrightarrow \forall_{a \in |\mathcal{M}|}(\mathcal{M} \models A[\eta^a_x])\\ &\leftrightarrow \forall_{a \in |\mathcal{M}|}(\mathcal{M}' \models A[\pi \circ \eta^a_x])\\ &\leftrightarrow \forall_{a \in |\mathcal{M}|}(\mathcal{M}' \models A[(\pi \circ \eta)^{\pi(a)}_x])\\ &\leftrightarrow \forall_{a' \in |\mathcal{M}'|}(\mathcal{M}' \models A[(\pi \circ \eta)^{a'}_x])\\ &\leftrightarrow \mathcal{M}' \models \forall_x A[\pi \circ \eta].\end{aligned}$$

For part "$\rightarrow$" of the next-to-last equivalence we have used the Coincidence Lemma from 2.2. □

The converse, i.e., that $\mathcal{M} \equiv \mathcal{M}'$ implies $\mathcal{M} \cong \mathcal{M}'$, is true for finite models, but not for infinite ones. This proves the impossibility to characterize models by first order axioms.

THEOREM. *For every infinite model $\mathcal{M}$ there is an elementarily equivalent model $\mathcal{M}_0$ not isomorphic to $\mathcal{M}$.*

PROOF. Let $=^{\mathcal{M}}$ be the equality on $D := |\mathcal{M}|$, and let $\mathcal{P}(D)$ denote the power set of D. For every $\alpha \in \mathcal{P}(D)$ choose a new constant c_α. In the language $\mathcal{L}' := \mathcal{L} \cup \{\, c_\alpha \mid \alpha \in \mathcal{P}(D) \,\}$ we consider the axiom system

$$\Gamma := \mathrm{Th}(\mathcal{M}) \cup \{\, c_\alpha \neq c_\beta \mid \alpha, \beta \in \mathcal{P}(D) \text{ and } \alpha \neq \beta \,\} \cup \mathrm{Eq}_{\mathcal{L}'}.$$

Every finite subset of Γ is satisfiable by an appropriate expansion of $\mathcal{M}$. Hence by the general compactness theorem also Γ is satisfiable, say by $\mathcal{M}'_0$. Let $\mathcal{M}_0 := \mathcal{M}'_0 \restriction \mathcal{L}$. We may assume that $=^{\mathcal{M}_0}$ is the equality on $|\mathcal{M}_0|$. $\mathcal{M}_0$

is not isomorphic to $\mathcal{M}$, for otherwise we would have an injection of $\mathcal{P}(D)$ into D and therefore a contradiction. □

2.3. Applications

2.3.1. Non-standard models. By what we just proved it is impossible to characterize an infinite model by a first order axiom system up to isomorphism. However, if we extend first order logic by also allowing quantification over sets X, we can formulate the following *Peano axioms*

$$\forall_n(\mathrm{S}n \neq 0),$$
$$\forall_{n,m}(\mathrm{S}n = \mathrm{S}m \to n = m),$$
$$\forall_X(0 \in X \to \forall_n(n \in X \to \mathrm{S}n \in X) \to \forall_n(n \in X)).$$

One can show easily that $(\mathbb{N}, 0, \mathrm{S})$ is up to isomorphism the unique model of the Peano axioms. A model which is elementarily equivalent, but not isomorphic to $\mathcal{N} := (\mathbb{N}, 0, \mathrm{S})$, is called a *non-standard model* of $\mathcal{N}$. In such non-standard models the principle of complete induction does not hold for all subsets of $|\mathcal{N}|$.

Theorem. *There are countable non-standard models of the natural numbers.*

Proof. Let x be a variable and $\Gamma := \mathrm{Th}(\mathcal{N}) \cup \{\, x \neq \underline{n} \mid n \in \mathbb{N} \,\}$, where $\underline{0} := 0$ and $\underline{n+1} := \mathrm{S}\underline{n}$. Clearly every finite subset of Γ is satisfiable, hence by compactness also Γ. By the theorem of Löwenheim and Skolem we then have a countable or finite $\mathcal{M}$ and an assignment η such that $\mathcal{M} \models \Gamma[\eta]$. Because of $\mathcal{M} \models \mathrm{Th}(\mathcal{N})$ we have $\mathcal{M} \equiv \mathcal{N}$ by 2.2.3; hence $\mathcal{M}$ is countable. Moreover $\eta(x) \neq^{\mathcal{M}} \underline{n}^{\mathcal{M}}$ for all $n \in \mathbb{N}$, hence $\mathcal{M} \ncong \mathcal{N}$. □

2.3.2. Archimedian ordered fields. We now consider some easy applications to well-known axiom systems. The axioms of *field theory* are (the equality axioms and)

$$x + (y + z) = (x + y) + z, \qquad x \cdot (y \cdot z) = (x \cdot y) \cdot z,$$
$$0 + x = x, \qquad 1 \cdot x = x,$$
$$(-x) + x = 0, \qquad x \neq 0 \to x^{-1} \cdot x = 1,$$
$$x + y = y + x, \qquad x \cdot y = y \cdot x,$$

and also

$$(x + y) \cdot z = (x \cdot z) + (y \cdot z),$$
$$1 \neq 0.$$

Fields are the models of this axiom system.

In the theory of ordered fields one has in addition a binary relation symbol $<$ and as axioms

$$\begin{aligned}
&x \not< x,\\
&x < y \to y < z \to x < z,\\
&x < y \vee x = y \vee y < x,\\
&x < y \to x + z < y + z,\\
&0 < x \to 0 < y \to 0 < x \cdot y.
\end{aligned}$$

Ordered fields are the models of this extended axiom system. An ordered field is called *archimedian ordered* if for every element a of the field there is a natural number n such that a is less than the n-fold multiple of the 1 in the field.

THEOREM. *For every archimedian ordered field there is an elementarily equivalent ordered field that is not archimedian ordered.*

PROOF. Let $\mathcal{K}$ be an archimedian ordered field, x a variable and

$$\Gamma := \mathrm{Th}(\mathcal{K}) \cup \{\, \underline{n} < x \mid n \in \mathbb{N} \,\}.$$

Clearly every finite subset of Γ is satisfiable, hence by the general compactness theorem also Γ. Therefore we have $\mathcal{M}$ and η such that $\mathcal{M} \models \Gamma[\eta]$. Because of $\mathcal{M} \models \mathrm{Th}(\mathcal{K})$ we obtain $\mathcal{M} \equiv \mathcal{K}$ and hence $\mathcal{M}$ is an ordered field. Moreover $1^{\mathcal{M}} \cdot n <^{\mathcal{M}} \eta(x)$ for all $n \in \mathbb{N}$, hence $\mathcal{M}$ is not archimedian ordered. □

2.3.3. Axiomatizable models. A class $\mathcal{S}$ of $\mathcal{L}$-models is (*finitely*) *axiomatizable* if there is a (finite) axiom system Γ such that $\mathcal{S} = \mathrm{Mod}_{\mathcal{L}}(\Gamma)$. Clearly $\mathcal{S}$ is finitely axiomatizable if and only if $\mathcal{S} = \mathrm{Mod}_{\mathcal{L}}(\{A\})$ for some formula A. If for every $\mathcal{M} \in \mathcal{S}$ there is an elementarily equivalent $\mathcal{M}' \notin \mathcal{S}$, then $\mathcal{S}$ cannot possibly be axiomatizable. By the theorem above we can conclude that the class of archimedian ordered fields is not axiomatizable. It also follows that the class of non archimedian ordered fields is not axiomatizable.

LEMMA. *Let $\mathcal{S}$ be a class of $\mathcal{L}$-models and Γ an axiom system.*

(a) *$\mathcal{S}$ is finitely axiomatizable if and only if $\mathcal{S}$ and the complement of $\mathcal{S}$ are axiomatizable.*
(b) *If $\mathrm{Mod}_{\mathcal{L}}(\Gamma)$ is finitely axiomatizable, then there is a finite $\Gamma_0 \subseteq \Gamma$ such that $\mathrm{Mod}_{\mathcal{L}}(\Gamma_0) = \mathrm{Mod}_{\mathcal{L}}(\Gamma)$.*

PROOF. (a). Let $\mathcal{S}^{\mathrm{C}}$ denote the complement of $\mathcal{S}$. For $\to$ assume $\mathcal{S} = \mathrm{Mod}_{\mathcal{L}}(\{A\})$. Then $\mathcal{M} \in \mathcal{S}^{\mathrm{C}} \leftrightarrow \mathcal{M} \models \neg A$, hence $\mathcal{S}^{\mathrm{C}} = \mathrm{Mod}_{\mathcal{L}}(\{\neg A\})$.

For the converse. assume $\mathcal{S} = \mathrm{Mod}_{\mathcal{L}}(\Gamma_1)$ and $\mathcal{S}^{\mathrm{C}} = \mathrm{Mod}_{\mathcal{L}}(\Gamma_2)$. Then $\Gamma_1 \cup \Gamma_2$ is not satisfiable, hence there is a finite $\Gamma \subseteq \Gamma_1$ such that $\Gamma \cup \Gamma_2$ is not satisfiable. One obtains

$$\mathcal{M} \in \mathcal{S} \to \mathcal{M} \models \Gamma \to \mathcal{M} \not\models \Gamma_2 \to \mathcal{M} \notin \mathcal{S}^{\mathrm{C}} \to \mathcal{M} \in \mathcal{S}.$$

Hence $\mathcal{S} = \mathrm{Mod}_{\mathcal{L}}(\Gamma)$.

(b). Let $\mathrm{Mod}_{\mathcal{L}}(\Gamma) = \mathrm{Mod}_{\mathcal{L}}(\{A\})$. Then $\Gamma \vdash_c A$, hence also $\Gamma_0 \vdash_c A$ for a finite $\Gamma_0 \subseteq \Gamma$. One obtains

$$\mathcal{M} \models \Gamma \to \mathcal{M} \models \Gamma_0 \to \mathcal{M} \models A \to \mathcal{M} \models \Gamma.$$

Hence $\mathrm{Mod}_{\mathcal{L}}(\Gamma_0) = \mathrm{Mod}_{\mathcal{L}}(\Gamma)$. □

2.3.4. Dense linear orders without end points. Finally we consider as an example of a complete theory the theory DO of dense linear orders without end points. The axioms are (the equality axioms and)

$$\begin{aligned} &x \not< x, && x < y \to \exists_z(x < z \wedge z < y),\\ &x < y \to y < z \to x < z, && \exists_y(x < y),\\ &x < y \vee x = y \vee y < x, && \exists_y(y < x). \end{aligned}$$

Lemma. *Every countable model of* DO *is isomorphic to the model* $(\mathbb{Q}, <)$ *of rational numbers.*

Proof. Let $\mathcal{M} = (D, \prec)$ be a countable model of DO; we can assume that $=^{\mathcal{M}}$ is the equality on D. Let $D = \{\, b_n \mid n \in \mathbb{N} \,\}$ and $\mathbb{Q} = \{\, a_n \mid n \in \mathbb{N} \,\}$, where we may assume $a_n \neq a_m$ and $b_n \neq b_m$ for $n < m$. We define recursively functions $f_n \subseteq \mathbb{Q} \times D$ as follows. Let $f_0 := \{(a_0, b_0)\}$. Assume we have already constructed f_n.

Case $n+1 = 2m$. Let j be minimal such that $b_j \notin \mathrm{ran}(f_n)$. Choose $a_i \notin \mathrm{dom}(f_n)$ such that for all $a \in \mathrm{dom}(f_n)$ we have $a_i < a \leftrightarrow b_j < f_n(a)$; such an a_i exists, since $\mathcal{M}$ and $(\mathbb{Q}, <)$ are models of DO. Let $f_{n+1} := f_n \cup \{(a_i, b_j)\}$.

Case $n + 1 = 2m + 1$. This is treated similarly. Let i be minimal such that $a_i \notin \mathrm{dom}(f_n)$. Choose $b_j \notin \mathrm{ran}(f_n)$ such that for all $a \in \mathrm{dom}(f_n)$ we have $a_i < a \leftrightarrow b_j < f_n(a)$; such a b_j exists, since $\mathcal{M}$ and $(\mathbb{Q}, <)$ are models of DO. Let $f_{n+1} := f_n \cup \{(a_i, b_j)\}$.

Then $\{b_0, \ldots, b_m\} \subseteq \mathrm{ran}(f_{2m})$ and $\{a_0, \ldots, a_{m+1}\} \subseteq \mathrm{dom}(f_{2m+1})$ by construction, and $f := \bigcup_n f_n$ is an isomorphism of $(\mathbb{Q}, <)$ onto $\mathcal{M}$. □

Theorem. *The theory* DO *is complete, and* $\mathrm{DO} = \mathrm{Th}(\mathbb{Q}, <)$.

Proof. Clearly $(\mathbb{Q}, <)$ is a model of DO. Hence by 2.2.3 it suffices to show that for every model $\mathcal{M}$ of DO we have $\mathcal{M} \equiv (\mathbb{Q}, <)$. So let $\mathcal{M}$ model of DO. By 2.2.3 there is a countable $\mathcal{M}'$ such that $\mathcal{M} \equiv \mathcal{M}'$. By the preceding lemma $\mathcal{M}' \cong (\mathbb{Q}, <)$, hence $\mathcal{M} \equiv \mathcal{M}' \equiv (\mathbb{Q}, <)$. □

A further example of a complete theory is the theory of algebraically closed fields. For a proof of this fact and for many more subjects of model theory we refer to the literature (e.g., Chang and Keisler (1990) or Ebbinghaus et al. (1996)).

CHAPTER 3

Recursion Theory

In this chapter we develop the basics of recursive function theory, or as it is more generally known, computability theory. Its history goes back to the seminal works of Turing, Kleene and others in the 1930's.

A computable function is one defined by a program whose operational semantics tell an idealized computer what to do to its storage locations as it proceeds deterministically from input to output, without any prior restrictions on storage space or computation time. We shall be concerned with various program-styles and the relationships between them, but the emphasis throughout will be on one underlying data-type, namely the natural numbers, since it is there that the most basic foundational connections between proof theory and computation are to be seen in their clearest light.

The two best-known models of machine computation are the Turing Machine and the (Unlimited) Register Machine of Shepherdson and Sturgis (1963). We base our development on the latter since it affords the quickest route to the results we want to establish.

3.1. Register Machines

3.1.1. Programs. A *register machine* stores natural numbers in registers denoted u, v, w, x, y, z possibly with subscripts, and it responds step by step to a *program* consisting of an ordered list of basic instructions:

$$\begin{array}{l} I_0 \\ I_1 \\ \vdots \\ I_{k-1} \end{array}$$

Each instruction has one of the following three forms whose meanings are obvious:

Zero: $x := 0$,
Succ: $x := x + 1$,
Jump: [**if** $x = y$ **then** I_n **else** I_m].

The instructions are obeyed in order starting with I_0 except when a conditional jump instruction is encountered, in which case the next instruction

will be either I_n or I_m according as the numerical contents of registers x and y are equal or not at that stage. The computation *terminates* when it runs out of instructions, that is when the next instruction called for is I_k. Thus if a program of length k contains a jump instruction as above then it must satisfy the condition $n, m \leq k$ and I_k means "halt". Notice of course that some programs do not terminate, for example the following one-liner:

$$[\textbf{if } x = x \textbf{ then } I_0 \textbf{ else } I_1]$$

3.1.2. Program constructs. We develop some shorthand for building up standard sorts of programs.

Transfer. "$x := y$" is the program

$$\begin{aligned} &x := 0\\ &[\textbf{if } x = y \textbf{ then } I_4 \textbf{ else } I_2]\\ &x := x + 1\\ &[\textbf{if } x = x \textbf{ then } I_1 \textbf{ else } I_1], \end{aligned}$$

which copies the contents of register y into register x.

Predecessor. The program "$x := y \dot{-} 1$" copies the modified predecessor of y into x, and simultaneously copies y into z:

$$\begin{aligned} &x := 0\\ &z := 0\\ &[\textbf{if } x = y \textbf{ then } I_8 \textbf{ else } I_3]\\ &z := z + 1\\ &[\textbf{if } z = y \textbf{ then } I_8 \textbf{ else } I_5]\\ &z := z + 1\\ &x := x + 1\\ &[\textbf{if } z = y \textbf{ then } I_8 \textbf{ else } I_5]. \end{aligned}$$

Composition. "P **;** Q" is the program obtained by concatenating program P with program Q. However in order to ensure that jump instructions in Q of the form "$[\textbf{if } x = y \textbf{ then } I_n \textbf{ else } I_m]$" still operate properly within Q they need to be re-numbered by changing the addresses n, m to $k+n, k+m$ respectively where k is the length of program P. Thus the effect of this program is to do P until it halts (if ever) and then do Q.

Conditional. "**if** $x = y$ **then** P **else** Q **fi**" is the program

$$\begin{aligned} &[\textbf{if } x = y \textbf{ then } I_1 \textbf{ else } I_{k+2}]\\ &\vdots P\\ &[\textbf{if } x = x \textbf{ then } I_{k+2+l} \textbf{ else } I_2]\\ &\vdots Q \end{aligned}$$

where k, l are the lengths of the programs P, Q respectively, and again their jump instructions must be appropriately renumbered by adding 1 to the

addresses in P and $k+2$ to the addresses in Q. Clearly if $x = y$ then program P is obeyed and the next jump instruction automatically bypasses Q and halts. If $x \neq y$ then program Q is performed.

For Loop. "**for** $i = 1 \dots x$ **do** P **od**" is the program

$$\begin{array}{l} i := 0 \\ [\textbf{if } x = i \textbf{ then } I_{k+4} \textbf{ else } I_2] \\ i := i + 1 \\ \vdots P \\ [\textbf{if } x = i \textbf{ then } I_{k+4} \textbf{ else } I_2] \end{array}$$

where again, k is the length of program P and the jump instructions in P must be appropriately re-addressed by adding 3. The intention of this new program is that it should iterate the program P x times (do nothing if $x = 0$). This requires the restriction that the register x and the "local" counting-register i are not re-assigned new values inside P.

While Loop. "**while** $x \neq 0$ **do** P **od**" is the program

$$\begin{array}{l} y := 0 \\ [\textbf{if } x = y \textbf{ then } I_{k+3} \textbf{ else } I_2] \\ \vdots P \\ [\textbf{if } x = y \textbf{ then } I_{k+3} \textbf{ else } I_2] \end{array}$$

where again, k is the length of program P and the jump instructions in P must be re-addressed by adding 2. This program keeps on doing P until (if ever) the register x becomes 0; it requires the restriction that the auxiliary register y is not re-assigned new values inside P.

3.1.3. Register machine computable functions. A register machine program P may have certain distinguished "input registers" and "output registers". It may also use other "working registers" for scratchwork and these will initially be set to zero. We write $P(x_1, \dots, x_k; y)$ to signify that program P has input registers $x_1, \dots, x_k$ and one output register y, which are distinct.

DEFINITION. The program $P(x_1, \dots, x_k; y)$ is said to *compute* the k-ary partial function $\varphi\colon \mathbb{N}^k \to \mathbb{N}$ if, starting with any numerical values $n_1, \dots, n_k$ in the input registers, the program terminates with the number m in the output register if and only if $\varphi(n_1, \dots, n_k)$ is defined with value m. In this case, the input registers hold their original values.

A function is *register machine computable* if there is some program which computes it.

Here are some examples.

Addition. "Add$(x, y; z)$" is the program

$$z := x \ \textbf{;}\ \textbf{for}\ i = 1, \ldots, y\ \textbf{do}\ z := z + 1\ \textbf{od}$$

which adds the contents of registers x and y into register z.

Subtraction. "Subt$(x, y; z)$" is the program

$$z := x \ \textbf{;}\ \textbf{for}\ i = 1, \ldots, y\ \textbf{do}\ w := z \dot{-} 1 \ \textbf{;}\ z := w\ \textbf{od}$$

which computes the modified subtraction function $x \dot{-} y$.

Bounded Sum. If $P(x_1, \ldots, x_k, w; y)$ computes the $k+1$-ary function φ then the program $Q(x_1, \ldots, x_k, z; x)$:

$x := 0$ **;**
for $i = 1, \ldots, z$ **do** $w := i \dot{-} 1$ **;** $P(\vec{x}, w; y)$ **;** $v := x$ **;** Add$(v, y; x)$ **od**

computes the function

$$\psi(x_1, \ldots, x_k, z) = \sum_{w<z} \varphi(x_1, \ldots, x_k, w)$$

which will be undefined if for some $w < z$, $\varphi(x_1, \ldots, x_k, w)$ is undefined.

Multiplication. Deleting "$w := i \dot{-} 1$ **;** P" from the last example gives a program Mult$(z, y; x)$ which places the product of y and z into x.

Bounded Product. If in the bounded sum example, the instruction $x := x + 1$ is inserted immediately after $x := 0$, and if Add$(v, y; x)$ is replaced by Mult$(v, y; x)$, then the resulting program computes the function

$$\psi(x_1, \ldots, x_k, z) = \prod_{w<z} \varphi(x_1, \ldots, x_k, w).$$

Composition. If $P_j(x_1, \ldots, x_k; y_j)$ computes φ_j for each $j = i, \ldots, n$ and if $P_0(y_1, \ldots, y_n; y_0)$ computes φ_0, then the program $Q(x_1, \ldots, x_k; y_0)$:

$$P_1(x_1, \ldots, x_k; y_1)\ \textbf{;}\ \ldots\ \textbf{;}\ P_n(x_1, \ldots, x_k; y_n)\ \textbf{;}\ P_0(y_1, \ldots, y_n; y_0)$$

computes the function

$$\psi(x_1, \ldots, x_k) = \varphi_0(\varphi_1(x_1, \ldots, x_k), \ldots, \varphi_n(x_1, \ldots, x_k))$$

which will be undefined if any of the φ-subterms on the right hand side is undefined.

Unbounded Minimization. If $P(x_1, \ldots, x_k, y; z)$ computes φ then the program $Q(x_1, \ldots, x_k; z)$:

$y := 0$ **;** $z := 0$ **;** $z := z + 1$ **;**
while $z \neq 0$ **do** $P(x_1, \ldots, x_k, y; z)$ **;** $y := y + 1$ **od ;**
$z := y \dot{-} 1$

computes the function

$$\psi(x_1, \ldots, x_k) = \mu_y(\varphi(x_1, \ldots, x_k, y) = 0)$$

that is, the *least number* y such that $\varphi(x_1, \ldots, x_k, y')$ is defined for every $y' \leq y$ and $\varphi(x_1, \ldots, x_k, y) = 0$.

3.2. Elementary Functions

3.2.1. Definition and simple properties. The *elementary functions* of Kalmár (1943) are those number-theoretic functions which can be defined explicitly by compositional terms built up from variables and the constants $0, 1$ by repeated applications of addition $+$, modified subtraction $\dot{-}$, bounded sums and bounded products.

By omitting bounded products, one obtains the *subelementary* functions.

The examples in the previous section show that all elementary functions are computable and totally defined. Multiplication and exponentiation are elementary since

$$m \cdot n = \textstyle\sum_{i<n} m \text{ and } m^n = \prod_{i<n} m$$

and hence by repeated composition, all exponential polynomials are elementary.

In addition the elementary functions are closed under

Definition by Cases.

$$f(\vec{n}\,) = \begin{cases} g_0(\vec{n}\,) & \text{if } h(\vec{n}\,) = 0 \\ g_1(\vec{n}\,) & \text{otherwise} \end{cases}$$

since f can be defined from g_0, g_1 and h by

$$f(\vec{n}\,) = g_0(\vec{n}\,) \cdot (1 \dot{-} h(\vec{n}\,)) + g_1(\vec{n}\,) \cdot (1 \dot{-} (1 \dot{-} h(\vec{n}\,))).$$

Bounded Minimization.

$$f(\vec{n}, m) = \mu_{k<m}(g(\vec{n}, k) = 0)$$

since f can be defined from g by

$$f(\vec{n}, m) = \sum_{i<m} \big(1 \dot{-} \sum_{k \leq i} (1 \dot{-} g(\vec{n}, k))\big).$$

Note: this definition gives value m if there is no $k < m$ such that $g(\vec{n}, k) = 0$. It shows that not only the elementary, but in fact the subelementary functions are closed under bounded minimization. Furthermore, we define $\mu_{k \leq m}(g(\vec{n}, k) = 0)$ as $\mu_{k<m+1}(g(\vec{n}, k) = 0)$.

Lemma.

(a) *For every elementary function* $f\colon \mathbb{N}^r \to \mathbb{N}$ *there is a number* k *such that for all* $\vec{n} = n_1, \ldots, n_r$,

$$f(\vec{n}\,) < 2_k(\max(\vec{n}\,))$$

where $2_0(m) := m$ *and* $2_{k+1}(m) := 2^{2_k(m)}$.

(b) *Hence the function $n \mapsto 2_n(1)$ is not elementary.*

PROOF. (a). By induction on the build-up of the compositional term defining f. The result clearly holds if f is any one of the base functions:

$$f(\vec{n}\,) = 0 \text{ or } 1 \text{ or } n_i \text{ or } n_i + n_j \text{ or } n_i \dot{-} n_j.$$

If f is defined from g by application of bounded sum or product:

$$f(\vec{n}, m) = \sum_{i<m} g(\vec{n}, i) \text{ or } \prod_{i<m} g(\vec{n}, i)$$

where $g(\vec{n}, i) < 2_k(\max(\vec{n}, i))$ then we have

$$f(\vec{n}, m) \leq (2_k(\max(\vec{n}, m)))^m < 2_{k+2}(\max(\vec{n}, m))$$

using $n^n < 2^{2^n}$ (since $n^n < (2^n)^n \leq 2^{2^n}$ for $n > 3$).

If f is defined from $g_0, g_1, \ldots, g_l$ by composition:

$$f(\vec{n}\,) = g_0(g_1(\vec{n}\,), \ldots, g_l(\vec{n}\,))$$

where for each $j \leq l$ we have $g_j(-) < 2_{k_j}(\max(-))$, then with $k = \max_j k_j$,

$$f(\vec{n}\,) < 2_k(2_k(\max(\vec{n}\,))) = 2_{2k}(\max(\vec{n}\,))$$

and this completes the first part.

(b). If $2_n(1)$ were an elementary function of n then by (a) there would be a positive k such that for all n,

$$2_n(1) < 2_k(n)$$

but then putting $n = 2_k(1)$ yields $2_{2_k(1)}(1) < 2_{2k}(1)$, a contradiction. □

3.2.2. Elementary relations. A relation R on $\mathbb{N}^k$ is said to be *elementary* if its characteristic function

$$c_R(\vec{n}\,) = \begin{cases} 1 & \text{if } R(\vec{n}\,) \\ 0 & \text{otherwise} \end{cases}$$

is elementary. In particular, the "equality" and "less than" relations are elementary since their characteristic functions can be defined as follows:

$$c_<(n, m) = 1 \dot{-} (1 \dot{-} (m \dot{-} n)), \quad c_=(n, m) = 1 \dot{-} (c_<(n, m) + c_<(m, n)).$$

Furthermore if R is elementary then so is the function

$$f(\vec{n}, m) = \mu_{k<m} R(\vec{n}, k)$$

since $R(\vec{n}, k)$ is equivalent to $1 \dot{-} c_R(\vec{n}, k) = 0$.

LEMMA. *The elementary relations are closed under applications of propositional connectives and bounded quantifiers.*

PROOF. For example, the characteristic function of $\neg R$ is

$$1 \dot{-} c_R(\vec{n}\,).$$

The characteristic function of $R_0 \wedge R_1$ is

$$c_{R_0}(\vec{n}\,) \cdot c_{R_1}(\vec{n}\,).$$

The characteristic function of $\forall_{i<m} R(\vec{n}, i)$ is

$$c_=(m, \mu_{i<m}(c_R(\vec{n}, i) = 0)). \qquad \square$$

EXAMPLES. The above closure properties enable us to show that many "natural" functions and relations of number theory are elementary; thus

$$\begin{aligned}
\lfloor \frac{n}{m} \rfloor \quad &= \mu_{k<n}(n < (k+1)m), \\
n \bmod m &= n \dot{-} \lfloor \frac{n}{m} \rfloor m, \\
\mathrm{Prime}(n) &\leftrightarrow 1 < n \wedge \neg \exists_{m<n}(1 < m \wedge n \bmod m = 0), \\
p_n \quad &= \mu_{m<2^{2^n}}(\mathrm{Prime}(m) \wedge n = \sum_{i<m} c_{\mathrm{Prime}}(i)),
\end{aligned}$$

so $p_0, p_1, p_2, \ldots$ gives the enumeration of primes in increasing order. The estimate $p_n \leq 2^{2^n}$ for the nth prime p_n can be proved by induction on n: For $n = 0$ this is clear, and for $n \geq 1$ we obtain

$$p_n \leq p_0 p_1 \cdots p_{n-1} + 1 \leq 2^{2^0} 2^{2^1} \cdots 2^{2^{n-1}} + 1 = 2^{2^n - 1} + 1 < 2^{2^n}.$$

3.2.3. The class $\mathcal{E}$.

DEFINITION. The class $\mathcal{E}$ consists of those number theoretic functions which can be defined from the initial functions: constant 0, successor S, projections (onto the ith coordinate), addition $+$, modified subtraction $\dot{-}$, multiplication $\cdot$ and exponentiation 2^x, by applications of composition and bounded minimization.

The remarks above show immediately that the characteristic functions of the equality and less than relations lie in $\mathcal{E}$, and that (by the proof of the lemma) the relations in $\mathcal{E}$ are closed under propositional connectives and bounded quantifiers.

Furthermore the above examples show that all the functions in the class $\mathcal{E}$ are elementary. We now prove the converse, which will be useful later.

LEMMA. *There are "pairing functions" π, π_1, π_2 in $\mathcal{E}$ with the following properties:*

(a) π *maps* $\mathbb{N} \times \mathbb{N}$ *bijectively onto* $\mathbb{N}$,
(b) $\pi(a, b) + b + 2 \leq (a + b + 1)^2$ *for* $a + b \geq 1$, *hence* $\pi(a, b) < (a + b + 1)^2$,
(c) $\pi_1(c), \pi_2(c) \leq c$,

(d) $\pi(\pi_1(c), \pi_2(c)) = c$,
(e) $\pi_1(\pi(a, b)) = a$,
(f) $\pi_2(\pi(a, b)) = b$.

PROOF. Enumerate the pairs of natural numbers as follows:

$$\begin{array}{cccccc} \vdots & & & & & \\ 6 & \dots & & & & \\ 3 & 7 & \dots & & & \\ 1 & 4 & 8 & \dots & & \\ 0 & 2 & 5 & 9 & \dots & \end{array}$$

At position $(0, b)$ we clearly have the sum of the lengths of the preceding diagonals, and on the next diagonal $a + b$ remains constant. Let $\pi(a, b)$ be the number written at position (a, b). Then we have

$$\pi(a, b) = \big(\sum_{i \le a+b} i\big) + a = \frac{1}{2}(a + b)(a + b + 1) + a.$$

Clearly $\pi\colon \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ is bijective. Moreover, $a, b \le \pi(a, b)$ and in case $\pi(a, b) \neq 0$ also $a < \pi(a, b)$. Let

$$\pi_1(c) := \mu_{x \le c} \exists_{y \le c} (\pi(x, y) = c),$$
$$\pi_2(c) := \mu_{y \le c} \exists_{x \le c} (\pi(x, y) = c).$$

Then clearly $\pi_i(c) \le c$ for $i \in \{1, 2\}$ and

$$\pi_1(\pi(a, b)) = a, \quad \pi_2(\pi(a, b)) = b, \quad \pi(\pi_1(c), \pi_2(c)) = c.$$

π, π_1 and π_2 are in $\mathcal{E}$ by definition. For $\pi(a, b)$ we have the estimate

$$\pi(a, b) + b + 2 \le (a + b + 1)^2 \quad \text{for } a + b \ge 1.$$

This follows with $n := a + b$ from

$$\frac{1}{2} n(n + 1) + n + 2 \le (n + 1)^2 \quad \text{for } n \ge 1,$$

which is equivalent to $n(n + 1) + 2(n + 1) \le 2((n + 1)^2 - 1)$ and hence to $(n + 2)(n + 1) \le 2n(n + 2)$, which holds for $n \ge 1$. □

The proof shows that π, π_1 and π_2 are in fact subelementary.

THEOREM (Gödel's β-function). *There is in $\mathcal{E}$ a function β with the following property: For every sequence $a_0, \dots, a_{n-1} < b$ of numbers less than b we can find a number $c \le 4 \cdot 4^{n(b+n+1)^4}$ such that $\beta(c, i) = a_i$ for all $i < n$.*

PROOF. Let $\pi'(a,b) := \pi(a,b)+1$ and $\pi'_i(a) := \pi_i(a \dot{-} 1)$ for $i = 1, 2$. Consider

$$a := \pi'(b,n) \quad \text{and} \quad d := \prod_{i<n}(1+\pi'(a_i,i)a!).$$

From $a!$ and d we can, for each given $i < n$, reconstruct the number a_i as the unique $x < b$ such that $1+\pi'(x,i)a!$ properly divides d. For clearly a_i is such an x, and if some $x < b$ were to satisfy the same condition, then because $1 \le \pi'(x,i) < a$ and the numbers $1+ka!$ are relatively prime for $k \le a$, we would have $\pi'(x,i) = \pi'(a_j, j)$ for some $j < n$. Hence $x = a_j$ and $i = j$, thus $x = a_i$. – Therefore

$$a_i = \mu_{x<b}\exists_{z<d}((1+\pi'(x,i)a!)z = d).$$

We can now define Gödel's β-function as

$$\beta(c,i) := \mu_{x<\pi'_1(c)}\exists_{z<\pi'_2(c)}((1+\pi'(x,i)\cdot\pi'_1(c))\cdot z = \pi'_2(c)).$$

Clearly β is in $\mathcal{E}$. Furthermore with $c := \pi'(a!, d)$ we see that $\beta(c,i) = a_i$. It is then not difficult to estimate the given bound on c, using $\pi'(b,n) < (b+n+1)^2$. □

The above definition of β shows that it is subelementary.

3.2.4. Closure properties of $\mathcal{E}$.

THEOREM. *The class $\mathcal{E}$ is closed under limited recursion. Thus if g, h, k are given functions in $\mathcal{E}$ and f is defined from them according to the schema*

$$\begin{aligned} f(\vec{m},0) &= g(\vec{m}),\\ f(\vec{m},n+1) &= h(n, f(\vec{m},n), \vec{m}),\\ f(\vec{m},n) &\le k(\vec{m},n), \end{aligned}$$

then f is in $\mathcal{E}$ also.

PROOF. Let f be defined from g, h and k in $\mathcal{E}$, by limited recursion as above. Using Gödel's β-function as in the last theorem we can find for any given $\vec{m}, n$ a number c such that $\beta(c,i) = f(\vec{m},i)$ for all $i \le n$. Let $R(\vec{m},n,c)$ be the relation

$$\beta(c,0) = g(\vec{m}) \wedge \forall_{i<n}(\beta(c,i+1) = h(i,\beta(c,i),\vec{m}))$$

and note by the remarks above that its characteristic function is in $\mathcal{E}$. It is clear, by induction, that if $R(\vec{m},n,c)$ holds then $\beta(c,i) = f(\vec{m},i)$, for all $i \le n$. Therefore we can define f explicitly by the equation

$$f(\vec{m},n) = \beta(\mu_c R(\vec{m},n,c), n).$$

f will lie in $\mathcal{E}$ if μ_c can be bounded by an $\mathcal{E}$ function. However, the theorem on Gödel's β-function gives a bound $4\cdot 4^{(n+1)(b+n+2)^4}$, where in this case b can be taken as the maximum of $k(\vec{m},i)$ for $i \le n$. But this can be defined

in $\mathcal{E}$ as $k(\vec{m}, i_0)$, where $i_0 = \mu_{i\leq n}\forall_{j\leq n}(k(\vec{m}, j) \leq k(\vec{m}, i))$. Hence μ_c can be bounded by an $\mathcal{E}$ function. □

REMARK. Note that it is in this proof only that the exponential function is required, in providing a bound for μ.

COROLLARY. *$\mathcal{E}$ is the class of all elementary functions.*

PROOF. It is sufficient merely to show that $\mathcal{E}$ is closed under bounded sums and bounded products. Suppose for instance, that f is defined from g in $\mathcal{E}$ by bounded summation: $f(\vec{m}, n) = \sum_{i<n} g(\vec{m}, i)$. Then f can be defined by limited recursion, as follows

$$\begin{aligned} f(\vec{m}, 0) &= 0 \\ f(\vec{m}, n+1) &= f(\vec{m}, n) + g(\vec{m}, n) \\ f(\vec{m}, n) &\leq n \cdot \max_{i<n} g(\vec{m}, i) \end{aligned}$$

and the functions (including the bound) from which it is defined are in $\mathcal{E}$. Thus f is in $\mathcal{E}$ by the theorem. If instead, f is defined by bounded product, then proceed similarly. □

3.2.5. Coding finite lists. Computation on lists is a practical necessity, so because we are basing everything here on the single data type $\mathbb{N}$ we must develop some means of "coding" finite lists or sequences of natural numbers into $\mathbb{N}$ itself. There are various ways to do this and we shall adopt one of the most traditional, based on the pairing functions π, π_1, π_2.

The empty sequence is coded by the number 0 and a sequence n_0, n_1, $\ldots$, n_{k-1} is coded by the "sequence number"

$$\langle n_0, n_1, \ldots, n_{k-1}\rangle = \pi'(\ldots\pi'(\pi'(0, n_0), n_1), \ldots, n_{k-1})$$

with $\pi'(a, b) := \pi(a, b) + 1$, thus recursively,

$$\begin{aligned} \langle\rangle &:= 0, \\ \langle n_0, n_1, \ldots, n_k\rangle &:= \pi'(\langle n_0, n_1, \ldots, n_{k-1}\rangle, n_k). \end{aligned}$$

Because of the surjectivity of π, every number a can be decoded uniquely as a sequence number $a = \langle n_0, n_1, \ldots, n_{k-1}\rangle$. If a is greater than zero, $\mathrm{hd}(a) := \pi_2(a \dot{-} 1)$ is the "head" (i.e., rightmost element) and $\mathrm{tl}(a) := \pi_1(a \dot{-} 1)$ is the "tail" of the list. The kth iterate of tl is denoted $\mathrm{tl}^{(k)}$ and since $\mathrm{tl}(a)$ is less than or equal to a, $\mathrm{tl}^{(k)}(a)$ is elementarily definable (by limited recursion). Thus we can define elementarily the "length" and "decoding" functions:

$$\begin{aligned} \mathrm{lh}(a) &:= \mu_{k\leq a}(\mathrm{tl}^{(k)}(a) = 0), \\ (a)_i &:= \mathrm{hd}(\mathrm{tl}^{(\mathrm{lh}(a)\dot{-}(i+1))}(a)). \end{aligned}$$

Then if $a = \langle n_0, n_1, \ldots, n_{k-1}\rangle$ it is easy to check that

$$\mathrm{lh}(a) = k \text{ and } (a)_i = n_i \text{ for each } i < k.$$

Furthermore $(a)_i = 0$ when $i \geq \mathrm{lh}(a)$. We shall write $(a)_{i,j}$ for $((a)_i)_j$ and $(a)_{i,j,k}$ for $(((a)_i)_j)_k$. This elementary coding machinery will be used at various crucial points in the following.

Note that our previous remarks show that the functions $\mathrm{lh}(\cdot)$ and $(a)_i$ are subelementary, and so is $\langle n_0, n_1, \ldots, n_{k-1}\rangle$ for each fixed k.

LEMMA (Estimate for sequence numbers).

$$(n+1)k \leq \langle \underbrace{n, \ldots, n}_{k}\rangle < (n+1)^{2^k} \quad \textit{for } n \geq 1.$$

PROOF. We prove a slightly strengthened form of the second estimate:

$$\langle \underbrace{n, \ldots, n}_{k}\rangle + n + 1 \leq (n+1)^{2^k},$$

by induction on k. For $k = 0$ the claim is clear. In the step $k \mapsto k+1$ we have

$$\begin{aligned}
\langle \underbrace{n, \ldots, n}_{k+1}\rangle + n + 1 &= \pi(\langle \underbrace{n, \ldots, n}_{k}\rangle, n) + n + 2 \\
&\leq (\langle \underbrace{n, \ldots, n}_{k}\rangle + n + 1)^2 \quad \text{by the lemma in 3.2.3} \\
&\leq (n+1)^{2^{k+1}} \quad \text{by induction hypothesis.}
\end{aligned}$$

For the first estimate the base case $k = 0$ is clear, and in the step we have

$$\begin{aligned}
\langle \underbrace{n, \ldots, n}_{k+1}\rangle &= \pi(\langle \underbrace{n, \ldots, n}_{k}\rangle, n) + 1 \\
&\geq \langle \underbrace{n, \ldots, n}_{k}\rangle + n + 1 \\
&\geq (n+1)(k+1) \quad \text{by induction hypothesis.}
\end{aligned}$$

□

Concatenation of sequence numbers $b * a$ is defined thus:

$$\begin{aligned}
b * \langle\rangle &:= b, \\
b * \langle n_0, n_1, \ldots, n_k\rangle &:= \pi(b * \langle n_0, n_1, \ldots, n_{k-1}\rangle, n_k) + 1.
\end{aligned}$$

To check that this operation is also elementary, define $h(b, a, i)$ by recursion on i as follows.

$$\begin{aligned}
h(b, a, 0) &= b, \\
h(b, a, i+1) &= \pi(h(b, a, i), (a)_i) + 1
\end{aligned}$$

and note that since

$$h(b,a,i) = \langle (b)_0, \ldots, (b)_{\mathrm{lh}(b)\dot{-}1}, (a)_0, \ldots, (a)_{i\dot{-}1}\rangle \quad \text{for } i \le \mathrm{lh}(a)$$

it follows from the estimate above that $h(a,b,i) \le (b+a)^{2^{\mathrm{lh}(b)+i}}$. Thus h is definable by limited recursion from elementary functions and hence is itself elementary. Finally

$$b * a = h(b,a,\mathrm{lh}(a)).$$

LEMMA. *The class $\mathcal{E}$ is closed under limited course-of-values recursion. Thus if h, k are given functions in $\mathcal{E}$ and f is defined from them according to the schema*

$$f(\vec{m},n) = h(n, \langle f(\vec{m},0), \ldots, f(\vec{m},n-1)\rangle, \vec{m})$$
$$f(\vec{m},n) \le k(\vec{m},n)$$

then f is in $\mathcal{E}$ also.

PROOF. $\bar{f}(\vec{m},n) := \langle f(\vec{m},0), \ldots, f(\vec{m},n-1)\rangle$ is definable by

$$\begin{aligned} \bar{f}(\vec{m},0) &= 0, \\ \bar{f}(\vec{m},n+1) &= \bar{f}(\vec{m},n) * \langle h(n,\bar{f}(\vec{m},n),\vec{m})\rangle \\ \bar{f}(\vec{m},n) &\le (\sum_{i<n} k(\vec{m},i) + 2)^{2^n}, \end{aligned}$$

using $\langle \underbrace{n,\ldots,n}_{k}\rangle < (n+1)^{2^k}$. But $f(\vec{m},n) = (\bar{f}(\vec{m},n+1))_n$. □

The next lemma gives closure of $\mathcal{E}$ under limited course-of-values recursion but with parameter substitution allowed. Here we are working at the extremity of elementary definability, but this generalized schema will be crucially important for the elementary arithmetization of syntax which is developed prior to Gödel's theorems in the next chapter (particularly in regard to the substitution function). Unfortunately this last closure property of $\mathcal{E}$ is rather complicated to state, because it requires notational details to do with iteration of parameter substitutions.

LEMMA. *The class $\mathcal{E}$ is closed under limited course-of-values recursion with parameter substitution. Suppose g, h, k, p_i and a_i (for $i \le l$) are all in $\mathcal{E}$ and let f be defined from them as follows.*

$$f(m,n) = \begin{cases} g(m) & \text{if } n = 0 \\ h(n, f(p_0(m,n), a_0(n)), \ldots, f(p_l(m,n), a_l(n)), m) & \text{otherwise} \end{cases}$$
$$f(m,n) \le k(m,n)$$

where $a_i(n) < n$ when $n > 0$. Then f is also in $\mathcal{E}$ provided that the iterated parameter function $p(\sigma, m, n)$ defined below is elementarily bounded.

For any sequence $\sigma := \langle i_0, i_1, \ldots, i_{r-1}\rangle$ of numbers $\leq l$ define $n(\sigma)$ by: $n(\langle\rangle) := n$, $n(\sigma * \langle i\rangle) := a_i(n(\sigma))$ if $n(\sigma) \neq 0$ and $:= 0$ otherwise. Then $p(\sigma, m, n)$ is given by the course-of-values recursion:

$$p(\langle\rangle, m, n) = m$$

$$p(\sigma * \langle i\rangle, m, n) = \begin{cases} p_i(p(\sigma, m, n), n(\sigma)) & \text{if } n(\sigma) \neq 0 \\ p(\sigma, m, n) & \text{if } n(\sigma) = 0. \end{cases}$$

PROOF. First note that since $p(\sigma, m, n)$ is defined by a course-of-values recursion and, by supposition, is elementarily bounded, it is itself in $\mathcal{E}$ by the last lemma. Similarly, $n(\sigma)$ is elementary.

We code the computation of $f(m, n)$ as a finitely branching tree of height $\leq n + 1$. Nodes are sequence numbers $\sigma = \langle i_0, i_1, \ldots, i_{r-1}\rangle$ with $i_j \leq l$ and each such node is bounded in value by $(l+1)^{2^{n+1}}$. At each node σ is attached the value of f at the current parameter substitution $p(\sigma, m, n)$ and the current stage $n(\sigma)$. Let $Q(m, n, z)$ be the elementary relation expressing the fact that z correctly encodes the computation tree for $f(m, n)$ with $(z)_\sigma$ being the correct value at current node σ. Thus $Q(m, n, z)$ is the following condition, for all nodes $\sigma \leq (l+1)^{2^{n+1}}$: If $n(\sigma) \neq 0$, $(z)_\sigma = h(n(\sigma), (z)_{\sigma*\langle 0\rangle}, \ldots, (z)_{\sigma*\langle l\rangle}, p(\sigma, m, n))$ and if $n(\sigma) = 0$ then $(z)_\sigma = g(p(\sigma, m, n))$. Clearly Q is an elementary relation, and if z is the least such that $Q(m, n, z)$ holds then $f(m, n) = (z)_{\langle\rangle}$. Therefore f will be elementary if z can be bounded by an elementary function. This is now easy because $z = \langle (z)_{\langle\rangle}, (z)_1, \ldots (z)_{(l+1)^{2^{n+1}}}\rangle$ where each $(z)_\sigma = f(p(\sigma, m, n), n(\sigma)) \leq k(p(\sigma, m, n), n(\sigma))$. Therefore

$$z \leq (\max\{\, k(p(\sigma, m, n), n(\sigma)) \mid \sigma \leq (l+1)^{2^{n+1}} \,\} + 1)^{2^{(l+1)^{2^{n+1}}}}$$

and this is elementary. □

3.3. The Normal Form Theorem

3.3.1. Program numbers. The three types of register machine instructions I can be coded by "instruction numbers" $\sharp I$ thus, where v_0, v_1, v_2, ... is a list of all variables used to denote registers:

If I is "$v_j := 0$" then $\sharp I = \langle 0, j\rangle$.
If I is "$v_j := v_j + 1$" then $\sharp I = \langle 1, j\rangle$.
If I is "**if** $v_j = v_l$ **then** I_m **else** I_n" then $\sharp I = \langle 2, j, l, m, n\rangle$.

Clearly, using the sequence coding and decoding apparatus above, we can check elementarily whether or not a given number is an instruction number.

Any register machine program $P = I_0, I_1, \ldots, I_{k-1}$ can then be coded by a "program number" or "index" $\sharp P$ thus:

$$\sharp P = \langle \sharp I_0, \sharp I_1, \ldots, \sharp I_{k-1} \rangle$$

and again (although it is tedious) we can elementarily check whether or not a given number is indeed of the form $\sharp P$ for some program P. Tradition has it that e is normally reserved as a variable over putative program numbers.

Standard program constructs such as those in 3.1 have associated "index-constructors", i.e., functions which, given indices of the subprograms, produce an index for the constructed program. The point is that for standard program constructs the associated index-constructor functions are elementary. For example there is an elementary index-constructor comp such that, given programs P_0, P_1 with indices e_0, e_1, $\mathrm{comp}(e_0, e_1)$ is an index of the program P_0 **;** P_1. A moment's thought should convince the reader that the appropriate definition of comp is as follows:

$$\mathrm{comp}(e_0, e_1) = e_0 * \langle r(e_0, e_1, 0), r(e_0, e_1, 1), \ldots, r(e_0, e_1, \mathrm{lh}(e_1) \dot{-} 1) \rangle$$

where $r(e_0, e_1, i) =$

$$\begin{cases} \langle 2, (e_1)_{i,1}, (e_1)_{i,2}, (e_1)_{i,3} + \mathrm{lh}(e_0), (e_1)_{i,4} + \mathrm{lh}(e_0) \rangle & \text{if } (e_1)_{i,0} = 2 \\ (e_1)_i & \text{otherwise} \end{cases}$$

re-addresses the jump instructions in P_1. Clearly r and hence comp are elementary functions.

Definition. Henceforth, $\varphi_e^{(r)}$ denotes the partial function computed by the register machine program with program number e, operating on the input registers $v_1, \ldots, v_r$ and with output register v_0. There is no loss of generality here, since the variables in any program can always be renamed so that $v_1, \ldots, v_r$ become the input registers and v_0 the output. If e is not a program number, or it is but does not operate on the right variables, then we adopt the convention that $\varphi_e^{(r)}(n_1, \ldots, n_r)$ is undefined for all inputs $n_1, \ldots, n_r$.

3.3.2. Normal form.

Theorem (Kleene's Normal Form). *For each arity r there is an elementary function U and an elementary relation T such that, for all e and all inputs $n_1, \ldots, n_r$,*

(a) $\varphi_e^{(r)}(n_1, \ldots, n_r)$ *is defined if and only if* $\exists_s T(e, n_1, \ldots, n_r, s)$,

(b) $\varphi_e^{(r)}(n_1, \ldots, n_r) = U(e, n_1, \ldots, n_r, \mu_s T(e, n_1, \ldots, n_r, s))$.

Proof. A computation of a register machine program $P(v_1, \ldots, v_r; v_0)$ on numerical inputs $\vec{n} = n_1, \ldots, n_r$ proceeds deterministically, step by step,

each step corresponding to the execution of one instruction. Let e be its program number, and let $v_0, \ldots, v_l$ be all the registers used by P, including the "working registers" so $r \leq l$.

The "state" of the computation at step s is defined to be the sequence number

$$\text{state}(e, \vec{n}, s) = \langle e, i, m_0, m_1, \ldots, m_l \rangle$$

where $m_0, m_1, \ldots, m_l$ are the values stored in the registers $v_0, v_1, \ldots, v_l$ after step s is completed, and the next instruction to be performed is the ith one, thus $(e)_i$ is its instruction number.

The "state transition function" $\text{tr}\colon \mathbb{N} \to \mathbb{N}$ computes the "next state". So suppose that $x = \langle e, i, m_0, m_1, \ldots, m_l \rangle$ is any putative state. Then in what follows, $e = (x)_0$, $i = (x)_1$, and $m_j = (x)_{j+2}$ for each $j \leq l$. The definition of $\text{tr}(x)$ is therefore as follows:

$$\text{tr}(x) = \langle e, i', m_0', m_1', \ldots, m_l' \rangle$$

where

(i) If $(e)_i = \langle 0, j \rangle$ where $j \leq l$ then $i' = i + 1$, $m_j' = 0$, and all other registers remain unchanged, i.e., $m_k' = m_k$ for $k \neq j$.

(ii) If $(e)_i = \langle 1, j \rangle$ where $j \leq l$ then $i' = i + 1$, $m_j' = m_j + 1$, and all other registers remain unchanged.

(iii) If $(e)_i = \langle 2, j_0, j_1, i_0, i_1 \rangle$ where $j_0, j_1 \leq l$ and $i_0, i_1 \leq \text{lh}(e)$ then $i' = i_0$ or $i' = i_1$ according as $m_{j_0} = m_{j_1}$ or not, and all registers remain unchanged, i.e., $m_j' = m_j$ for all $j \leq l$.

(iv) Otherwise, if e is not a program number, or if it refers to a register v_k with $l < k$, or if $\text{lh}(e) \leq i$, then $\text{tr}(x)$ simply repeats the same state x so $i' = i$, and $m_j' = m_j$ for every $j \leq l$.

Clearly tr is an *elementary* function, since it is defined by elementarily decidable cases, with (a great deal of) elementary decoding and re-coding involved in each case.

Consequently, the "state function" $\text{state}(e, \vec{n}, s)$ is also *elementary* because it can be defined by iterating the transition function by limited recursion on s as follows:

$$\begin{aligned} \text{state}(e, \vec{n}, 0) &= \langle e, 0, 0, n_1, \ldots, n_r, 0, \ldots, 0 \rangle \\ \text{state}(e, \vec{n}, s+1) &= \text{tr}(\text{state}(e, \vec{n}, s)) \\ \text{state}(e, \vec{n}, s) &\leq h(e, \vec{n}, s) \end{aligned}$$

where for the bounding function h we can take

$$h(e, \vec{n}, s) = \langle e, e \rangle * \langle \max(\vec{n}\,) + s, \ldots, \max(\vec{n}\,) + s \rangle.$$

This is because the maximum value of any register at step s cannot be greater than $\max(\vec{n}\,) + s$. Now this expression clearly is elementary, since

$\langle m, \ldots, m\rangle$ with i occurrences of m is definable by a limited recursion with bound $(m+i)^{2^i}$, as is easily seen by induction on i.

Now recall that if program P has program number e then computation terminates when instruction $I_{\mathrm{lh}(e)}$ is encountered. Thus we can define the "termination relation" $T(e, \vec{n}, s)$ meaning "computation terminates at step s", by

$$T(e, \vec{n}, s) := ((\mathrm{state}(e, \vec{n}, s))_1 = \mathrm{lh}(e)).$$

Clearly T is elementary and

$$\varphi_e^{(r)}(\vec{n}\,) \text{ is defined} \leftrightarrow \exists_s T(e, \vec{n}, s).$$

The output on termination is the value of register v_0, so if we define the "output function" $U(e, \vec{n}, s)$ by

$$U(e, \vec{n}, s) := (\mathrm{state}(e, \vec{n}, s))_2$$

then U is also elementary and

$$\varphi_e^{(r)}(\vec{n}\,) = U(e, \vec{n}, \mu_s T(e, \vec{n}, s)). \qquad \square$$

3.3.3. Σ_1^0-definable relations and μ-recursive functions. A relation R of arity r is said to be Σ_1^0-*definable* if there is an elementary relation E, say of arity $r+l$, such that for all $\vec{n} = n_1, \ldots, n_r$,

$$R(\vec{n}\,) \leftrightarrow \exists_{k_1} \ldots \exists_{k_l} E(\vec{n}, k_1, \ldots, k_l).$$

A partial function φ is said to be Σ_1^0-*definable* if its graph

$$\{\, (\vec{n}, m) \mid \varphi(\vec{n}\,) \text{ is defined and } = m \,\}$$

is Σ_1^0-definable.

To say that a non-empty relation R is Σ_1^0-definable is equivalent to saying that the set of all sequences $\langle \vec{n}\rangle$ satisfying R can be enumerated (possibly with repetitions) by some elementary function $f\colon \mathbb{N} \to \mathbb{N}$. Such relations are called *elementarily enumerable*. For choose any fixed sequence $\langle a_1, \ldots, a_r\rangle$ satisfying R and define

$$f(m) = \begin{cases} \langle (m)_1, \ldots, (m)_r\rangle & \text{if } E((m)_1, \ldots, (m)_{r+l}) \\ \langle a_1, \ldots, a_r\rangle & \text{otherwise.} \end{cases}$$

Conversely if R is elementarily enumerated by f then

$$R(\vec{n}\,) \leftrightarrow \exists_m (f(m) = \langle \vec{n}\rangle)$$

is a Σ_1^0-definition of R.

The *μ-recursive functions* are those (partial) functions which can be defined from the initial functions: constant 0, successor S, projections (onto the ith coordinate), addition $+$, modified subtraction $\dot{-}$ and multiplication

$\cdot$, by applications of composition and unbounded minimization. Note that it is through unbounded minimization that partial functions may arise.

LEMMA. *Every elementary function is μ-recursive.*

PROOF. By simply removing the bounds on μ in the lemmas in 3.2.3 one obtains μ-recursive definitions of the pairing functions π, π_1, π_2 and of Gödel's β-function. Then by removing all mention of bounds from the Theorem in 3.2.4 one sees that the μ-recursive functions are closed under (unlimited) primitive recursive definitions: $f(\vec{m}, 0) = g(\vec{m})$, $f(\vec{m}, n+1) = h(n, f(\vec{m}, n))$. Thus one can μ-recursively define bounded sums and bounded products, and hence all elementary functions. □

3.3.4. Computable functions.

DEFINITION. The *while-programs* are those programs which can be built up from assignment statements $x := 0$, $x := y$, $x := y + 1$, $x := y \dot{-} 1$, by Conditionals, Composition, For-Loops and While-Loops as in 3.1 (on program constructs).

THEOREM. *The following are equivalent:*

(a) φ *is register machine computable,*
(b) φ *is Σ^0_1-definable,*
(c) φ *is μ-recursive,*
(d) φ *is computable by a while program.*

PROOF. The Normal Form Theorem shows immediately that every register machine computable function $\varphi_e^{(r)}$ is Σ^0_1-definable since

$$\varphi_e^{(r)}(\vec{n}\,) = m \leftrightarrow \exists_s \big(T(e, \vec{n}, s) \wedge U(e, \vec{n}, s) = m \big)$$

and the relation $T(e, \vec{n}, s) \wedge U(e, \vec{n}, s) = m$ is clearly elementary. If φ is Σ^0_1-definable, say

$$\varphi(\vec{n}\,) = m \leftrightarrow \exists_{k_1} \ldots \exists_{k_l} E(\vec{n}, m, k_1, \ldots, k_l)$$

then φ can be defined μ-recursively by

$$\varphi(\vec{n}\,) = (\mu_m E(\vec{n}, (m)_0, (m)_1, \ldots, (m)_l)\,)_0 \ ,$$

using the fact (above) that elementary functions are μ-recursive. The examples of computable functionals in 3.1 show how the definition of any μ-recursive function translates automatically into a while program. Finally, 3.1 shows how to implement any while program on a register machine. □

Henceforth *computable* (or *recursive*) means "register machine computable" or any of its equivalents.

COROLLARY. *The function $\varphi_e^{(r)}(n_1, \ldots, n_r)$ is a computable partial function of the $r+1$ variables $e, n_1, \ldots, n_r$.*

PROOF. Immediate from the Normal Form. □

LEMMA. *A relation R is computable if and only if both R and its complement $\mathbb{N}^n \setminus R$ are Σ_1^0-definable.*

PROOF. We can assume that both R and $\mathbb{N}^n \setminus R$ are not empty, and (for simplicity) also $n = 1$.

"$\rightarrow$". By the theorem above every computable relation is Σ_1^0-definable, and with R clearly its complement is computable.

"$\leftarrow$". Let $f, g \in \mathcal{E}$ enumerate R and $\mathbb{N} \setminus R$, respectively. Then

$$h(n) := \mu_i(f(i) = n \vee g(i) = n)$$

is a total μ-recursive function, and $R(n) \leftrightarrow f(h(n)) = n$. □

3.3.5. Undecidability of the halting problem. The above corollary says that there is a single "universal" program which, given numbers e and $\vec{n}$, computes $\varphi_e^{(r)}(\vec{n}\,)$ if it is defined. However we cannot decide in advance whether or not it will be defined. There is no program which, given e and $\vec{n}$, computes the total function

$$h(e, \vec{n}\,) = \begin{cases} 1 & \text{if } \varphi_e^{(r)}(\vec{n}\,) \text{ is defined,} \\ 0 & \text{if } \varphi_e^{(r)}(\vec{n}\,) \text{ is undefined.} \end{cases}$$

For suppose there were such a program. Then the function

$$\psi(\vec{n}\,) = \mu_m(h(n_1, \vec{n}\,) = 0)$$

would be computable, say with fixed program number e_0, and therefore

$$\varphi_{e_0}^{(r)}(\vec{n}\,) = \begin{cases} 0 & \text{if } h(n_1, \vec{n}\,) = 0 \\ \text{undefined} & \text{if } h(n_1, \vec{n}\,) = 1. \end{cases}$$

But then fixing $n_1 = e_0$ gives:

$$\varphi_{e_0}^{(r)}(\vec{n}\,) \text{ defined} \leftrightarrow h(e_0, \vec{n}\,) = 0 \leftrightarrow \varphi_{e_0}^{(r)}(\vec{n}\,) \text{ undefined},$$

a contradiction. Hence the relation $R(e, \vec{n}\,)$ which holds if and only if $\varphi_e^{(r)}(\vec{n}\,)$ is defined, is not recursive. It is however Σ_1^0-definable.

CHAPTER 4

Gödel's Theorems

4.1. Gödel Numbers

We will assign numbers – so-called Gödel numbers, GN for short – to the syntactical constructs developed in chapter 1: terms, formulas and derivations. Using the elementary sequence-coding and decoding machinery developed earlier we will be able to construct the code number of a composed object from its parts, and conversely to disassemble the code number of a composed object into the code numbers of its parts.

4.1.1. Gödel numbers of terms, formulas and derivations. Let $\mathcal{L}$ be a countable first order language. Assume that we have injectively assigned to every n-ary relation symbol R a *symbol number* $\mathrm{sn}(R)$ of the form $\langle 1, n, i\rangle$ and to every n-ary function symbol f a symbol number $\mathrm{sn}(f)$ of the form $\langle 2, n, j\rangle$. Call $\mathcal{L}$ *elementarily presented* if the set $\mathrm{Symb}_{\mathcal{L}}$ of all these symbol numbers is elementary. In what follows we shall always assume that the languages $\mathcal{L}$ considered are elementarily presented. In particular this applies to every language with finitely many relation and function symbols.

Let $\mathrm{sn}(\mathrm{Var}) := \langle 0\rangle$. For every $\mathcal{L}$-term r we define recursively its Gödel number $\ulcorner r \urcorner$ by

$$\begin{aligned} \ulcorner x_i \urcorner &:= \langle \mathrm{sn}(\mathrm{Var}), i\rangle, \\ \ulcorner f r_1 \dots r_n \urcorner &:= \langle \mathrm{sn}(f), \ulcorner r_1 \urcorner, \dots, \ulcorner r_n \urcorner\rangle. \end{aligned}$$

Assign numbers to the logical symbols by $\mathrm{sn}(\to) := \langle 3, 0\rangle$ und $\mathrm{sn}(\forall) := \langle 3, 1\rangle$. For simplicity we leave out the logical connectives $\wedge$, $\vee$ and $\exists$ here; they could be treated similarly. We define for every $\mathcal{L}$-formula A its Gödel number $\ulcorner A \urcorner$ by

$$\begin{aligned} \ulcorner R r_1 \dots r_n \urcorner &:= \langle \mathrm{sn}(R), \ulcorner r_1 \urcorner, \dots, \ulcorner r_n \urcorner\rangle, \\ \ulcorner A \to B \urcorner &:= \langle \mathrm{sn}(\to), \ulcorner A \urcorner, \ulcorner B \urcorner\rangle, \\ \ulcorner \forall_{x_i} A \urcorner &:= \langle \mathrm{sn}(\forall), i, \ulcorner A \urcorner\rangle. \end{aligned}$$

We define symbol numbers for the names of the natural deduction rules: $\mathrm{sn}(\mathrm{AssVar}) := \langle 4, 0\rangle$, $\mathrm{sn}(\to^+) := \langle 4, 1\rangle$, $\mathrm{sn}(\to^-) := \langle 4, 2\rangle$, $\mathrm{sn}(\forall^+) := \langle 4, 3\rangle$,

$\mathrm{sn}(\forall^-) := \langle 4, 4\rangle$. For a derivation M we define its Gödel number $\ulcorner M \urcorner$ by

$$\begin{aligned}
\ulcorner u_i^A \urcorner &:= \langle \mathrm{sn}(\mathrm{AssVar}), i, \ulcorner A \urcorner\rangle,\\
\ulcorner \lambda_{u_i^A} M \urcorner &:= \langle \mathrm{sn}(\to^+), i, \ulcorner A \urcorner, \ulcorner M \urcorner\rangle,\\
\ulcorner MN \urcorner &:= \langle \mathrm{sn}(\to^-), \ulcorner M \urcorner, \ulcorner N \urcorner\rangle,\\
\ulcorner \lambda_{x_i} M \urcorner &:= \langle \mathrm{sn}(\forall^+), i, \ulcorner M \urcorner\rangle,\\
\ulcorner Mr \urcorner &:= \langle \mathrm{sn}(\forall^-), \ulcorner M \urcorner, \ulcorner r \urcorner\rangle.
\end{aligned}$$

It will be helpful in the sequel to have some general estimates on Gödel numbers, which we provide here. For a term r or formula A we define its *sum of maximal sequence lengths* $\|r\|$ or $\|A\|$ by

$$\begin{aligned}
&\|x_i\| := 2,\\
&\|fr_0 \dots r_{k-1}\| := k + 1 + \max(\|r_i\|),
\end{aligned}
\qquad
\begin{aligned}
&\|Rr_0 \dots r_{k-1}\| := k + 1 + \max(\|r_i\|),\\
&\|A \to B\| := 3 + \max(\|A\|, \|B\|),\\
&\|\forall_{x_i} A\| := 3 + \|A\|
\end{aligned}$$

and its *symbol bound* $\mathrm{sb}(r)$ or $\mathrm{sb}(A)$ by

$$\begin{aligned}
\mathrm{sb}(x_i) &:= 1 + \max(\mathrm{sn}(\mathrm{Var}), i),\\
\mathrm{sb}(f) &:= 1 + \mathrm{sn}(f),\\
\mathrm{sb}(fr_0 \dots r_k) &:= \max(\mathrm{sn}(f), \max(\mathrm{sb}(r_i))),\\
\mathrm{sb}(R) &:= 1 + \mathrm{sn}(R),\\
\mathrm{sb}(Rr_0 \dots r_k) &:= \max(\mathrm{sn}(R), \max(\mathrm{sb}(r_i))),\\
\mathrm{sb}(A \to B) &:= \max(\mathrm{sn}(\to), \mathrm{sb}(A), \mathrm{sb}(B)),\\
\mathrm{sb}(\forall_{x_i} A) &:= \max(\mathrm{sn}(\forall), i, \mathrm{sb}(A)).
\end{aligned}$$

LEMMA. $\|r\| \le \ulcorner r \urcorner < \mathrm{sb}(r)^{2^{\|r\|}}$ *and* $\|A\| \le \ulcorner A \urcorner < \mathrm{sb}(A)^{2^{\|A\|}}$.

PROOF. We prove $\|r\| \le \ulcorner r \urcorner$ by induction on r. *Case* x_i.

$$\|x_i\| = 2 \le \langle \mathrm{sn}(\mathrm{Var}), i\rangle = \ulcorner x_i \urcorner.$$

Case $fr_0 \dots r_{k-1}$. First note that $k + \sum_{i<k} n_i \le \langle n_0, \dots, n_{k-1}\rangle$ can be proved easily, by induction on k. Hence

$$\begin{aligned}
\ulcorner fr_0 \dots r_{k-1} \urcorner &= \langle \mathrm{sn}(f), \ulcorner r_0 \urcorner, \dots, \ulcorner r_{k-1} \urcorner\rangle\\
&\ge k + 1 + \max(\ulcorner r_i \urcorner)\\
&\ge k + 1 + \max(\|r_i\|) \quad \text{by induction hypothesis}\\
&= \|fr_0 \dots r_{k-1}\|.
\end{aligned}$$

The proof of $\|A\| \le \ulcorner A \urcorner$ is similar. For $\ulcorner r \urcorner < \mathrm{sb}(r)^{2^{\|r\|}}$ we again use induction on r. For a variable x_i we obtain by the estimate in 3.2.5

$$\ulcorner x_i \urcorner = \langle \mathrm{sn}(\mathrm{Var}), i\rangle < \mathrm{sb}(x_i)^{2^2} = \mathrm{sb}(x_i)^{2^{\|x_i\|}}$$

and for a constant f

$$\ulcorner f \urcorner = \langle \mathrm{sn}(f) \rangle = \langle \mathrm{sb}(f) \dot{-} 1 \rangle < \mathrm{sb}(f)^2 = \mathrm{sb}(f)^{2^{\|f\|}}.$$

For a term $r := fr_0 \dots r_{k-1}$ built with a function symbol f or arity $k > 0$ we have

$$\begin{aligned}
&\ulcorner f r_0 \dots r_{k-1} \urcorner \\
&= \langle \mathrm{sn}(f), \ulcorner r_0 \urcorner, \dots, \ulcorner r_{k-1} \urcorner \rangle \\
&\leq \langle \underbrace{n \dot{-} 1, n \dot{-} 1, \dots, n \dot{-} 1}_{k+1} \rangle \quad \text{with } n := \mathrm{sb}(r)^{2^{\max \|r_i\|}}, \text{ by ind. hyp.} \\
&< n^{2^{k+1}} \quad \text{by the estimate in 3.2.5} \\
&= \mathrm{sb}(r)^{2^{k+1+\max \|r_i\|}} = \mathrm{sb}(r)^{2^{\|r\|}}.
\end{aligned}$$

The proof of $\ulcorner A \urcorner < \mathrm{sb}(A)^{2^{\|A\|}}$ is again similar, but we spell out the quantifier case $A := \forall_{x_i} B$:

$$\begin{aligned}
\ulcorner \forall_{x_i} B \urcorner &= \langle \mathrm{sn}(\forall), i, \ulcorner B \urcorner \rangle \\
&\leq \langle n \dot{-} 1, n \dot{-} 1, n \dot{-} 1 \rangle \quad \text{with } n := \mathrm{sb}(A)^{2^{\|B\|}}, \text{ by ind. hyp.} \\
&< n^{2^3} \quad \text{by the estimate in 3.2.5} \\
&= \mathrm{sb}(A)^{2^{3+\|B\|}} = \mathrm{sb}(A)^{2^{\|A\|}}.
\end{aligned}$$

□

4.1.2. Elementary functions on Gödel numbers. We shall define an elementary predicate Deriv such that $\mathrm{Deriv}(d)$ if and only if d is the Gödel number of a derivation. To this end we need a number of auxiliary functions and relations, which will all be elementary and have the properties described. (The convention is that relations are capitalized and functions are lower case). First we need some basic notions:

$\mathrm{Ter}(t)$	t is GN of a term,
$\mathrm{For}(a)$	a is GN of a formula,
$\alpha\text{-Eq}(x, y)$	the terms/formulas with GNs x, y are α-equal,
$\mathrm{FV}(i, y)$	the variable x_i is free in the term or formula with GN y,
$\mathrm{fmla}(d)$	GN of the formula derived by the derivation with GN d.

By the *context* of a derivation M we mean the set $\{u_{i_0}^{A_0}, \dots, u_{i_{n-1}}^{A_{n-1}}\}$ of its free assumption variables, where $i_0 < \dots < i_{n-1}$. Its Gödel number is defined to be the least number c such that $\forall_{\nu<n}((c)_{i_\nu} = \ulcorner A_\nu \urcorner)$.

$\mathrm{ctx}(d)$	GN of the context of the derivation with GN d,
$\mathrm{Cons}(c_1, c_2)$	the contexts with GN c_1, c_2 are consistent.

Then Deriv can be defined by course-of-values recursion, using the next-to-last lemma in 3.2.5.

$$\begin{aligned}\mathrm{Deriv}(d) := {} & ((d)_0 = \mathrm{sn}(\mathrm{AssVar}) \wedge \mathrm{lh}(d) = 3 \wedge \mathrm{For}((d)_2)) \vee \\ & ((d)_0 = \mathrm{sn}(\rightarrow^+) \wedge \mathrm{lh}(d) = 4 \wedge \mathrm{For}((d)_2) \wedge \mathrm{Deriv}((d)_3) \wedge \\ & \quad ((\mathrm{ctx}((d)_3))_{(d)_1} \neq 0 \rightarrow (\mathrm{ctx}((d)_3))_{(d)_1} = (d)_2)) \vee \\ & ((d)_0 = \mathrm{sn}(\rightarrow^-) \wedge \mathrm{lh}(d) = 3 \wedge \mathrm{Deriv}((d)_1) \wedge \mathrm{Deriv}((d)_2) \wedge \\ & \quad \mathrm{Cons}(\mathrm{ctx}((d)_1), \mathrm{ctx}((d)_2)) \wedge \\ & \quad (\mathrm{fmla}((d)_1))_0 = \mathrm{sn}(\rightarrow) \wedge (\mathrm{fmla}((d)_1))_1 = \mathrm{fmla}((d)_2)) \vee \\ & ((d)_0 = \mathrm{sn}(\forall^+) \wedge \mathrm{lh}(d) = 3 \wedge \mathrm{Deriv}((d)_2) \wedge \forall_{i<\mathrm{lh}(\mathrm{ctx}((d)_2))}(\\ & \quad (\mathrm{ctx}((d)_2))_i \neq 0 \rightarrow \neg\mathrm{FV}((d)_1, (\mathrm{ctx}((d)_2))_i))) \vee \\ & ((d)_0 = \mathrm{sn}(\forall^-) \wedge \mathrm{lh}(d) = 3 \wedge \mathrm{Deriv}((d)_1) \wedge \mathrm{Ter}((d)_2) \wedge \\ & \quad (\mathrm{fmla}((d)_1))_0 = \mathrm{sn}(\forall)).\end{aligned}$$

Still further auxiliary functions are needed. A *substitution* is a map $x_{i_0} \mapsto r_0, \ldots, x_{i_{n-1}} \mapsto r_{n-1}$ with $i_0 < \cdots < i_{n-1}$ from variables to terms; its Gödel number is the least number s such that $\forall_{\nu<n}((s)_{i_\nu} = \ulcorner r_\nu \urcorner)$. Hence $(s)_{i_\nu} = 0$ indicates that s leaves x_{i_ν} unchanged.

$\mathrm{union}(c_1, c_2)$	GN of the union of the consistent contexts with GN c_1, c_2,
$\mathrm{remove}(c, i)$	GN of result of removing u_i from the context with GN c,
$\mathrm{sub}(x, s)$	GN of the result of applying the substitution with GN s to the term or formula with GN x,
$\mathrm{update}(s, i, t)$	GN of the result of updating the substitution with GN s by changing its entry at i to the term with GN t.

We now give definitions of all these; from the form of the definitions it will be clear that they have the required properties, and are elementary.

Update. This can be defined explicitly, using the bounded least number operator:

$$\begin{aligned}&\mathrm{update}(s, i, t) := \\ &\quad \mu_{x<h(\max(s,t),\max(\mathrm{lh}(s),i))}((x)_i = t \wedge \forall_{k<\max(\mathrm{lh}(s),i)}(k \neq i \rightarrow (x)_k = (s)_k))\end{aligned}$$

where $h(n, k) := (n + 1)^{2^k}$ is the elementary function defined earlier with the property $\langle n, \ldots, n\rangle \leq h(n, k)$.

Substitution. The substitution function defined next takes a formula or term with GN x and applies to it a substitution with GN s to produce a new formula with GN y. The substitution works by assigning specific terms to the free variables, but in order to avoid clashing it must also reassign

new variables to the universally bound ones. This occurs in the final clause of the definition where, to be on the safe side, we (recursively) assign to a bound variable the new variable with index $x + i(s)$, where $i(s)$ is the maximum index of any variable occurring in a value term $(s)_j$ of s. Notice that $i(s) \leq s$. We define substitution by a limited course-of-values recursion with parameter substitutions:

$$\mathrm{sub}(x, s) :=$$
$$\begin{cases} x & \text{if } (x)_0 = \mathrm{sn}(\mathrm{Var}) \wedge (s)_{(x)_1} = 0, \\ (s)_{(x)_1} & \text{if } (x)_0 = \mathrm{sn}(\mathrm{Var}) \wedge (s)_{(x)_1} \neq 0, \\ \mu_{y \leq k(x,s)}(\mathrm{lh}(x) = \mathrm{lh}(y) \wedge (x)_0 = (y)_0 \wedge \forall_{i<l}(\mathrm{sub}((x)_{i+1}, s) = (y)_{i+1})) & \\ \quad \text{if } (x)_{0,0} = 1 \vee (x)_{0,0} = 2 \vee (x)_0 = \mathrm{sn}(\rightarrow), & \\ \langle \mathrm{sn}(\forall), x + i(s), \mathrm{sub}((x)_2, \mathrm{update}(s, (x)_1, \langle \mathrm{sn}(\mathrm{Var}), x + i(s) \rangle)) \rangle & \\ \quad \text{if } (x)_0 = \mathrm{sn}(\forall), & \\ 0 \quad \text{otherwise}, & \end{cases}$$
$$\mathrm{sub}(x, s) \leq k(x, s),$$

where it is assumed that the relation and function symbols in the given language $\mathcal{L}$ all have arity $\leq l$. The bound $k(x, s)$ and a bound for the iterated parameter updates remain to be provided, so that the last lemma in 3.2.5 can be applied. Then sub will be elementary.

First notice that as s is continually updated by the recursion, for the sake of (the formula or term with GN) x, the first update assigns to a bound variable in x a "new" variable with index $x+i(s)$. The next update will then assign to a bound variable in some subformula x' of x a new variable with index $x' + x + i(s)$ etcetera. The final update will therefore be a sequence of length $\leq x^2 + i(s)$, whose entries are all $< \max(s, \langle \mathrm{sn}(\mathrm{Var}), x^2 + i(s) \rangle)$. Thus a bound for all iterated updates starting from s and x is this last expression to the power of $2^{x^2+i(s)}$, which is elementary.

Using the lemma in 4.1.1 above one can see that if x is the GN of a term or a formula X and s is the GN of a substitution S, so that we may write $\mathrm{sub}(x, s) = \ulcorner X[S] \urcorner$, then $\mathrm{sb}(X[S]) \leq \max(s, x, x^2 + i(s)) \leq x^2 + s$ and, clearly, $\|X[S]\| \leq x + s$. The lemma then gives an elementary bound $k(x, s) := (x^2 + s)^{2^{x+s}}$ for $\mathrm{sub}(x, s)$.

Remove, union, consistency, context. Removal of an assumption variable from a context is defined by

$$\mathrm{remove}(c, i) := \mu_{x \leq c}((x)_i = 0 \wedge \forall_{j<\mathrm{lh}(c)}(j \neq i \rightarrow (x)_j = (c)_j)).$$

The union of two consistent contexts can again be defined by the bounded μ-operator:

$$\mathrm{union}(c_1,c_2) := \mu_{c\le c_1*c_2}\forall_{i<\max(\mathrm{lh}(c_1),\mathrm{lh}(c_2))}((c)_i = \max((c_1)_i,(c_2)_i)).$$

Consistency of two contexts is defined by

$$\mathrm{Cons}(c_1,c_2) :=$$
$$\forall_{i<\max(\mathrm{lh}(c_1),\mathrm{lh}(c_2))}((c_1)_i \neq 0 \to (c_2)_i \neq 0 \to \alpha\text{-}\mathrm{Eq}((c_1)_i,(c_2)_i)).$$

The context of a derivation is defined by

$$\begin{aligned}\mathrm{ctx}(d) := \mu_{c\le d}(&((d)_0 = \mathrm{sn}(\mathrm{AssVar}) \wedge (c)_{(d)_1} = (d)_2) \vee\\ &((d)_0 = \mathrm{sn}(\to^+) \wedge c = \mathrm{remove}(\mathrm{ctx}((d)_3),(d)_1)) \vee\\ &((d)_0 = \mathrm{sn}(\to^-) \wedge c = \mathrm{union}(\mathrm{ctx}((d)_1),\mathrm{ctx}((d)_2))) \vee\\ &((d)_0 = \mathrm{sn}(\forall^+) \wedge c = \mathrm{ctx}((d)_2)) \vee\\ &((d)_0 = \mathrm{sn}(\forall^-) \wedge c = \mathrm{ctx}((d)_1))).\end{aligned}$$

Formulas, terms. The end formula of a derivation is defined by

$$\begin{aligned}\mathrm{fmla}(d) := \mu_{a\le f(d)}(&((d)_0 = \mathrm{sn}(\mathrm{AssVar}) \wedge a = (d)_2) \vee\\ &((d)_0 = \mathrm{sn}(\to^+) \wedge a = \langle \mathrm{sn}(\to),(d)_2,\mathrm{fmla}((d)_3)\rangle) \vee\\ &((d)_0 = \mathrm{sn}(\to^-) \wedge a = (\mathrm{fmla}((d)_1))_2) \vee\\ &((d)_0 = \mathrm{sn}(\forall^+) \wedge a = \langle \mathrm{sn}(\forall),(d)_1,\mathrm{fmla}((d)_2)\rangle) \vee\\ &((d)_0 = \mathrm{sn}(\forall^-) \wedge\\ &\qquad \mathrm{sub}((\mathrm{fmla}((d)_1))_2,\mu_{s\le d}((s)_{(\mathrm{fmla}((d)_1))_1}{=}(d)_2)) = a)),\end{aligned}$$

where the elementary bound $f(d)$ remains to be provided. Clearly it suffices to have an elementary estimate of $\ulcorner A(r)\urcorner$ in terms of $a = \ulcorner \forall_x A(x)\urcorner$ and $b = \ulcorner r\urcorner$. For the GN s of the substitution assigning r to x we have $s \le a^{2^b}$. Hence $\ulcorner A(r)\urcorner = \mathrm{sub}(a,s) \le k(a,a^{2^b}) \le k(d,d^{2^d}) =: f(d)$.

Notice that this is the only place in our definitions of auxiliary functions and relations where the substitution function is needed.

Freeness of a variable x_i in a term or formula is defined by

$$\begin{aligned}\mathrm{FV}(i,y) := &((y)_0 = \mathrm{sn}(\mathrm{Var}) \wedge (y)_1 = i) \vee\\ &((y)_{0,0} = 1 \wedge \exists_{j<\mathrm{lh}(y)\dot{-}1}\mathrm{FV}(i,(y)_{j+1})) \vee\\ &((y)_{0,0} = 2 \wedge \exists_{j<\mathrm{lh}(y)\dot{-}1}\mathrm{FV}(i,(y)_{j+1})) \vee\\ &((y)_0 = \mathrm{sn}(\to) \wedge (\mathrm{FV}(i,(y)_1) \vee \mathrm{FV}(i,(y)_2))) \vee\\ &((y)_0 = \mathrm{sn}(\forall) \wedge i \neq (y)_1 \wedge \mathrm{FV}(i,(y)_2)).\end{aligned}$$

To define α-equality (i.e., equality up to renaming of bound variables) of formulas we use a relation $\mathrm{Corr}(n,m,s,t)$ due to Robert Stärk. The intuitive

meaning is this: two numbers n, m (indices of variables) are "correlated" w.r.t. coded lists s, t (of mutually inverted pairs of indices) if one of the following holds.

(i) There is a first element $\langle n, v\rangle$ of the form $\langle n, \dots\rangle$ in s and a first element $\langle m, u\rangle$ of the form $\langle m, \dots\rangle$ in t, and $v = m$, $u = n$.
(ii) There is no element of the form $\langle n, \dots\rangle$ in s and no element of the form $\langle m, \dots\rangle$ in t, and $n = m$.

We define Corr by

$$\begin{aligned}\mathrm{Corr}(n,m,s,t) := \exists_{i<\mathrm{lh}(s)}\exists_{j<\mathrm{lh}(t)}((s)_i = \langle n,(s)_{i,1}\rangle \wedge \forall_{i'<i}(s)_{i',0} \neq n \wedge \\ (t)_j = \langle m,(t)_{j,1}\rangle \wedge \forall_{j'<j}(t)_{j',0} \neq m \wedge \\ (s)_{i,1} = m \wedge (t)_{j,1} = n) \vee \\ (n = m \wedge \forall_{i<\mathrm{lh}(s)}(s)_{i,0} \neq n \wedge \forall_{j<\mathrm{lh}(t)}(t)_{j,0} \neq m).\end{aligned}$$

Now define α-Eq$'$ by

$$\begin{aligned}&\alpha\text{-Eq}'(a,b,s,t) := \\ &((a)_0 = (b)_0 = \mathrm{sn}(\mathrm{Var}) \wedge \mathrm{Corr}((a)_1,(b)_1,s,t)) \vee \\ &((a)_0 = (b)_0 \wedge \mathrm{Symb}_{\mathcal{L}}((a)_0) \wedge \forall_{i<(a)_{0,1}}\alpha\text{-Eq}'((a)_{i+1},(b)_{i+1},s,t)) \vee \\ &((a)_0 = (b)_0 = \mathrm{sn}(\to) \wedge \alpha\text{-Eq}'((a)_1,(b)_1,s,t) \wedge \alpha\text{-Eq}'((a)_2,(b)_2,s,t)) \vee \\ &((a)_0 = (b)_0 = \mathrm{sn}(\forall) \wedge \alpha\text{-Eq}'((a)_2,(b)_2,\langle\langle(a)_1,(b)_1\rangle\rangle * s,\langle\langle(b)_1,(a)_1\rangle\rangle * t)).\end{aligned}$$

α-Eq$'$ is an elementary relation because it is here defined by course-of-values recursion with parameter substitution, where iterates of the (quadratic) parameter updates are elementarily bounded. Finally $\alpha\text{-Eq}(x,y) := \alpha\text{-Eq}'(x,y,\langle\rangle,\langle\rangle)$.

The sets of formulas and terms are defined by

$$\begin{aligned}&\mathrm{For}(a) := \\ &((a)_{0,0} = 1 \wedge \mathrm{Symb}_{\mathcal{L}}((a)_0) \wedge \mathrm{lh}(a) = (a)_{0,1} + 1 \wedge \forall_{j<(a)_{0,1}}\mathrm{Ter}((a)_{j+1})) \vee \\ &((a)_0 = \mathrm{sn}(\to) \wedge \mathrm{lh}(a) = 3 \wedge \mathrm{For}((a)_1) \wedge \mathrm{For}((a)_2)) \vee \\ &((a)_0 = \mathrm{sn}(\forall) \wedge \mathrm{lh}(a) = 3 \wedge \mathrm{For}((a)_2)),\end{aligned}$$

$$\begin{aligned}\mathrm{Ter}(t) := ((t)_0 = \mathrm{sn}(\mathrm{Var}) \wedge \mathrm{lh}(t) = 2) \vee \\ ((t)_{0,0} = 2 \wedge \mathrm{Symb}_{\mathcal{L}}((t)_0) \wedge \mathrm{lh}(t) = (t)_{0,1}{+}1 \wedge \forall_{j<(t)_{0,1}}\mathrm{Ter}((t)_{j+1})).\end{aligned}$$

Recall that for simplicity we have left out the logical connectives $\wedge$, $\vee$ and $\exists$. They could be added easily, including an extension of the notion of a derivation to also allow their axioms as listed in 1.1.7.

4.1.3. Axiomatized theories. Let $\mathcal{L}$ be an elementarily presented language with $=$ in $\mathcal{L}$. Call a relation *recursive* if its (total) characteristic function is recursive. A set S of formulas is called *recursive* (*elementary*, *primitive recursive*, Σ^0_1*-definable*), if $\ulcorner S \urcorner := \{ \ulcorner A \urcorner \mid A \in S \}$ is recursive (elementary, primitive recursive, Σ^0_1-definable). Clearly the sets $\mathrm{Stab}_{\mathcal{L}}$ of stability axioms and $\mathrm{Eq}_{\mathcal{L}}$ of $\mathcal{L}$-equality axioms are elementary. A theory T with $L(T) \subseteq \mathcal{L}$ is *recursively (elementarily, primitive recursively) axiomatizable*, if there is a recursive (elementary, primitive recursive) set S of closed $\mathcal{L}$-formulas such that $T = \{ A \in \overline{\mathcal{L}} \mid S \cup \mathrm{Eq}_{\mathcal{L}} \vdash A \}$.

THEOREM. *For theories T with $L(T) \subseteq \mathcal{L}$ the following are equivalent.*

(a) *T is recursively axiomatizable.*
(b) *T is primitive recursively axiomatizable.*
(c) *T is elementarily axiomatizable.*
(d) *T is Σ^0_1-definable.*

PROOF. (d) $\to$ (c). Let $\ulcorner T \urcorner$ be recursively enumerable. Then there is an elementary f such that $\ulcorner T \urcorner = \mathrm{ran}(f)$. Let $f(n) = \ulcorner A_n \urcorner$. We define an elementary function g with the property $g(n) = \ulcorner A_0 \wedge \cdots \wedge A_n \urcorner$ by

$$\begin{aligned} g(0) &:= f(0), \\ g(n+1) &:= g(n) \mathbin{\dot\wedge} f(n+1), \\ g(n) &< m_n^{2^{3n}} \quad \text{where } m_n := 1 + \max(\mathrm{sn}(\wedge), \max_{i \le n} f(i)) \end{aligned}$$

with $a \mathbin{\dot\wedge} b := \langle \mathrm{sn}(\wedge), a, b \rangle$. The estimate is proved by induction on n. The base case is clear, and in the step we have

$$\begin{aligned} g(n+1) &= \langle \mathrm{sn}(\wedge), g(n), f(n+1) \rangle \\ &\le \langle m_{n+1} \dot- 1, m_n^{2^{3n}} \dot- 1, m_{n+1} \dot- 1 \rangle \quad \text{by induction hypothesis} \\ &< (m_{n+1}^{2^{3n}})^{2^3} \quad \text{by the estimate in 3.2.5} \\ &= m_{n+1}^{2^{3(n+1)}}. \end{aligned}$$

For $S := \{ A_0 \wedge \cdots \wedge A_n \mid n \in \mathbb{N} \}$ we have $\ulcorner S \urcorner = \mathrm{ran}(g)$, and this set is elementary because of $a \in \mathrm{ran}(g) \leftrightarrow \exists_{n<a}(a = g(n))$. T is elementarily axiomatizable, since $T = \{ A \in \overline{\mathcal{L}} \mid S \cup \mathrm{Eq}_{\mathcal{L}} \vdash A \}$.

(c) $\to$ (b) and (b) $\to$ (a) are clear.

(a) $\to$ (d). Let T be axiomatized by S with $\ulcorner S \urcorner$ recursive. Then

$$\begin{aligned} a \in \ulcorner T \urcorner \leftrightarrow \exists_d (&\mathrm{Deriv}(d) \wedge \mathrm{fmla}(d) = a \wedge \forall_{i<a} \neg \mathrm{FV}(i, a) \wedge \\ &\forall_{i < \mathrm{lh}(\mathrm{ctx}(d))} ((\mathrm{ctx}(d))_i \in \ulcorner \mathrm{Eq}_{\mathcal{L}} \urcorner \cup \ulcorner S \urcorner)). \end{aligned}$$

Hence $\ulcorner T \urcorner$ is Σ^0_1-definable. □

Call a theory T in our elementarily presented language $\mathcal{L}$ *axiomatized* if it is given by a Σ^0_1-definable axiom system Ax_T. By the theorem just proved we can even assume that Ax_T is elementary. For such axiomatized theories we define a binary relation Prf_T by

$$\mathrm{Prf}_T(d,a) := \mathrm{Deriv}(d) \wedge \mathrm{fmla}(d){=}a \wedge \forall_{i<\mathrm{lh}(\mathrm{ctx}(d))}((\mathrm{ctx}(d))_i \in \ulcorner \mathrm{Eq}_{\mathcal{L}} \urcorner \cup \ulcorner \mathrm{Ax}_T \urcorner).$$

Clearly Prf_T is elementary and $\mathrm{Prf}_T(d,a)$ if and only if d is the GN of a derivation of the formula with GN a from a context composed of equality axioms and formulas from Ax_T. A theory T is *consistent* if $\bot \notin T$; otherwise T is *inconsistent*. A theory T is *complete* if for every closed formula A we have $A \in T$ or $\neg A \in T$, and *incomplete* otherwise.

Corollary. *Let T be a consistent theory. If T is axiomatized and complete then T is recursive.*

Proof. We define the characteristic function $c_{\ulcorner T \urcorner}$ of $\ulcorner T \urcorner$ as follows. $c_{\ulcorner T \urcorner}(a)$ is 0 if $\neg\mathrm{For}(a)$ or $\exists_{i<a}\mathrm{FV}(i,a)$. Otherwise it is defined by

$$c_{\ulcorner T \urcorner}(a) = (\mu_x((\mathrm{Prf}_T((x)_0,a) \wedge (x)_1 = 1) \vee (\mathrm{Prf}_T((x)_0, \dot\neg a) \wedge (x)_1 = 0)))_1$$

with $\dot\neg a := \langle \mathrm{sn}(\rightarrow), a, \mathrm{sn}(\bot)\rangle$. Completeness of T implies that $c_{\ulcorner T \urcorner}$ is total, and consistency that it indeed is the characteristic function of $\ulcorner T \urcorner$. □

4.1.4. Undefinability of the notion of truth. Let $\mathcal{M}$ be an $\mathcal{L}$-structure. A relation $R \subseteq |\mathcal{M}|^n$ is called *definable* in $\mathcal{M}$ if there is an $\mathcal{L}$-formula $A(x_1,\dots,x_n)$ such that

$$R = \{\, (a_1,\dots,a_n) \in |\mathcal{M}|^n \mid \mathcal{M} \models A(x_1,\dots x_n)[x_1 := a_1,\dots,x_n := a_n] \,\}.$$

We assume in this section that $|\mathcal{M}| = \mathbb{N}$, 0 is a constant in $\mathcal{L}$ and S is a unary function symbol in $\mathcal{L}$ with $0^{\mathcal{M}} = 0$ and $\mathrm{S}^{\mathcal{M}}(a) = a+1$. Recall that for every $a \in \mathbb{N}$ the *numeral* $\underline{a} \in \mathrm{Ter}_{\mathcal{L}}$ is defined by $\underline{0} := 0$ and $\underline{n+1} := \mathrm{S}\underline{n}$. Observe that in this case the definability of $R \subseteq \mathbb{N}^n$ by $A(x_1,\dots,x_n)$ is equivalent to

$$R = \{\, (a_1,\dots,a_n) \in \mathbb{N}^n \mid \mathcal{M} \models A(\underline{a_1},\dots,\underline{a_n}) \,\}.$$

Furthermore let $\mathcal{L}$ be an elementarily presented language. We always assume in this section that every elementary relation is definable in $\mathcal{M}$. A set S of formulas is called *definable* in $\mathcal{M}$ if $\ulcorner S \urcorner := \{\, \ulcorner A \urcorner \mid A \in S \,\}$ is.

We shall show that already from these assumptions it follows that the notion of truth for $\mathcal{M}$, more precisely the set $\mathrm{Th}(\mathcal{M})$ of all closed formulas valid in $\mathcal{M}$, is undefinable in $\mathcal{M}$. From this it will follow that the notion of truth is in fact undecidable, for otherwise the set $\mathrm{Th}(\mathcal{M})$ would be recursive (Church's Thesis), hence recursively enumerable, and hence definable, because we have assumed already that all elementary relations are definable

in $\mathcal{M}$ and so their projections are definable also. For the proof we shall need the following Fixed Point Lemma, which will be generalized in 4.2.2.

LEMMA (Semantical Fixed Point Lemma). *If every elementary relation is definable in $\mathcal{M}$, then for every $\mathcal{L}$-formula $B(z)$ we can find a closed $\mathcal{L}$-formula A such that*

$$\mathcal{M} \models A \quad \textit{if and only if} \quad \mathcal{M} \models B(\underline{\ulcorner A \urcorner}).$$

PROOF. Let s be the elementary function satisfying for every formula $C = C(z)$ with $z := x_0$,

$$s(\ulcorner C \urcorner, k) = \mathrm{sub}(\ulcorner C \urcorner, \langle \ulcorner \underline{k} \urcorner \rangle) = \ulcorner C(\underline{k}) \urcorner$$

where sub is the substitution function already defined in 4.1.2. Hence in particular

$$s(\ulcorner C \urcorner, \ulcorner C \urcorner) = \ulcorner C(\underline{\ulcorner C \urcorner}) \urcorner.$$

By assumption the graph G_s of s is definable in $\mathcal{M}$, by $A_s(x_1, x_2, x_3)$ say. Let

$$C := \exists_x (B(x) \wedge A_s(z, z, x)), \quad A := C(\underline{\ulcorner C \urcorner}),$$

and therefore

$$A = \exists_x (B(x) \wedge A_s(\underline{\ulcorner C \urcorner}, \underline{\ulcorner C \urcorner}, x)).$$

Hence $\mathcal{M} \models A$ if and only if $\exists_{a \in \mathbb{N}}((\mathcal{M} \models B(\underline{a})) \wedge a = \ulcorner C(\underline{\ulcorner C \urcorner}) \urcorner)$, which is the same as $\mathcal{M} \models B(\underline{\ulcorner A \urcorner})$. □

THEOREM (Tarski's Undefinability Theorem). *Assume that every elementary relation is definable in $\mathcal{M}$. Then* $\mathrm{Th}(\mathcal{M})$ *is undefinable in $\mathcal{M}$, hence in particular not Σ^0_1-definable.*

PROOF. Assume that $\ulcorner \mathrm{Th}(\mathcal{M}) \urcorner$ is definable by $B_W(z)$. Then for all closed formulas A

$$\mathcal{M} \models A \quad \text{if and only if} \quad \mathcal{M} \models B_W(\underline{\ulcorner A \urcorner}).$$

Now consider the formula $\neg B_W(z)$ and choose by the Fixed Point Lemma a closed $\mathcal{L}$-formula A such that

$$\mathcal{M} \models A \quad \text{if and only if} \quad \mathcal{M} \models \neg B_W(\underline{\ulcorner A \urcorner}).$$

This contradicts the equivalence above.

We already have noticed that all Σ^0_1-definable relations are definable in $\mathcal{M}$. Hence it follows that $\ulcorner \mathrm{Th}(\mathcal{M}) \urcorner$ cannot be Σ^0_1-definable. □

4.2. The Notion of Truth in Formal Theories

We now want to generalize the arguments of the previous section. There we have made essential use of the notion of truth in a structure $\mathcal{M}$, i.e., of the relation $\mathcal{M} \models A$. The set of all closed formulas A such that $\mathcal{M} \models A$ has been called the theory of $\mathcal{M}$, denoted $\mathrm{Th}(\mathcal{M})$.

Now instead of $\mathrm{Th}(\mathcal{M})$ we shall start more generally from an arbitrary theory T. We consider the question as to whether in T there is a *notion of truth* (in the form of a *truth formula* $B(z)$), such that $B(z)$ "means" that z is "true". A consequence is that we have to explain all the notions used without referring to semantical concepts at all.

(i) z ranges over closed formulas (or sentences) A, or more precisely over their Gödel numbers $\ulcorner A \urcorner$.
(ii) A "true" is to be replaced by $T \vdash A$.
(iii) C "equivalent" to D is to be replaced by $T \vdash C \leftrightarrow D$.

Hence the question now is whether there is a truth formula $B(z)$ such that $T \vdash A \leftrightarrow B(\underline{\ulcorner A \urcorner})$ for all sentences A. The result will be that this is impossible, under rather weak assumptions on the theory T. Technically, the issue will be to replace the notion of definability by the notion of "representability" within a formal theory. We begin with a discussion of this notion.

In this section we assume that $\mathcal{L}$ is an elementarily presented language with 0, S and $=$ in $\mathcal{L}$, and T an $\mathcal{L}$-theory containing the equality axioms $\mathrm{Eq}_{\mathcal{L}}$.

4.2.1. Representable relations and functions.

Definition. A relation $R \subseteq \mathbb{N}^n$ is *representable* in T if there is a formula $A(x_1, \ldots, x_n)$ such that

$$\begin{aligned} T &\vdash A(\underline{a_1}, \ldots, \underline{a_n}) && \text{if } (a_1, \ldots, a_n) \in R, \\ T &\vdash \neg A(\underline{a_1}, \ldots, \underline{a_n}) && \text{if } (a_1, \ldots, a_n) \notin R. \end{aligned}$$

A function $f\colon \mathbb{N}^n \to \mathbb{N}$ is called *representable* in T if there is a formula $A(x_1, \ldots, x_n, y)$ representing the graph $G_f \subseteq \mathbb{N}^{n+1}$ of f, i.e., such that

$$T \vdash A(\underline{a_1}, \ldots, \underline{a_n}, \underline{f(a_1, \ldots, a_n)}), \tag{4.1}$$

$$T \vdash \neg A(\underline{a_1}, \ldots, \underline{a_n}, \underline{c}) \qquad \text{if } c \neq f(a_1, \ldots, a_n) \tag{4.2}$$

and such that in addition

$$T \vdash A(\underline{a_1}, \ldots, \underline{a_n}, y) \wedge A(\underline{a_1}, \ldots, \underline{a_n}, z) \to y{=}z \text{ for all } a_1, \ldots, a_n \in \mathbb{N}. \tag{4.3}$$

Note that in case $T \vdash \underline{b} \neq \underline{c}$ for $b < c$ condition (4.2) follows from (4.1) and (4.3).

LEMMA. *If the characteristic function c_R of a relation $R \subseteq \mathbb{N}^n$ is representable in T, then so is the relation R itself.*

PROOF. For simplicity assume $n = 1$. Let $A(x, y)$ be a formula representing c_R. We show that $A(x, \underline{1})$ represents the relation R. Assume $a \in R$. Then $c_R(a) = 1$, hence $(a, 1) \in G_{c_R}$, hence $T \vdash A(\underline{a}, \underline{1})$. Conversely, assume $a \notin R$. Then $c_R(a) = 0$, hence $(a, 1) \notin G_{c_R}$, hence $T \vdash \neg A(\underline{a}, \underline{1})$. □

4.2.2. Undefinability of the notion of truth in formal theories.

LEMMA (Fixed Point Lemma). *Assume that all elementary functions are representable in T. Then for every formula $B(z)$ we can find a closed formula A such that*

$$T \vdash A \leftrightarrow B(\underline{\ulcorner A \urcorner}).$$

PROOF. The proof is very similar to the proof of the Semantical Fixed Point Lemma. Let s be the elementary function introduced there and $A_s(x_1, x_2, x_3)$ a formula representing s in T. Let

$$C := \exists_x(B(x) \wedge A_s(z, z, x)), \quad A := C(\underline{\ulcorner C \urcorner}),$$

and therefore

$$A = \exists_x(B(x) \wedge A_s(\underline{\ulcorner C \urcorner}, \underline{\ulcorner C \urcorner}, x)).$$

Because of $s(\ulcorner C \urcorner, \ulcorner C \urcorner) = \ulcorner C(\underline{\ulcorner C \urcorner}) \urcorner = \ulcorner A \urcorner$ we can prove in T

$$A_s(\underline{\ulcorner C \urcorner}, \underline{\ulcorner C \urcorner}, x) \leftrightarrow x = \underline{\ulcorner A \urcorner},$$

hence by definition of A also

$$A \leftrightarrow \exists_x(B(x) \wedge x = \underline{\ulcorner A \urcorner})$$

and therefore

$$A \leftrightarrow B(\underline{\ulcorner A \urcorner}).$$ □

Note that for $T = \mathrm{Th}(\mathcal{M})$ we obtain the Semantical Fixed Point Lemma above as a special case.

THEOREM. *Let T be a consistent theory such that all elementary functions are representable in T. Then there cannot exist a formula $B(z)$ defining the notion of truth, i.e., such that for all closed formulas A*

$$T \vdash A \leftrightarrow B(\underline{\ulcorner A \urcorner}).$$

PROOF. Assume we would have such a $B(z)$. Consider the formula $\neg B(z)$ and choose by the Fixed Point Lemma a closed formula A such that

$$T \vdash A \leftrightarrow \neg B(\underline{\ulcorner A \urcorner}).$$

For this A we obtain $T \vdash A \leftrightarrow \neg A$, contradicting the consistency of T. □

With $T := \mathrm{Th}(\mathcal{M})$ Tarski's Undefinability Theorem is a special case.

4.3. Undecidability and Incompleteness

Consider a consistent formal theory T with the property that all recursive functions are representable in T. This is a very weak assumption, as we shall show in the next section: it is always satisfied if the theory allows to develop a certain minimum of arithmetic. We shall show that such a theory necessarily is undecidable. First we shall prove a (weak) First Incompleteness Theorem saying that every axiomatized such theory must be incomplete, and then we prove a sharpened form of this theorem due to Gödel and then Rosser, which explicitly provides a closed formula A such that neither A nor $\neg A$ is provable in the theory T.

In this section let $\mathcal{L}$ again be an elementarily presented language with 0, S, $=$ in $\mathcal{L}$ and T a theory containing the equality axioms $\mathrm{Eq}_{\mathcal{L}}$.

4.3.1. Undecidability.

THEOREM. *Assume that T is a consistent theory such that all recursive functions are representable in T. Then T is not recursive.*

PROOF. Assume that T is recursive. By assumption there exists a formula $B(z)$ representing $\ulcorner T \urcorner$ in T. Choose by the Fixed Point Lemma a closed formula A such that

$$T \vdash A \leftrightarrow \neg B(\underline{\ulcorner A \urcorner}).$$

We shall prove $(*)$ $T \nvdash A$ and $(**)$ $T \vdash A$; this is the desired contradiction.

Ad $(*)$. Assume $T \vdash A$. Then $A \in T$, hence $\ulcorner A \urcorner \in \ulcorner T \urcorner$, hence $T \vdash B(\underline{\ulcorner A \urcorner})$ (because $B(z)$ represents in T the set $\ulcorner T \urcorner$). By the choice of A it follows that $T \vdash \neg A$, which contradicts the consistency of T.

Ad $(**)$. By $(*)$ we know $T \nvdash A$. Therefore $A \notin T$, hence $\ulcorner A \urcorner \notin \ulcorner T \urcorner$ and therefore $T \vdash \neg B(\underline{\ulcorner A \urcorner})$. By the choice of A it follows that $T \vdash A$. □

4.3.2. Incompleteness.

THEOREM (First Incompleteness Theorem). *Assume that T is an axiomatized consistent theory with the property that all recursive functions are representable in T. Then T is incomplete.*

PROOF. This is an immediate consequence of the fact that every axiomatized consistent theory which is complete is also recursive (a corollary in 4.1.3), and the Undecidability Theorem above. □

As already mentioned, we now sharpen the Incompleteness Theorem in the sense that we actually produce a formula A such that neither A nor $\neg A$ is provable. Gödel's first incompleteness theorem provided such an A under the assumption that the theory satisfied a stronger condition than mere consistency, namely "ω-consistency". Rosser then improved Gödel's result

by showing, with a somewhat more complicated formula, that consistency is all that is required.

THEOREM (Gödel-Rosser). *Let T be axiomatized and consistent. Assume that there is a formula $L(x,y)$ – written $x<y$ – such that*

$$T \vdash \forall_{x<\underline{n}}(x = \underline{0} \vee \cdots \vee x = \underline{n-1}), \tag{4.4}$$

$$T \vdash \forall_x(x = \underline{0} \vee \cdots \vee x = \underline{n} \vee \underline{n} < x). \tag{4.5}$$

Assume also that every elementary function is representable in T. Then we can find a closed formula A such that neither A nor $\neg A$ is provable in T.

PROOF. We first define $\mathrm{Refut}_T \subseteq \mathbb{N}\times\mathbb{N}$ by

$$\mathrm{Refut}_T(d,a) := \mathrm{Prf}_T(d, \dot{\neg} a).$$

Then Refut_T is elementary and $\mathrm{Refut}_T(d,a)$ if and only if d is the GN of a derivation of the negation of a formula with GN a from a context composed of equality axioms and formulas from Ax_T. Let $B_{\mathrm{Prf}_T}(x_1,x_2)$ and $B_{\mathrm{Refut}_T}(x_1,x_2)$ be formulas representing Prf_T and Refut_T, respectively. Choose by the Fixed Point Lemma a closed formula A such that

$$T \vdash A \leftrightarrow \forall_x(B_{\mathrm{Prf}_T}(x, \underline{\ulcorner A \urcorner}) \to \exists_{y<x} B_{\mathrm{Refut}_T}(y, \underline{\ulcorner A \urcorner})).$$

A expresses its own underivability, in the form (due to Rosser): "For every proof of me there is a shorter proof of my negation".

We shall show $(*)$ $T \nvdash A$ and $(**)$ $T \nvdash \neg A$. Ad $(*)$. Assume $T \vdash A$. Choose n such that

$$\mathrm{Prf}_T(n, \ulcorner A \urcorner).$$

Then we also have

$$\text{not } \mathrm{Refut}_T(m, \ulcorner A \urcorner) \qquad \text{for all } m,$$

since T is consistent. Hence

$$T \vdash B_{\mathrm{Prf}_T}(\underline{n}, \underline{\ulcorner A \urcorner}),$$
$$T \vdash \neg B_{\mathrm{Refut}_T}(\underline{m}, \underline{\ulcorner A \urcorner}) \qquad \text{for all } m.$$

By (4.4) we can conclude

$$T \vdash B_{\mathrm{Prf}_T}(\underline{n}, \underline{\ulcorner A \urcorner}) \wedge \forall_{y<\underline{n}} \neg B_{\mathrm{Refut}_T}(y, \underline{\ulcorner A \urcorner}).$$

Hence

$$T \vdash \exists_x(B_{\mathrm{Prf}_T}(x, \underline{\ulcorner A \urcorner}) \wedge \forall_{y<x} \neg B_{\mathrm{Refut}_T}(y, \underline{\ulcorner A \urcorner})),$$
$$T \vdash \neg A.$$

This contradicts the assumed consistency of T.

Ad $(**)$. Assume $T \vdash \neg A$. Choose n such that

$$\mathrm{Refut}_T(n, \ulcorner A \urcorner).$$

Then we also have

$$\text{not } \mathrm{Prf}_T(m, \ulcorner A \urcorner) \qquad \text{for all } m,$$

since T is consistent. Hence

$$T \vdash B_{\mathrm{Refut}_T}(\underline{n}, \underline{\ulcorner A \urcorner}),$$
$$T \vdash \neg B_{\mathrm{Prf}_T}(\underline{m}, \underline{\ulcorner A \urcorner}) \qquad \text{for all } m.$$

This implies

$$T \vdash \forall_x (B_{\mathrm{Prf}_T}(x, \underline{\ulcorner A \urcorner}) \to \exists_{y<x} B_{\mathrm{Refut}_T}(y, \underline{\ulcorner A \urcorner})),$$

as can be seen easily by cases on x, using (4.5). Hence $T \vdash A$. But this again contradicts the assumed consistency of T. □

Finally we formulate a variant of this theorem which does not assume that the theory T talks about numbers only. Call T a *theory with defined natural numbers* if there is a formula $N(x)$ – written Nx – such that $T \vdash N0$ and $T \vdash \forall_{x \in N} N(\mathrm{S}x)$ where $\forall_{x \in N} A$ is short for $\forall_x (Nx \to A)$. Representing a function in such a theory of course means that the free variables in (4.3) are relativized to N:

$$T \vdash \forall_{y,z \in N} (A(\underline{a_1}, \ldots, \underline{a_n}, y) \wedge A(\underline{a_1}, \ldots, \underline{a_n}, z) \to y{=}z) \text{ for all } a_1, \ldots, a_n \in \mathbb{N}.$$

Theorem (Gödel-Rosser). *Assume that T is an axiomatized consistent theory with defined natural numbers, and that there is a formula $L(x, y)$ – written $x < y$ – such that*

$$T \vdash \forall_{x \in N} (x < \underline{n} \to x = \underline{0} \vee \ \cdots \vee \ x = \underline{n-1}),$$
$$T \vdash \forall_{x \in N} (x = \underline{0} \vee \ \cdots \vee \ x = \underline{n} \vee \ \underline{n} < x).$$

Assume also that every elementary function is representable in T. Then one can find a closed formula A such that neither A nor $\neg A$ is provable in T.

Proof. As for the Gödel-Rosser Theorem above; just relativize all quantifiers to N. □

4.4. Representability

We show in this section that already very simple theories have the property that all recursive functions are representable in them.

4.4.1. Weak arithmetical theories.

THEOREM. *Let $\mathcal{L}$ be an elementarily presented language with* 0, S, $=$ *in $\mathcal{L}$ and T a consistent theory with defined natural numbers containing the equality axioms $\mathrm{Eq}_{\mathcal{L}}$ and the ex-falso-quodlibet axiom $\forall_{x,y\in N}(\bot \to x = y)$. Assume that there is a formula $L(x, y)$ – written $x < y$ – such that*

$$T \vdash \mathrm{S}\underline{a} \neq 0 \qquad \text{for all } a \in \mathbb{N}, \tag{4.6}$$

$$T \vdash \mathrm{S}\underline{a} = \mathrm{S}\underline{b} \to \underline{a} = \underline{b} \qquad \text{for all } a, b \in \mathbb{N}, \tag{4.7}$$

(4.8) *the functions $+$ and $\cdot$ are representable in T,*

$$T \vdash \forall_{x\in N}(x \not< 0), \tag{4.9}$$

$$T \vdash \forall_{x\in N}(x < \mathrm{S}\underline{b} \to x < \underline{b} \vee x = \underline{b}) \qquad \text{for all } b \in \mathbb{N}, \tag{4.10}$$

$$T \vdash \forall_{x\in N}(x < \underline{b} \vee x = \underline{b} \vee \underline{b} < x) \qquad \text{for all } b \in \mathbb{N}. \tag{4.11}$$

Then T fulfills the assumptions of the Gödel-Rosser Theorem relativized to N, i.e.,

$$T \vdash \forall_{x\in N}(x < \underline{a} \to x = \underline{0} \vee \cdots \vee x = \underline{a-1}) \qquad \text{for all } a \in \mathbb{N}, \tag{4.12}$$

$$T \vdash \forall_{x\in N}(x = \underline{0} \vee \cdots \vee x = \underline{a} \vee \underline{a} < x) \qquad \text{for all } a \in \mathbb{N}, \tag{4.13}$$

and every recursive function is representable in T.

PROOF. (4.12) can be proved easily by induction on a. The base case follows from (4.9), and the step from the induction hypothesis and (4.10). (4.13) immediately follows from the trichotomy law (4.11), using (4.12).

For the representability of recursive functions, first note that the formulas $x = y$ and $x < y$ actually do represent in T the equality and the less-than relations, respectively. From (4.6) and (4.7) we can see immediately that $T \vdash \underline{a} \neq \underline{b}$ when $a \neq b$. Assume $a \not< b$. We show $T \vdash \underline{a} \not< \underline{b}$ by induction on b. $T \vdash \underline{a} \not< 0$ follows from (4.9). In the step we have $a \not< b + 1$, hence $a \not< b$ and $a \neq b$, hence by induction hypothesis and the representability (above) of the equality relation, $T \vdash \underline{a} \not< \underline{b}$ and $T \vdash \underline{a} \neq \underline{b}$, hence by (4.10) $T \vdash \underline{a} \not< \mathrm{S}\underline{b}$. Now assume $a < b$. Then $T \vdash \underline{a} \neq \underline{b}$ and $T \vdash \underline{b} \not< \underline{a}$, hence by (4.11) $T \vdash \underline{a} < \underline{b}$.

We now show by induction on the definition of μ-recursive functions, that every recursive function is representable in T. Recall (from 4.2.1) that the second condition (4.2) in the definition of representability of a function automatically follows from the other two (and hence need not be checked further). This is because $T \vdash \underline{a} \neq \underline{b}$ for $a \neq b$.

The *initial functions* constant 0, successor and projection (onto the i-th coordinate) are trivially represented by the formulas $0 = y$, $\mathrm{S}x = y$ and $x_i = y$ respectively. Addition and multiplication are represented in

T by assumption. Recall that the one remaining initial function of μ-recursiveness is $\dot{-}$, but this is definable from the characteristic function of $<$ by $a \dot{-} b = \mu_i(b + i \geq a) = \mu_i(c_<(b + i, a) = 0)$. We now show that the characteristic function of $<$ is representable in T. (It will then follow that $\dot{-}$ is representable, once we have shown that the representable functions are closed under μ.) We show that

$$A(x_1, x_2, y) := (x_1 < x_2 \wedge y = 1) \vee (x_1 \not< x_2 \wedge y = 0)$$

represents $c_<$. First notice that $\forall_{y,z \in N}(A(\underline{a_1}, \underline{a_2}, y) \wedge A(\underline{a_1}, \underline{a_2}, z) \to y = z)$ already follows logically from the equality axiom and the ex-falso-quodlibet axiom for equality (by cases on the alternatives of A). Assume $a_1 < a_2$. Then $T \vdash \underline{a_1} < \underline{a_2}$, hence $T \vdash A(\underline{a_1}, \underline{a_2}, 1)$. Now assume $a_1 \not< a_2$. Then $T \vdash \underline{a_1} \not< \underline{a_2}$, hence $T \vdash A(\underline{a_1}, \underline{a_2}, 0)$.

For the *composition* case, suppose f is defined from $h, g_1, \ldots, g_m$ by

$$f(\vec{a}\,) = h(g_1(\vec{a}\,), \ldots, g_m(\vec{a}\,)).$$

By induction hypothesis we already have representing formulas $A_{g_i}(\vec{x}, y_i)$ and $A_h(\vec{y}, z)$. As representing formula for f we take

$$A_f := \exists_{\vec{y} \in N}(A_{g_1}(\vec{x}, y_1) \wedge \cdots \wedge A_{g_m}(\vec{x}, y_m) \wedge A_h(\vec{y}, z)).$$

Assume $f(\vec{a}\,) = c$. Then there are $b_1, \ldots, b_m$ such that $T \vdash A_{g_i}(\underline{\vec{a}}, \underline{b_i})$ for each i, and $T \vdash A_h(\underline{\vec{b}}, \underline{c})$ so by logic $T \vdash A_f(\underline{\vec{a}}, \underline{c})$. It remains to show uniqueness $T \vdash \forall_{z_1, z_2 \in N}(A_f(\underline{\vec{a}}, z_1) \wedge A_f(\underline{\vec{a}}, z_2) \to z_1 = z_2)$. But this follows by logic from the induction hypothesis for g_i, which gives

$$T \vdash \forall_{y_{1i}, y_{2i} \in N}(A_{g_i}(\underline{\vec{a}}, y_{1i}) \wedge A_{g_i}(\underline{\vec{a}}, y_{2i}) \to y_{1i} = y_{2i} = \underline{g_i(\vec{a}\,)})$$

and the induction hypothesis for h, which gives

$$T \vdash \forall_{z_1, z_2 \in N}(A_h(\underline{\vec{b}}, z_1) \wedge A_h(\underline{\vec{b}}, z_2) \to z_1 = z_2) \quad \text{with } b_i = g_i(\vec{a}\,).$$

For the μ case, suppose f is defined from g (taken here to be binary for notational convenience) by $f(a) = \mu_i(g(i, a) = 0)$, assuming $\forall_a \exists_i (g(i, a) = 0)$. By induction hypothesis we have a formula $A_g(y, x, z)$ representing g. In this case we represent f by the formula

$$A_f(x, y) := Ny \wedge A_g(y, x, 0) \wedge \forall_{v \in N}(v < y \to \exists_{u \in N; u \neq 0} A_g(v, x, u)).$$

We first show the representability condition (4.1), that is $T \vdash A_f(\underline{a}, \underline{b})$ when $f(a) = b$. Because of the form of A_f this follows from the assumed representability of g together with $T \vdash \forall_{v \in N}(v < \underline{b} \to v = \underline{0} \vee \cdots \vee v = \underline{b - 1})$.

We now tackle the uniqueness condition (4.3). Given a, let $b := f(a)$ (thus $g(b, a) = 0$ and b is the least such). It suffices to show

$$T \vdash \forall_{y \in N}(A_f(\underline{a}, y) \to y = \underline{b}).$$

We prove $T \vdash \forall_{y\in N}(y < \underline{b} \to \neg A_f(\underline{a}, y))$ and $T \vdash \forall_{y\in N}(\underline{b} < y \to \neg A_f(\underline{a}, y))$, and then appeal to the trichotomy law and the ex-falso-quodlibet axiom for equality.

We first show $T \vdash \forall_{y\in N}(y < \underline{b} \to \neg A_f(\underline{a}, y))$. Now since, for any $i < b$, $T \vdash \neg A_g(\underline{i}, \underline{a}, 0)$ by the assumed representability of g, we obtain immediately $T \vdash \neg A_f(\underline{a}, \underline{i})$. Hence because of $T \vdash \forall_{y\in N}(y < \underline{b} \to y = \underline{0} \vee \cdots \vee y = \underline{b-1})$ the claim follows.

Secondly, $T \vdash \forall_{y\in N}(\underline{b} < y \to \neg A_f(\underline{a}, y))$ follows almost immediately from $T \vdash \forall_{y\in N}(\underline{b} < y \to A_f(\underline{a}, y) \to \exists_{u\in N;u\neq 0} A_g(\underline{b}, \underline{a}, u))$ and the uniqueness for g, $T \vdash \forall_{u\in N}(A_g(\underline{b}, \underline{a}, u) \to u = 0)$. □

4.4.2. Robinson's theory Q. We conclude this section by considering a special and particularly simple arithmetical theory due originally to Robinson (1950). Let $\mathcal{L}_1$ be the language given by 0, S, $+$, $\cdot$ and $=$, and let Q be the theory determined by the axioms $\mathrm{Eq}_{\mathcal{L}_1}$, ex-falso-quodlibet for equality $\bot \to x = y$ and

$$\mathrm{S}x \neq 0, \tag{4.14}$$

$$\mathrm{S}x = \mathrm{S}y \to x = y, \tag{4.15}$$

$$x + 0 = x, \tag{4.16}$$

$$x + \mathrm{S}y = \mathrm{S}(x + y), \tag{4.17}$$

$$x \cdot 0 = 0, \tag{4.18}$$

$$x \cdot \mathrm{S}y = x \cdot y + x, \tag{4.19}$$

$$\exists_z(x + \mathrm{S}z = y) \vee x = y \vee \exists_z(y + \mathrm{S}z = x). \tag{4.20}$$

THEOREM (Robinson's Q). *Every consistent theory $T \supseteq Q$ fulfills the assumptions of the Gödel-Rosser Theorem w.r.t. the definition $L(x, y) := \exists_z(x + \mathrm{S}z = y)$ of the $<$-relation. In particular, every recursive function is representable in T.*

PROOF. We show that T satisfies the conditions of the previous theorem. For (4.6) and (4.7) this is clear. For (4.8) we can take $x+y = z$ and $x\cdot y = z$ as representing formulas. For (4.9) we have to show $\neg\exists_z(x + \mathrm{S}z = 0)$; this follows from (4.17) and (4.14). For the proof of (4.10) we need the auxiliary proposition

$$x = 0 \vee \exists_y(x = 0 + \mathrm{S}y), \tag{4.21}$$

which will be attended to below. Assume $x+\mathrm{S}z = \mathrm{S}\underline{b}$, hence also $\mathrm{S}(x+z) = \mathrm{S}\underline{b}$ and therefore $x + z = \underline{b}$. We must show $\exists_{y'}(x + \mathrm{S}y' = \underline{b}) \vee x = \underline{b}$. But this follows from (4.21) for z. In case $z = 0$ we obtain $x = \underline{b}$, and in case $\exists_y(z = 0 + \mathrm{S}y)$ we have $\exists_{y'}(x + \mathrm{S}y' = \underline{b})$, since $0 + \mathrm{S}y = \mathrm{S}(0 + y)$. Thus (4.10) is proved. (4.11) follows immediately from (4.20). For the proof of

(4.21) we use (4.20) with $y = 0$. It clearly suffices to exclude the first case $\exists_z(x + \mathrm{S}z = 0)$. But this means $\mathrm{S}(x + z) = 0$, contradicting (4.14). □

COROLLARY (Essential undecidability of Q). *Every consistent theory $T \supseteq Q$ in an elementarily presented language is non-recursive.*

PROOF. This follows from the theorem above and the Undecidability Theorem in 4.3.1. □

COROLLARY (Undecidability of logic). *The set of formulas derivable in the classical fragment of minimal logic is non-recursive.*

PROOF. Otherwise Q would be recursive, because a formula A is derivable in Q if and only if the implication $B \to A$ is derivable, where B is the conjunction of the finitely many axioms and equality axioms of Q. □

REMARK. Note that it suffices that the underlying language contains one binary relation symbol (for $=$), one constant symbol (for 0), one unary function symbol (for S) and two binary functions symbols (for $+$ and $\cdot$). The study of decidable fragments of first order logic is one of the oldest research areas of Mathematical Logic. For more information see Börger et al. (1997).

4.4.3. Σ_1-formulas. Reading the above proof of representability, one can see that the representing formulas used are of a restricted form, having no unbounded universal quantifiers and therefore defining Σ^0_1-relations. This will be of crucial importance for our proof of Gödel's Second Incompleteness Theorem to follow, but in addition we need to make a syntactically precise definition of the class of formulas involved, more specific and apparently more restrictive than the notion of Σ_1-formula used earlier. However, as proved in the corollary below, we can still represent all recursive functions even in the weak theory Q by means of Σ_1-formulas in this more restrictive sense. Consequently provable Σ_1-ness will be the same whichever definition we take.

DEFINITION. For the remainder of this chapter, the Σ_1-formulas of the language $\mathcal{L}_1$ will be those generated inductively by the following clauses:

(a) Only atomic formulas of the restricted forms $x = y$, $x \neq y$, $0 = x$, $\mathrm{S}x = y$, $x + y = z$ and $x \cdot y = z$ are allowed as Σ_1-formulas.
(b) If A and B are Σ_1-formulas, then so are $A \wedge B$ and $A \vee B$.
(c) If A is a Σ_1-formula, then so is $\forall_{x<y}A$, which is an abbreviation for $\forall_x(\exists_z(x + \mathrm{S}z = y) \to A)$.
(d) If A is a Σ_1-formula, then so is $\exists_x A$.

COROLLARY. *Every recursive function is representable in Q by a Σ_1-formula in the language $\mathcal{L}_1$.*

PROOF. This can be seen immediately by inspecting the proof of the theorem above on weak arithmetical theories. Only notice that because of the equality axioms $\exists_z(x+\mathrm{S}z = y)$ is equivalent to $\exists_z\exists_w(\mathrm{S}z = w \wedge x+w = y)$ and $A(0)$ is equivalent to $\exists_x(0 = x \wedge A(x))$. □

4.5. Unprovability of Consistency

We have seen in the theorem of Gödel-Rosser how, for every axiomatized consistent theory T safisfying certain weak assumptions, we can construct an undecidable sentence A meaning "For every proof of me there is a shorter proof of my negation". Because A is unprovable, it is clearly true.

Gödel's Second Incompleteness Theorem provides a particularly interesting alternative to A, namely a formula Con_T expressing the consistency of T. Again it turns out to be unprovable and therefore true. We shall prove this theorem in a sharpened form due to Löb.

4.5.1. Formalized Σ_1-completeness. We prove an auxiliary proposition, expressing the completeness of Q with respect to Σ_1-formulas.

LEMMA. *Let $A(x_1, \ldots, x_n)$ be a Σ_1-formula in the language $\mathcal{L}_1$ determined by* 0, S, $+$, $\cdot$ *und* $=$. *Assume that $\mathcal{N}_1 \models A(\underline{a_1}, \ldots, \underline{a_n})$ where $\mathcal{N}_1$ is the standard model of $\mathcal{L}_1$. Then $Q \vdash A(\underline{a_1}, \ldots, \underline{a_n})$.*

PROOF. By induction on the Σ_1-formulas of the language $\mathcal{L}_1$. For atomic formulas, the cases have been dealt with either in the earlier parts of the proof of the theorem above on weak arithmetical theories, or (for $x + y = z$ and $x \cdot y = z$) they follow from the recursion equations (4.16) - (4.19).

Cases $A \wedge B$, $A \vee B$. The claim follows immediately from the induction hypothesis.

Case $\forall_{x<y}A(x, y, z_1, \ldots, z_n)$; for simplicity assume $n = 1$. Suppose $\mathcal{N}_1 \models (\forall_{x<y}A)(\underline{b}, \underline{c})$. Then also $\mathcal{N}_1 \models A(\underline{i}, \underline{b}, \underline{c})$ for each $i < b$ and hence by induction hypothesis $Q \vdash A(\underline{i}, \underline{b}, \underline{c})$. Now by the theorem above on Robinson's Q

$$Q \vdash \forall_{x<\underline{b}}(x = \underline{0} \vee \cdots \vee x = \underline{b-1}),$$

hence

$$Q \vdash (\forall_{x<y}A)(\underline{b}, \underline{c}).$$

Case $\exists_x A(x, y_1, \ldots, y_n)$; for simplicity take $n = 1$. Assume $\mathcal{N}_1 \models (\exists_x A)(\underline{b})$. Then $\mathcal{N}_1 \models A(\underline{a}, \underline{b})$ for some $a \in \mathbb{N}$, hence by induction hypothesis $Q \vdash A(\underline{a}, \underline{b})$ and therefore $Q \vdash (\exists_x A)(\underline{b})$. □

LEMMA (Formalized Σ_1-Completeness). *In an appropriate theory T of arithmetic with induction, we can formally prove for any Σ_1-formula A*

$$A(\vec{x}\,) \to \exists_p \mathrm{Prf}_T(p, \ulcorner A(\dot{\vec{x}}\,) \urcorner).$$

Here $\mathrm{Prf}_T(p,z)$ *is a suitable* Σ_1*-formula which represents in Robinson's* Q *the recursive relation "a is the Gödel number of a proof in* T *of the formula with Gödel number* b*". Also* $\ulcorner A(\dot{x})\urcorner$ *is a term which represents, in* Q*, the numerical function mapping a number* a *to the Gödel number of* $A(\underline{a})$*.*

PROOF. We have not been precise about the theory T in which this result is to be formalized, but we shall content ourselves at this stage with merely pointing out, as we proceed, the basic properties that are required. Essentially T will be an extension of Q, together with induction formalized by the axiom schema

$$B(0) \wedge \forall_x(B(x) \to B(\mathrm{S}x)) \to \forall_x B(x)$$

and it will be assumed that T has sufficiently many basic functions available to deal with the construction of appropriate Gödel numbers.

The proof proceeds by induction on the build-up of the Σ_1-formula $A(\vec{x}\,)$.

We consider three atomic cases, leaving the others to the reader. Suppose $A(x)$ is the formula $0 = x$. We show $T \vdash 0 = x \to \exists_p \mathrm{Prf}_T(p, \ulcorner 0 = \dot{x}\urcorner)$, by induction on x. The base case merely requires the construction of a numeral representing the Gödel number of the axiom $0 = 0$, and the induction step is trivial because $T \vdash \mathrm{S}x \neq 0$. Secondly suppose A is the formula $x + y = z$. We show $T \vdash \forall_z(x + y = z \to \exists_p \mathrm{Prf}_T(p, \ulcorner \dot{x} + \dot{y} = \dot{z}\urcorner))$ by induction on y. If $y = 0$, the assumption gives $x = z$ and one requires only the Gödel number for the axiom $\forall_x(x + 0 = x)$ which, when applied to the Gödel number of the x-th numeral, gives $\exists_p \mathrm{Prf}_T(p, \ulcorner \dot{x} + 0 = \dot{z}\urcorner)$. If y is a successor $\mathrm{S}u$, then the assumption gives $z = \mathrm{S}v$ where $x + u = v$, so by the induction hypothesis we already have a p such that $\mathrm{Prf}_T(p, \ulcorner \dot{x} + \dot{u} = \dot{v}\urcorner)$. Applying the successor to both sides, one then easily obtains from p a p' such that $\mathrm{Prf}_T(p', \ulcorner \dot{x} + \dot{y} = \dot{z}\urcorner)$. Thirdly suppose A is the formula $x \neq y$. We show $T \vdash \forall_y(x \neq y \to \exists_p \mathrm{Prf}_T(p, \ulcorner \dot{x} \neq \dot{y}\urcorner))$ by induction on x. The base case $x = 0$ requires a subinduction on y. If $y = 0$, then the claim is trivial (by ex-falso). If $y = \mathrm{S}u$, we have to produce a Gödel number p such that $\mathrm{Prf}_T(p, \ulcorner 0 \neq \mathrm{S}\dot{u}\urcorner)$, but this is just an axiom. Now consider the step case $x = \mathrm{S}v$. Again we need an auxiliary induction on y. Its base case is dealt with exactly as before, and when $y = \mathrm{S}u$ it uses the induction hypothesis for $v \neq u$ together with the injectivity of the successor.

The cases where A is built up by conjunction or disjunction are rather trivial. One only requires, for example in the conjunction case, a function which combines the Gödel numbers of the proofs of the separate conjuncts into a single Gödel number of a proof of the conjunction A itself.

Now consider the case $\exists_y A(y,x)$ (with just one parameter x for simplicity). By the induction hypothesis we already have $T \vdash A(y,x) \to \exists_p \mathrm{Prf}_T(p, \ulcorner A(\dot{y},\dot{x})\urcorner)$. But any Gödel number p such that $\mathrm{Prf}_T(p, \ulcorner A(\dot{y},\dot{x})\urcorner)$

can easily be transformed (by formally applying the $\exists^+$-rule) into a Gödel number p' such that $\mathrm{Prf}_T(p', \ulcorner \exists_y A(y, \dot{x}) \urcorner)$. Therefore we obtain as required, $T \vdash \exists_y A(y,x) \to \exists_{p'} \mathrm{Prf}_T(p', \ulcorner \exists_y A(y,\dot{x}) \urcorner)$.

Finally suppose the Σ_1-formula is of the form $\forall_{u<y} A(u,x)$. We show

$$T \vdash \forall_{u<y} A(u,x) \to \exists_p \mathrm{Prf}_T(p, \ulcorner \forall_{u<\dot{y}} A(u,\dot{x}) \urcorner).$$

By the induction hypothesis

$$T \vdash A(u,x) \to \exists_p \mathrm{Prf}_T(p, \ulcorner A(\dot{u},\dot{x}) \urcorner)$$

so by logic

$$T \vdash \forall_{u<y} A(u,x) \to \forall_{u<y} \exists_p \mathrm{Prf}_T(p, \ulcorner A(\dot{u},\dot{x}) \urcorner).$$

The required result now follows immediately from the auxiliary lemma:

$$T \vdash \forall_{u<y} \exists_p \mathrm{Prf}_T(p, \ulcorner A(\dot{u},\dot{x}) \urcorner) \to \exists_q \mathrm{Prf}_T(q, \ulcorner \forall_{u<\dot{y}} A(u,\dot{x}) \urcorner).$$

It remains only to prove this, which we do by induction on y (inside T). In case $y = 0$ a proof of $u < 0 \to A$ is trivial (by ex-falso-quodlibet, since A is a Σ_1-formula), so the required Gödel number q is easily constructed. For the step case $y = \mathrm{S}z$ the assumption gives $\forall_{u<z} \exists_p \mathrm{Prf}_T(p, \ulcorner A(\dot{u},\dot{x}) \urcorner)$, from which follows $\exists_q \mathrm{Prf}_T(q, \ulcorner \forall_{u<\dot{z}} A(u,\dot{x}) \urcorner)$ by the induction hypothesis. Also $\exists_{p'} \mathrm{Prf}_T(p', \ulcorner A(\dot{z},\dot{x}) \urcorner)$. Now we only have to combine p' and q to obtain (by means of an appropriate "simple" function) a Gödel number q' so that $\mathrm{Prf}_T(q', \ulcorner \forall_{u<\dot{y}} A(u,\dot{x}) \urcorner)$. □

4.5.2. Derivability conditions. Let T be an axiomatized consistent theory with $T \supseteq Q$, and possessing "enough" induction to formalize Σ_1-completeness as we have just done. Define, from the associated formula Prf_T, the following $\mathcal{L}_1$-formulas:

$$\begin{aligned} \mathrm{Thm}_T(x) &:= \exists_y \mathrm{Prf}_T(y,x), \\ \mathrm{Con}_T &:= \neg \exists_y \mathrm{Prf}_T(y, \underline{\ulcorner \bot \urcorner}). \end{aligned}$$

Then $\mathrm{Thm}_T(x)$ defines in $\mathcal{N}_1$ the set of formulas provable in T, and we have $\mathcal{N}_1 \models \mathrm{Con}_T$ if and only if T is consistent. We write $\Box A$ for $\mathrm{Thm}_T(\underline{\ulcorner A \urcorner})$; hence Con_T can be written $\neg\Box\bot$. Now consider the following two *derivability conditions* for T, due to Hilbert and Bernays (1970):

(4.22) $\quad T \vdash A \to \Box A \quad (A \text{ closed } \Sigma_1\text{-formula of the language } \mathcal{L}_1)$,

(4.23) $\quad T \vdash \Box(A \to B) \to \Box A \to \Box B$.

(4.22) is just a special case of formalized Σ_1-completeness for closed formulas, and (4.23) requires only that the theory T has a term that constructs, from the Gödel number of a proof of $A \to B$ and the Gödel number of a proof of A, the Gödel number of a proof of B, and furthermore this fact must be provable in T.

THEOREM (Gödel's Second Incompleteness Theorem). *Let T be an axiomatized consistent extension of Q, satisfying the derivability conditions* (4.22) *und* (4.23). *Then $T \nvdash \mathrm{Con}_T$.*

PROOF. Let $C := \bot$ in Löb's Theorem below, which is a generalization of Gödel's original result. □

THEOREM (Löb). *Let T be an axiomatized consistent extension of Q satisfying the derivability conditions* (4.22) *and* (4.23). *Then for any closed $\mathcal{L}_1$-formula C, if $T \vdash \Box C \to C$, then already $T \vdash C$.*

PROOF. Assume $T \vdash \Box C \to C$. We must show $T \vdash C$. Choose A by the Fixed Point Lemma such that

$$Q \vdash A \leftrightarrow (\Box A \to C). \tag{4.24}$$

First we show $T \vdash \Box A \to C$. We obtain

$T \vdash A \to \Box A \to C$	by (4.24)
$T \vdash \Box(A \to \Box A \to C)$	by Σ_1-completeness
$T \vdash \Box A \to \Box(\Box A \to C)$	by (4.23)
$T \vdash \Box A \to \Box\Box A \to \Box C$	again by (4.23)
$T \vdash \Box A \to \Box C$	since $T \vdash \Box A \to \Box\Box A$ by (4.22).

Therefore the assumption $T \vdash \Box C \to C$ implies $T \vdash \Box A \to C$. Hence $T \vdash A$ by (4.24), and then $T \vdash \Box A$ by Σ_1-completeness. But $T \vdash \Box A \to C$ as we have just shown, therefore $T \vdash C$. □

REMARK. It follows that if T is any axiomatized consistent extension of Q satisfying the derivability conditions (4.22) und (4.23), then the reflection schema

$$\Box C \to C \quad \text{for closed } \mathcal{L}_1\text{-formulas } C$$

is not derivable in T. For by Löb's Theorem, it cannot be derivable when C is underivable.

By adding to Q the induction schema for all formulas we obtain *Peano-Arithmetic* PA, which is the most natural example of a theory T to which the results above apply. However, various weaker fragments of PA, obtained by restricting the classes of induction formulas, would serve equally well as examples of such T.

4.6. Notes

The fundamental paper on incompleteness is Gödel (1931). This paper already contains the β-function crucially needed for the representation theorem; the fixed point lemma is used implicitly. Gödel's first incompleteness

theorem uses the formula "I am not provable", a fixed point of $\neg \mathrm{Thm}_T(x)$. To prove independence of this proposition from the underlying theory T one needs ω-consistency of T (which is automatically fulfilled if T is a subtheory of the theory of the standard model). Rosser (1936) found the sharpening presented here, using the formula "For every proof of me there is a shorter proof of my negation". Löb's theorem is based on the formula A, which says "If I am provable, then C". Undefinability of the notion of truth was proved originally by Tarski (1939), and undecidability of predicate logic is a result of Church (1936). The arithmetical theory Q is due to Robinson (1950).

There is also much more work on general reflection principles, which we only have touched in the most simple case. One must mention here Smoryński (1991), Feferman (1960) and Girard (1987).

The volumes of Gödel's Collected Works edited by Feferman et al. (1986, 1990, 1995, 2002, 2002) provide excellent commentaries on his massive contributions to logic.

CHAPTER 5

Set Theory

5.1. Cumulative Type Structures

Set theory can be viewed as a framework within which mathematics can be given a foundation. Here we want to develop set theory as a formal theory within mathematical logic. But first it is necessary to have an intuitive picture of the notion of a set, to be described by the axioms.

5.1.1. Cantor's definition. Cantor in 1895 gave the following definition:

> Unter einer "Menge" verstehen wir jede Zusammenfassung M von bestimmten wohlunterschiedenen Objekten m unserer Anschauung oder unseres Denkens (welche die Elemente von M genannt werden) zu einem Ganzen.

One can try to make this definition more precise, as follows. Let V be the collection of all objects "unserer Anschauung oder unseres Denkens". Let $A(x)$ denote properties of objects x from V. Then one can form the set $\{ x \mid A(x) \}$, the set of all objects x of V with the property $A(x)$. According to Cantor's definition $\{ x \mid A(x) \}$ is again an object in V.

Examples for properties: (1) x is a natural number. (2) x is a set. (3) x is a point, y is a line and x lies on y. (4) y is a set and x is an element of y, shortly: $\mathrm{Set}(y) \wedge x \in y$.

However, Cantor's definition cannot be accepted in its original form, for it leads to contradictions. The most well known is *Russell's antinomy*: Let $x_0 := \{ x \mid \mathrm{Set}(x) \wedge x \notin x \}$. Then

$$x_0 \in x_0 \leftrightarrow \mathrm{Set}(x_0) \wedge x_0 \notin x_0 \leftrightarrow x_0 \notin x_0,$$

for x_0 is a set.

5.1.2. Shoenfield's principle. The root for this contradiction is the fact that in Cantor's definition we accept the concept of a finished totality of all sets. However, this is neither necessary nor does it mirror the usual practice of mathematics. It completely suffices to form a set only if all its elements "are available" already. This leads to the concept of a stepwise construction of sets, or more precisely to the *cumulative type structure*: We

start with certain "urelements", that form the sets of level 0. Then on an arbitrary level we can form all sets whose elements belong to earlier levels.

If for instance we take as urelements the natural numbers, then $\{27, \{5\}\}$ belongs to level 2.

The following natural questions pose themselves: (1) Which urelements should we choose? (2) How far do the levels reach?

Ad (1). For the purposes of mathematics it is completely sufficient not to assume any urelements at all; then one speaks of *pure sets*. This will be done in the following.

Level 0:	$-$
Level 1:	$\emptyset$
Level 2:	$\emptyset, \{\emptyset\}$
Level 3:	$\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \{\emptyset, \{\emptyset\}\}$
	and so on.

Ad (2). Shoenfield (1967) formulated the following principle:

> Consider a collection $\mathcal{S}$ of levels. If a situation can be conceived where all the levels from $\mathcal{S}$ are constructed, then there exists a level which is past all those levels.

From this admittedly rather vague principle we shall draw exact consequences, which will be fixed as axioms.

By a *set* we intuitively understand an object that belongs to some level of the cumulative type structure. By a *class* we mean an arbitrary collection of sets.

So every set clearly is a class. Moreover there are classes that are not sets, for instance the class V of all sets.

5.2. Axiomatic Set Theory

In set theory – as in any axiomatic theory – we have to explicitly state all used properties, including the "obvious" ones.

5.2.1. Extensionality, equality. The language of set theory has a single non-logical symbol, the *element* relation $\in$. So the only atomic formulas are of the form $x \in y$ (x is an element of y). Equality $x = y$ is defined by

$$(x = y) := \forall_z (z \in x \leftrightarrow z \in y).$$

To ensure compatibility of the $\in$-relation with equality we need an axiom:

Axiom (Extensionality).

$$x = y \to x \in z \to y \in z.$$

REMARK. If alternatively equality is to be used as a primitive symbol, one must require the equality axioms and in addition

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

As classes in our axiomatic theory we only allow *definable* collections of sets. By "definable" we mean definable by a formula in the language of set theory. More precisely: If $A(x)$ is a formula, then

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

denotes the *class* of all sets x with the property $A(x)$.

Instead of classes we could have used properties or more precisely formulas as well. However, classes allow for a simpler and more suggestive formulation of many of the propositions we want to consider.

If $A(x)$ is the formula $x = x$, then $\{\, x \mid A(x) \,\}$ is called the *all class* or the (set theoretic) *universe*, denoted by V. If $A(x)$ is the formula $x \notin x$, then $\{\, x \mid A(x) \,\}$ is called the *Russell class*.

We now give some definitions that will be used all over in the following. A set b is an element of the class $\{\, x \mid A(x) \,\}$ if $A(b)$ holds:

$$b \in \{\, x \mid A(x) \,\} := A(b).$$

Two classes $\mathcal{A}$, $\mathcal{B}$ are *equal* if they have the same elements:

$$(\mathcal{A} = \mathcal{B}) := \forall_x(x \in \mathcal{A} \leftrightarrow x \in \mathcal{B}).$$

If $\mathcal{A}$ is a class and b a set, then $\mathcal{A}$ and b are called equal if they have the same elements:

$$(\mathcal{A} = b) := \forall_x(x \in \mathcal{A} \leftrightarrow x \in b).$$

In this case we identify the class $\mathcal{A}$ with this set b. Instead of "$\mathcal{A}$ is set" we also write $\mathcal{A} \in V$. A class $\mathcal{B}$ is an element of a set a (of a class $\mathcal{A}$, resp.) if $\mathcal{B}$ is equal to an element x of a (of $\mathcal{A}$, resp.).

$$(\mathcal{B} \in a) := \exists_x(x \in a \wedge \mathcal{B} = x),$$
$$(\mathcal{B} \in \mathcal{A}) := \exists_x(x \in \mathcal{A} \wedge \mathcal{B} = x).$$

A class $\mathcal{A}$ is a *proper class* if $\mathcal{A}$ is not a set:

$$\mathcal{A} \text{ proper class} := \forall_x(x \neq \mathcal{A}).$$

REMARK. Every set b is a class, since

$$b = \{\, x \mid x \in b \,\}.$$

The Russell class is a proper class, for if $\{\, x \mid x \notin x \,\} = x_0$, we would have

$$x_0 \in x_0 \leftrightarrow x_0 \notin x_0.$$

So the Russell construction is not an antinomy any more, but simply says that there are sets and (proper) classes.

Let $\mathcal{A}$, $\mathcal{B}$ be classes (proper classes or sets) and $a, b, a_1, \ldots, a_n$ sets. We define

$$\{a_1, \ldots, a_n\} := \{\, x \mid x = a_1 \vee \cdots \vee x = a_n \,\},$$

$\emptyset := \{\, x \mid x \neq x \,\}$	empty class,
$V := \{\, x \mid x = x \,\}$	all class,
$\mathcal{A} \subseteq \mathcal{B} := \forall_x(x \in \mathcal{A} \to x \in \mathcal{B})$	$\mathcal{A}$ is subclass of $\mathcal{B}$,
$\mathcal{A} \subsetneq \mathcal{B} := \mathcal{A} \subseteq \mathcal{B} \wedge \mathcal{A} \neq \mathcal{B}$	$\mathcal{A}$ is proper subclass of $\mathcal{B}$,
$\mathcal{A} \cap \mathcal{B} := \{\, x \mid x \in \mathcal{A} \wedge x \in \mathcal{B} \,\}$	intersection,
$\mathcal{A} \cup \mathcal{B} := \{\, x \mid x \in \mathcal{A} \vee x \in \mathcal{B} \,\}$	union,
$\mathcal{A} \setminus \mathcal{B} := \{\, x \mid x \in \mathcal{A} \wedge x \notin \mathcal{B} \,\}$	difference,
$\bigcup \mathcal{A} := \{\, x \mid \exists_y(y \in \mathcal{A} \wedge x \in y) \,\}$	big union,
$\bigcap \mathcal{A} := \{\, x \mid \forall_y(y \in \mathcal{A} \to x \in y) \,\}$	big intersection,
$\mathcal{P}(\mathcal{A}) := \{\, x \mid x \subseteq \mathcal{A} \,\}$	power class of $\mathcal{A}$.

In particular $a \cup b = \bigcup\{a, b\}$ and $a \cap b = \bigcap\{a, b\}$, and $\bigcap \emptyset$ is the all class. Moreover $\mathcal{P}(\mathcal{A})$ is the class of all subclasses of $\mathcal{A}$ that happen to be sets.

5.2.2. Pairs, relations, functions, unions. Ordered pairs are defined by means of a little trick due to Kuratowski:

$$(a, b) := \{\, x \mid x = \{a\} \vee x = \{a, b\} \,\} \quad \text{(ordered) pair,}$$

so $(a, b) = \{\{a\}, \{a, b\}\}$. To make sure that (a, b) is not the empty class, we have to require axiomatically that $\{a\}$ and $\{a, b\}$ are sets:

Axiom (Pairing).

$$\{x, y\} \textit{ is a set.}$$

In the cumulative type structure the pairing axiom clearly holds, because for any two levels S_1 and S_2 by the Shoenfield principle there must be a level S coming after S_1 and S_2.

Explicitly the pairing axiom is $\forall_x \forall_y \exists_z \forall_u (u \in z \leftrightarrow u = x \vee u = y)$. In particular it follows that for every set a the singleton class $\{a\}$ is a set. It also follows that $(a, b) = \{\{a\}, \{a, b\}\}$ is a set.

Moreover we define

$$\{\, (x, y) \mid A(x, y) \,\} := \{\, z \mid \exists_{x,y}(A(x, y) \wedge z = (x, y)) \,\}$$

and

$\mathcal{A} \times \mathcal{B} := \{\, (x, y) \mid x \in \mathcal{A} \wedge y \in \mathcal{B} \,\}$	cartesian product of $\mathcal{A}$, $\mathcal{B}$,
$\mathrm{dom}(\mathcal{A}) := \{\, x \mid \exists_y((x, y) \in \mathcal{A}) \,\}$	domain of $\mathcal{A}$,

$\mathrm{rng}(\mathcal{A}) := \{\, y \mid \exists_x((x,y) \in \mathcal{A}) \,\}$	range of $\mathcal{A}$,
$\mathcal{A}{\restriction}\mathcal{B} := \{\, (x,y) \mid (x,y) \in \mathcal{A} \wedge x \in \mathcal{B} \,\}$	restriction of $\mathcal{A}$ to $\mathcal{B}$,
$\mathcal{A}[\mathcal{B}] := \{\, y \mid \exists_x(x \in \mathcal{B} \wedge (x,y) \in \mathcal{A}) \,\}$	image of $\mathcal{B}$ under $\mathcal{A}$,
$\mathcal{A}^{-1} := \{\, (y,x) \mid (x,y) \in \mathcal{A} \,\}$,	inverse of $\mathcal{A}$,
$\mathcal{A} \circ \mathcal{B} := \{\, (x,z) \mid \exists_y((x,y) \in \mathcal{B} \wedge (y,z) \in \mathcal{A}) \,\}$	composition of $\mathcal{A}$, $\mathcal{B}$.

Without any difficulty we can introduce the usual notions concerning relations and functions. For classes $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ we define

(a) $\mathcal{A}$ is a *relation* iff $\mathcal{A} \subseteq V \times V$. Hence a relation is a class of pairs. Instead of $(a,b) \in \mathcal{A}$ we also write $a\mathcal{A}b$.
(b) $\mathcal{A}$ is a *relation on* $\mathcal{B}$ iff $\mathcal{A} \subseteq \mathcal{B} \times \mathcal{B}$.
(c) $\mathcal{A}$ is a *function* iff $\mathcal{A}$ is a relation and

$$\forall_{x,y,z}((x,y) \in \mathcal{A} \to (x,z) \in \mathcal{A} \to y = z).$$

(d) $\mathcal{A}\colon \mathcal{B} \to \mathcal{C}$ iff $\mathcal{A}$ is a function such that $\mathrm{dom}(\mathcal{A}) = \mathcal{B}$ and $\mathcal{A}[\mathcal{B}] \subseteq \mathcal{C}$. We then call $\mathcal{A}$ a *function from* $\mathcal{B}$ *to* $\mathcal{C}$.
(e) $\mathcal{A}\colon \mathcal{B} \to_{\mathrm{onto}} \mathcal{C}$ iff $\mathcal{A}\colon \mathcal{B} \to \mathcal{C}$ and $\mathcal{A}[\mathcal{B}] = \mathcal{C}$. We then call $\mathcal{A}$ a *surjective function from* $\mathcal{B}$ *onto* $\mathcal{C}$.
(f) $\mathcal{A}$ is *injective* iff $\mathcal{A}$ and $\mathcal{A}^{-1}$ are functions.
(g) $\mathcal{A}\colon \mathcal{B} \leftrightarrow \mathcal{C}$ iff $\mathcal{A}\colon \mathcal{B} \to_{\mathrm{onto}} \mathcal{C}$ and $\mathcal{A}$ is injective. Then $\mathcal{A}$ is called *bijective function from* $\mathcal{B}$ *onto* $\mathcal{C}$.

For the further development of set theory more axioms are necessary, in particular

Axiom (Union).

$$\bigcup x \text{ is a set.}$$

The union axiom holds in the cumulative type structure. To see this, consider a level S where x is formed. An arbitrary element $v \in x$ then is available at an earlier level S_v already. Similarly every element $u \in v$ is present at a level $S_{v,u}$ before S_v. But all these u make up $\bigcup x$. Hence also $\bigcup x$ can be formed at level S.

Explicitly the union axiom is $\forall_x \exists_y \forall_z (z \in y \leftrightarrow \exists_u (u \in x \wedge z \in u))$.

We now can extend the previous definition by

$$\mathcal{A}(x) := \bigcup\{\, y \mid (x,y) \in \mathcal{A} \,\} \quad \text{application.}$$

If $\mathcal{A}$ is a function and $(x,y) \in \mathcal{A}$, then $\mathcal{A}(x) = \bigcup\{y\} = y$ and we write $\mathcal{A}\colon x \mapsto y$.

5.2.3. Separation, power set, replacement axioms.

AXIOM (Separation). *For every class $\mathcal{A}$,*

$$\mathcal{A} \subseteq x \to \exists_y(\mathcal{A} = y).$$

So the separation scheme says that every subclass $\mathcal{A}$ of a set x is a set. It is valid in the cumulative type structure, since on the same level where x is formed we can also form the set y, whose elements are just the elements of the class $\mathcal{A}$.

Notice that the separation scheme consists of infinitely many axioms.

AXIOM (Power set).

$$\mathcal{P}(x) \text{ is a set.}$$

The power set axiom holds in the cumulative type structure. To see this, consider a level S where x is formed. Then also every subset $y \subseteq x$ has been formed at level S. On the next level S' (which exists by the Shoenfield principle) we can form $\mathcal{P}(x)$.

Explicitly the power set axiom is $\forall_x \exists_y \forall_z (z \in y \leftrightarrow z \subseteq x)$.

LEMMA. *$a \times b$ is a set.*

PROOF. We show $a \times b \subseteq \mathcal{P}(\mathcal{P}(a \cup b))$. So let $x \in a$ and $y \in b$. Then

$$\begin{aligned} \{x\}, \{x, y\} &\subseteq a \cup b \\ \{x\}, \{x, y\} &\in \mathcal{P}(a \cup b) \\ \{\, \{x\}, \{x, y\} \,\} &\subseteq \mathcal{P}(a \cup b) \\ (x, y) = \{\, \{x\}, \{x, y\} \,\} &\in \mathcal{P}(\mathcal{P}(a \cup b)) \end{aligned}$$

The claim now follows from the union axiom, the pairing axiom, the power set axiom and the separation scheme. □

AXIOM (Replacement). *For every class $\mathcal{A}$,*

$$\mathcal{A} \text{ is a function } \to \forall_x(\mathcal{A}[x] \text{ is a set}).$$

Also the replacement scheme holds in the cumulative type structure; however, this requires some more thought. Consider all elements u of the set $x \cap \mathrm{dom}(\mathcal{A})$. For every such u we know that $\mathcal{A}(u)$ is a set, hence is formed at a level S_u of the cumulative type structure. Because $x \cap \mathrm{dom}(\mathcal{A})$ is a set, we can imagine a situation where all S_u for $u \in x \cap \mathrm{dom}(\mathcal{A})$ are constructed. Hence by the Shoenfield principle there must be a level S coming after all these S_u. In S we can form $\mathcal{A}[x]$.

LEMMA. *The replacement scheme implies the separation scheme.*

PROOF. Let $\mathcal{A} \subseteq x$ and $\mathcal{B} := \{\, (u, v) \mid u = v \wedge u \in \mathcal{A} \,\}$. Then $\mathcal{B}$ is a function and we have $\mathcal{B}[x] = \mathcal{A}$. □

This does not yet conclude our list of axioms of set theory: later we will require the infinity axiom, the regularity axiom and the axiom of choice.

5.3. Recursion, Induction, Ordinals

We want to develop a general framework for recursive definitions and inductive proofs. Both will be done by means of so-called well-founded relations. To carry this through, we introduce as an auxiliary notion that of a transitively well-founded relation; later we will see that it is equivalent to the notion of a well-founded relation. We then define the natural numbers in the framework of set theory, and will obtain induction and recursion on the natural numbers as special cases of the corresponding general theorems for transitively well-founded relations. By recursion on natural numbers we can then define the transitive closure of a set, and by means of this notion we will be able to show that well-founded relations coincide with the transitively well-founded relations.

Then we study particular well-founded relations. We first show that arbitrary classes together with the $\in$-relation are up to isomorphism the only well-founded extensional relations (isomorphy theorem of Mostowski). Then we consider linear well-founded orderings, called well-orderings. Since they will always be extensional, they must be isomorphic to certain classes with the $\in$-relation, which will be called ordinal classes. Ordinals can then be defined as those ordinal classes that happen to be sets.

5.3.1. Recursion on transitively well-founded relations. Let $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ denote classes. For an arbitrary relation $\mathcal{R}$ on $\mathcal{A}$ we define

(a) $\hat{x}^{\mathcal{R}} := \{\, y \mid y\mathcal{R}x \,\}$ is the class of $\mathcal{R}$-*predecessors* of x. We shall write $\hat{x}$ instead of $\hat{x}^{\mathcal{R}}$, if $\mathcal{R}$ is clear from the context.

(b) $\mathcal{B} \subseteq \mathcal{A}$ is called $\mathcal{R}$-*transitive* if

$$\forall_x(x \in \mathcal{B} \to \hat{x} \subseteq \mathcal{B}).$$

Hence $\mathcal{B} \subseteq \mathcal{A}$ is $\mathcal{R}$-transitive iff $y\mathcal{R}x$ and $x \in \mathcal{B}$ imply $y \in \mathcal{B}$.

(c) Let $\mathcal{B} \subseteq \mathcal{A}$. $x \in \mathcal{B}$ is an $\mathcal{R}$-*minimal element* of $\mathcal{B}$ if $\hat{x} \cap \mathcal{B} = \emptyset$.

(d) $\mathcal{R}$ is a *transitively well-founded relation* on $\mathcal{A}$ if

(i) for every $x \in \mathcal{A}$ there is an $\mathcal{R}$-transitive set $b \subseteq \mathcal{A}$ such that $\hat{x} \subseteq b$, and

(ii) every nonempty subset of $\mathcal{A}$ has an $\mathcal{R}$-minimal element, i.e.

$$\forall_a(a \subseteq \mathcal{A} \to a \neq \emptyset \to \exists_x(x \in a \wedge \hat{x} \cap a = \emptyset)).$$

We shall almost everywhere omit $\mathcal{R}$ if it is clear from the context.

Remark. Let $\mathcal{R}$ be a relation on $\mathcal{A}$. $\mathcal{R}$ is a *transitive relation on* $\mathcal{A}$ if for all $x, y, z \in \mathcal{A}$

$$x\mathcal{R}y \to y\mathcal{R}z \to x\mathcal{R}z.$$

We have the following connection to the notion of $\mathcal{R}$-transitivity for classes: Let $\mathcal{R}$ be a relation on $\mathcal{A}$. Then

$\mathcal{R}$ is a transitive relation on $\mathcal{A} \leftrightarrow$ for every $y \in \mathcal{A}$, $\hat{y}$ is $\mathcal{R}$-transitive.

PROOF. $\rightarrow$. Let $\mathcal{R}$ be a transitive relation on $\mathcal{A}$, $y \in \mathcal{A}$ and $x \in \hat{y}$, hence $x\mathcal{R}y$. We must show $\hat{x} \subseteq \hat{y}$. So let $z\mathcal{R}x$. We must show $z\mathcal{R}y$. But this follows from the transitivity of $\mathcal{R}$. $\leftarrow$. Let $x, y, z \in \mathcal{A}$, $x\mathcal{R}y$ and $y\mathcal{R}z$. We must show $x\mathcal{R}z$. We have $x\mathcal{R}y$ and $y \in \hat{z}$. Since $\hat{z}$ is $\mathcal{R}$-transitive, we obtain $x \in \hat{z}$, hence $x\mathcal{R}z$. □

LEMMA. *Let $\mathcal{R}$ be a transitively well-founded relation on $\mathcal{A}$. Then*

(a) *Every nonempty subclass $\mathcal{B} \subseteq \mathcal{A}$ has an $\mathcal{R}$-minimal element.*
(b) *For every $x \in \mathcal{A}$, $\hat{x}$ is a set.*

PROOF. (a). Let $\mathcal{B} \subseteq \mathcal{A}$ and $z \in \mathcal{B}$. We may assume that z is not $\mathcal{B}$-minimal, i.e., $\hat{z} \cap \mathcal{B} \neq \emptyset$. By part (i) of the definition of transitively well-founded relations there exists an $\mathcal{R}$-transitive superset $b \subseteq \mathcal{A}$ of $\hat{z}$. Because of $\hat{z} \cap \mathcal{B} \neq \emptyset$ we have $b \cap \mathcal{B} \neq \emptyset$. By part (ii) of the same definition there exists an $\mathcal{R}$-minimal $x \in b \cap \mathcal{B}$, i.e., $\hat{x} \cap b \cap \mathcal{B} = \emptyset$. Since b is $\mathcal{R}$-transitive, from $x \in b$ we obtain $\hat{x} \subseteq b$. Therefore $\hat{x} \cap \mathcal{B} = \emptyset$ and hence x is an $\mathcal{R}$-minimal element of $\mathcal{B}$.

(b). This is a consequence of the separation scheme. □

5.3.2. Induction and recursion theorems. We write $\forall_{x \in \mathcal{A}} \dots$ for $\forall_x(x \in \mathcal{A} \rightarrow \dots)$ and similarly $\exists_{x \in \mathcal{A}} \dots$ for $\exists_x(x \in \mathcal{A} \wedge \dots)$.

THEOREM (Induction theorem). *Let $\mathcal{A}$ and $\mathcal{B}$ be arbitrary classes, and $\mathcal{R}$ a transitively well-founded relation on $\mathcal{A}$. Then*

$$\forall_{x \in \mathcal{A}}(\hat{x} \subseteq \mathcal{B} \rightarrow x \in \mathcal{B}) \rightarrow \mathcal{A} \subseteq \mathcal{B}.$$

PROOF. Assume $\mathcal{A} \setminus \mathcal{B} \neq \emptyset$. Let x be a minimal element of $\mathcal{A} \setminus \mathcal{B}$. It suffices to show $\hat{x} \subseteq \mathcal{B}$, for then by assumption we obtain $x \in \mathcal{B}$, hence a contradiction. Let $z \in \hat{x}$. By the choice of x we have $z \notin \mathcal{A} \setminus \mathcal{B}$, hence $z \in \mathcal{B}$ (because $z \in \mathcal{A}$ holds, since $\mathcal{R}$ is a relation on $\mathcal{A}$). □

THEOREM (Recursion theorem). *Let $\mathcal{R}$ be a transitively well-founded relation on $\mathcal{A}$ and $\mathcal{G}\colon V \rightarrow V$. Then there exists exactly one function $\mathcal{F}\colon \mathcal{A} \rightarrow V$ such that*

$$\forall_{x \in \mathcal{A}}\big(\mathcal{F}(x) = \mathcal{G}(\mathcal{F}{\restriction}\hat{x})\big).$$

PROOF. First observe that $\mathcal{F}{\restriction}\hat{x} \subseteq \hat{x} \times \mathcal{F}[\hat{x}]$, hence $\mathcal{F}{\restriction}\hat{x}$ is a set.

Uniqueness. Given $\mathcal{F}_1$, $\mathcal{F}_2$. Consider

$$\{\, x \mid x \in \mathcal{A} \wedge \mathcal{F}_1(x) = \mathcal{F}_2(x) \,\} =: \mathcal{B}.$$

By the induction theorem it suffices to show $\forall_{x\in\mathcal{A}}(\hat{x} \subseteq \mathcal{B} \to x \in \mathcal{B})$. Let $x \in \mathcal{A}$ and $\hat{x} \subseteq \mathcal{B}$. Then

$$\begin{aligned}&\mathcal{F}_1{\restriction}\hat{x} = \mathcal{F}_2{\restriction}\hat{x}\\ &\mathcal{G}\,(\mathcal{F}_1{\restriction}\hat{x}) = \mathcal{G}\,(\mathcal{F}_2{\restriction}\hat{x})\\ &\mathcal{F}_1(x) = \mathcal{F}_2(x)\\ &x \in \mathcal{B}.\end{aligned}$$

Existence. Let

$$\mathcal{B} := \{\, f \mid f \text{ function, } \mathrm{dom}(f) \text{ } \mathcal{R}\text{-transitive subset of } \mathcal{A},\ \forall_{x\in\mathrm{dom}(f)}(f(x) = \mathcal{G}(f{\restriction}\hat{x}))\,\}$$

and

$$\mathcal{F} := \bigcup \mathcal{B}.$$

We first show that

$$f, g \in \mathcal{B} \to x \in \mathrm{dom}(f) \cap \mathrm{dom}(g) \to f(x) = g(x).$$

Let $f, g \in \mathcal{B}$. We prove the claim by induction on x, i.e., by an application of the induction theorem to

$$\{\, x \mid x \in \mathrm{dom}(f) \cap \mathrm{dom}(g) \to f(x) = g(x)\,\}.$$

Let $x \in \mathrm{dom}(f) \cap \mathrm{dom}(g)$. Then

$$\begin{aligned}&\hat{x} \subseteq \mathrm{dom}(f) \cap \mathrm{dom}(g), && \text{since } \mathrm{dom}(f), \mathrm{dom}(g) \text{ are } \mathcal{R}\text{-transitive}\\ &f{\restriction}\hat{x} = g{\restriction}\hat{x} && \text{by induction hypothesis}\\ &\mathcal{G}\,(f{\restriction}\hat{x}) = \mathcal{G}\,(g{\restriction}\hat{x})\\ &f(x) = g(x).\end{aligned}$$

Therefore $\mathcal{F}$ is a function. Now this immediately implies $f \in \mathcal{B} \to x \in \mathrm{dom}(f) \to \mathcal{F}(x) = f(x)$; hence we have shown

$$\mathcal{F}(x) = \mathcal{G}\,(\mathcal{F}{\restriction}\hat{x}) \quad \text{for all } x \in \mathrm{dom}(\mathcal{F}). \tag{5.1}$$

We now show

$$\mathrm{dom}(\mathcal{F}) = \mathcal{A}.$$

$\subseteq$ is clear. $\supseteq$. Use the induction theorem. Let $\hat{y} \subseteq \mathrm{dom}(\mathcal{F})$. We must show $y \in \mathrm{dom}(\mathcal{F})$. This is proved indirectly; so assume $y \notin \mathrm{dom}(\mathcal{F})$. Let b be $\mathcal{R}$-transitive such that $\hat{y} \subseteq b \subseteq \mathcal{A}$. Define

$$g := \mathcal{F}{\restriction}b \cup \{\, (y, \mathcal{G}(\mathcal{F}{\restriction}\hat{y}))\,\}.$$

It clearly suffices to show $g \in \mathcal{B}$, for because of $y \in \mathrm{dom}(g)$ this implies $y \in \mathrm{dom}(\mathcal{F})$ and hence the desired contradiction.

g is a function: This is clear, since $y \notin \mathrm{dom}(\mathcal{F})$ by assumption.

$\mathrm{dom}(g)$ is $\mathcal{R}$-transitive: We have $\mathrm{dom}(g) = (b \cap \mathrm{dom}(\mathcal{F})) \cup \{y\}$. First notice that $\mathrm{dom}(\mathcal{F})$ as a union of $\mathcal{R}$-transitive sets is $\mathcal{R}$-transitive itself. Moreover, since b is $\mathcal{R}$-transitive, also $b \cap \mathrm{dom}(\mathcal{F})$ is $\mathcal{R}$-transitive. Now let $z\mathcal{R}x$ and $x \in \mathrm{dom}(g)$. We must show $z \in \mathrm{dom}(g)$. In case $x \in b \cap \mathrm{dom}(\mathcal{F})$ also $z \in b \cap \mathrm{dom}(\mathcal{F})$ (since $b \cap \mathrm{dom}(\mathcal{F})$ is $\mathcal{R}$-transitive, as we just observed), hence $z \in \mathrm{dom}(g)$. In case $x = y$ we have $z \in \hat{y}$, hence $z \in b$ and $z \in \mathrm{dom}(\mathcal{F})$ by the choice of b and y, hence again $z \in \mathrm{dom}(g)$.

$\forall_{x \in \mathrm{dom}(g)}(g(x) = \mathcal{G}(g{\restriction}\hat{x}))$: In case $x \in b \cap \mathrm{dom}(\mathcal{F})$ we have

$$\begin{aligned} g(x) &= \mathcal{F}(x) \\ &= \mathcal{G}(\mathcal{F}{\restriction}\hat{x}) \quad \text{by (5.1)} \\ &= \mathcal{G}(g{\restriction}\hat{x}) \quad \text{since } \hat{x} \subseteq b \cap \mathrm{dom}(\mathcal{F}), \text{ for } b \cap \mathrm{dom}(\mathcal{F}) \text{ is } \mathcal{R}\text{-transitive.} \end{aligned}$$

In case $x = y$ we have $g(x) = \mathcal{G}(\mathcal{F}{\restriction}\hat{x}) = \mathcal{G}(g{\restriction}\hat{x})$, for $\hat{x} = \hat{y} \subseteq b \cap \mathrm{dom}(\mathcal{F})$ by the choice of y. □

5.3.3. Natural numbers. Zermelo defined the natural numbers within set theory as follows: $0 = \emptyset$, $1 = \{\emptyset\}$, $2 = \{\{\emptyset\}\}$, $3 = \{\{\{\emptyset\}\}\}$ and so on. A disadvantage of this definition is that it cannot be generalized to the transfinite. Later, John von Neumann proposed to represent the number n by a certain set consisting of exactly n elements, namely

$$n := \{\, 0, 1, \ldots, n-1 \,\}.$$

So $0 = \emptyset$ and $n + 1 = \{\, 0, 1, \ldots, n \,\} = \{\, 0, 1, \ldots, n-1 \,\} \cup \{n\}$. Generally we define

$$0 := \emptyset, \quad x + 1 := x \cup \{x\}.$$

In particular, $1 := 0 + 1$, $2 := 1 + 1$, $3 := 2 + 1$ and so on.

In order to know that the class of all natural numbers constructed in this way is a set, we need another axiom:

Axiom (Infinity).

$$\exists_x(\emptyset \in x \wedge \forall_y(y \in x \to y \cup \{y\} \in x)).$$

The infinity axiom holds in the cumulative type structure. To see this, observe that $0 = \emptyset$, $1 := \emptyset \cup \{\emptyset\}$, $2 := 1 \cup \{1\}$ and so on are formed at levels S_0, S_1, S_2 …, and we can conceive a situation where all these levels are completed. By the Shoenfield principle there must be a level – call it S_ω – which is past all these levels. At S_ω we can form ω.

We call a class $\mathcal{A}$ *inductive* if

$$\emptyset \in \mathcal{A} \wedge \forall_y(y \in \mathcal{A} \to y \cup \{y\} \in \mathcal{A}).$$

So the infinity axiom says that there is an inductive set. Define

$$\omega := \bigcap \{\, x \mid x \text{ is inductive} \,\}.$$

Clearly ω is a set, with the properties $0 \in \omega$ and $y \in \omega \to y + 1 \in \omega$. ω is called the set of *natural numbers*.

Let n, m denote natural numbers. $\forall_n A(n)$ is short for $\forall_x(x \in \omega \to A(x))$, similarly $\exists_n A(n)$ for $\exists_x(x \in \omega \wedge A(x))$ and $\{ n \mid A(n) \}$ for $\{ x \mid x \in \omega \wedge A(x) \}$.

THEOREM (Induction on ω).

(a) $x \subseteq \omega \to 0 \in x \to \forall_n(n \in x \to n + 1 \in x) \to x = \omega$.

(b) *For every formula* $A(x)$,

$$A(0) \to \forall_n(A(n) \to A(n + 1)) \to \forall_n A(n).$$

PROOF. (a). x is inductive, hence $\omega \subseteq x$. (b). Let $\mathcal{A} := \{ n \mid A(n) \}$. Then $\mathcal{A} \subseteq \omega$ (so $\mathcal{A}$ is set), and by assumption

$$0 \in \mathcal{A},$$
$$n \in \mathcal{A} \to n + 1 \in \mathcal{A}.$$

By (a), $\mathcal{A} = \omega$. □

We now show that for natural numbers the relation $\in$ has all the properties of $<$, and the relation $\subseteq$ all the properties of $\leq$.

A class $\mathcal{A}$ is called *transitive* if it is $\mathcal{E}$-transitive w.r.t. the special relation $\mathcal{E} := \{ (x, y) \mid x \in y \}$ on V, i.e., if $\forall_x(x \in \mathcal{A} \to x \subseteq \mathcal{A})$. Therefore $\mathcal{A}$ is transitive iff

$$y \in x \in \mathcal{A} \to y \in \mathcal{A}.$$

LEMMA (Transitivity). (a) *n is transitive.*

(b) *ω is transitive.*

PROOF. (a). Induction on n. 0 is transitive. $n \to n + 1$. By induction hypothesis, n is transitive. We must show that $n+1$ is transitive. We argue as follows:

$$\begin{aligned} &y \in x \in n + 1 \\ &y \in x \in n \cup \{n\} \\ &y \in x \in n \vee y \in x = n \\ &y \in n \vee y \in n \qquad \text{by induction hypothesis} \\ &y \in n \cup \{n\} = n + 1. \end{aligned}$$

(b). We show $\forall_x(x \in n \to x \in \omega)$, by induction on n. 0: Clear. $n \to n+1$. By induction hypothesis we have $\forall_x(x \in n \to x \in \omega)$. So assume $x \in n + 1$. Then $x \in n \vee x = n$, hence $x \in \omega$. □

LEMMA. $n \notin n$.

PROOF. Induction on n. 0. Clear. $n \to n+1$: By induction hypothesis is $n \notin n$. Assume

$$\begin{aligned} &n+1 \in n+1 \\ &n+1 \in n \vee n+1 = n \\ &n \in n+1 \in n \vee n \in n+1 = n \\ &n \in n \quad \text{for } n \text{ is transitive by the transitivity lemma.} \end{aligned}$$

This is a contradiction to the induction hypothesis. □

LEMMA. (a) $n \subseteq m+1 \leftrightarrow n \subseteq m \vee n = m+1$.
(b) $n \subseteq m \leftrightarrow n \in m \vee n = m$.
(c) $n \subseteq m \vee m \subseteq n$.
(d) $n \in m \vee n = m \vee m \in n$.

PROOF. (a). $\leftarrow$ follows from $m \subseteq m+1$. $\rightarrow$. Assume $n \subseteq m+1$. *Case* $m \in n$. We show $n = m+1$. $\subseteq$ holds by assumption. $\supseteq$.

$$\begin{aligned} &p \in m+1 \\ &p \in m \vee p = m \\ &p \in n. \end{aligned}$$

Case $m \notin n$. We show $n \subseteq m$.

$$\begin{aligned} &p \in n \\ &p \in m+1 \\ &p \in m \vee p = m, \end{aligned}$$

but $p = m$ is impossible because of $m \notin n$.

(b). $\leftarrow$ follows from transitivity of m. $\rightarrow$. Induction on m. 0. Clear. $m \to m+1$.

$$\begin{aligned} &n \subseteq m+1 \\ &n \subseteq m \vee n = m+1 && \text{by (a)} \\ &n \in m \vee n = m \vee n = m+1 && \text{by induction hypothesis} \\ &n \in m+1 \vee n = m+1. \end{aligned}$$

(c). Induction on n. 0. Clear. $n \to n+1$: *Case* $m \subseteq n$. Clear. *Case* $n \subseteq m$. Then

$$\begin{aligned} &n \in m \vee n = m && \text{by (b)} \\ &n, \{n\} \subseteq m \vee m \subseteq n+1 \\ &n+1 \subseteq m \vee m \subseteq n+1 \end{aligned}$$

(d). Follows from (c) and (b). □

THEOREM (Peano axioms). (a) $n+1 \neq \emptyset$.
(b) $n+1 = m+1 \to n = m$.
(c) $x \subseteq \omega \to 0 \in x \to \forall_n(n \in x \to n+1 \in x) \to x = \omega$.

PROOF. (a). Clear. (c). This is part (a) of the theorem on ω-induction. (b).

$$\begin{aligned} & n+1 = m+1 \\ & n \in m+1 \wedge m \in n+1 \\ & (n \in m \wedge m \in n) \vee n = m \\ & n \in n \vee n = m \\ & n = m. \end{aligned}$$

□

We now treat different forms of induction.

THEOREM (Course-of-values induction on ω).
(a) $x \subseteq \omega \to \forall_n(\forall_m(m \in n \to m \in x) \to n \in x) \to x = \omega$.
(b) $\forall_n(\forall_m(m \in n \to A(m)) \to A(n)) \to \forall_n A(n)$.

PROOF. (b). Assume $\forall_n(\forall_m(m \in n \to A(m)) \to A(n))$; we shall say in this case that $A(n)$ is *progressive*. We show $\forall_m(m \in n \to A(m))$, by induction on n. 0. Clear. $n \to n+1$. By induction hypothesis $\forall_m(m \in n \to A(m))$. Let $m \in n+1$. Then $m \in n \vee m = n$. In case $m \in n$ we obtain $A(m)$ by induction hypothesis, and in case $m = n$ we can infer $A(n)$ from the progressiveness of A, using the induction hypothesis.
(a). From (b), with $A(y) := y \in x$. □

THEOREM (Principle of least element for ω).
(a) $\emptyset \neq x \subseteq \omega \to \exists_n(n \in x \wedge n \cap x = \emptyset)$.
(b) $\exists_n A(n) \to \exists_n(A(n) \wedge \neg\exists_m(m \in n \wedge A(m)))$.

PROOF. (b). By course-of-values induction

$$\forall_n(\forall_m(m \in n \to \neg A(m)) \to \neg A(n)) \to \forall_n \neg A(n).$$

Contraposition gives

$$\begin{aligned} & \exists_n A(n) \to \exists_n(A(n) \wedge \forall_m(m \in n \to \neg A(m))) \\ & \exists_n A(n) \to \exists_n(A(n) \wedge \neg\exists_m(m \in n \wedge A(m))). \end{aligned}$$

(a). From (b), using $A(y) := (y \in x)$. □

We now consider recursion on natural numbers, which can be treated as a special case of the recursion theorem. To this end, we identify $\in$ with the relation $\mathcal{E} = \{\, (x,y) \mid x \in y \,\}$ and prove the following lemma:

LEMMA. *$\in \cap(\omega \times \omega)$ is a transitively well-founded relation on ω.*

PROOF. We show both conditions, from the definition of transitively well-founded relations. (i). Clear, since n is transitive. (ii). Let $\emptyset \neq a \subseteq \omega$. We must show $\exists_n(n \in a \wedge n \cap a = \emptyset)$. But this is the above principle of the least element. □

THEOREM (Course-of-values recursion on ω). *Let $\mathcal{G}\colon V \to V$. Then there is exactly one function $f\colon \omega \to V$ such that*

$$\forall_n\,(f(n) = \mathcal{G}(f{\restriction}n)).$$

PROOF. By the recursion theorem there is a unique $\mathcal{F}\colon \omega \to V$ such that $\forall_n\,(\mathcal{F}(n) = \mathcal{G}(\mathcal{F}{\restriction}n))$. By replacement, $\operatorname{rng}(\mathcal{F}) = \mathcal{F}[\omega]$ is a set. Since the product of two sets is a set, by separation $\mathcal{F} \subseteq \omega \times \mathcal{F}[\omega]$ is a set. □

COROLLARY (ω-recursion). *Let $\mathcal{G}\colon V \to V$ and a be a set. Then there is exactly one function $f\colon \omega \to V$ such that*

$$\begin{aligned} &f(0) = a,\\ &\forall_n(f(n+1) = \mathcal{G}(f(n))). \end{aligned}$$

PROOF. First observe that $\bigcup(n+1) = n$, because of

$$\begin{aligned} x \in \bigcup(n+1) &\leftrightarrow \exists_y(x \in y \in n+1)\\ &\leftrightarrow \exists_m(x \in m \in n+1)\\ &\leftrightarrow \exists_m(x \in m \subseteq n)\\ &\leftrightarrow x \in n. \end{aligned}$$

For the given $\mathcal{G}$ we will construct $\mathcal{G}'$ such that $\mathcal{G}'(f{\restriction}(n+1)) = \mathcal{G}(f(n))$. We define a function $\mathcal{G}'\colon V \to V$ satisfying

$$\mathcal{G}'(x) = \begin{cases} \mathcal{G}(x(\bigcup \operatorname{dom}(x))), & \text{if } x \neq \emptyset;\\ a, & \text{if } x = \emptyset, \end{cases}$$

by

$$\mathcal{G}' = \{\,(x,y) \mid (x \neq \emptyset \to y = \mathcal{G}(x(\bigcup \operatorname{dom}(x)))) \wedge (x = \emptyset \to y = a)\,\}.$$

Then there is a unique function $f\colon \omega \to V$ such that

$$\begin{aligned} f(n+1) &= \mathcal{G}'(f{\restriction}(n+1))\\ &= \mathcal{G}((f{\restriction}(n+1))(\underbrace{\bigcup(n+1)}_{n}))\\ &= \mathcal{G}(f(n)), \end{aligned}$$

and

$$f(0) = \mathcal{G}'(\underbrace{f{\restriction}0}_{\emptyset}) = a.$$ □

We now define

$$s_m(0) = m, \quad s_m(n+1) = s_m(n) + 1.$$

By ω-recursion for every m there is such a function, and it is uniquely determined. We define

$$m + n := s_m(n).$$

Because of $s_m(1) = s_m(0+1) = s_m(0) + 1 = m + 1$, for $n = 1$ this definition is compatible with the previous terminology. Moreover, we have $m + 0 = m$ and $m + (n + 1) = (m + n) + 1$.

LEMMA. (a) $m + n \in \omega$.
(b) $(m + n) + p = m + (n + p)$.
(c) $m + n = n + m$.

PROOF. (a). Induction on n. 0. Clear. $n \to n + 1$. $m + (n + 1) = (m + n) + 1$, and by induction hypothesis $m + n \in \omega$.

(b). Induction on p. 0. Clear. $p \to p + 1$.

$$\begin{aligned}(m + n) + (p + 1) &= [(m + n) + p] + 1 && \text{by definition}\\ &= [m + (n + p)] + 1 && \text{by induction hypothesis}\\ &= m + [(n + p) + 1]\\ &= m + [n + (p + 1)].\end{aligned}$$

(c). We first prove two auxiliary propositions.

(i) $0 + n = n$. The proof is by induction on n. 0. Clear. $n \to n + 1$. $0 + (n + 1) = (0 + n) + 1 = n + 1$.

(ii) $(m + 1) + n = (m + n) + 1$. Again the proof is by induction on n. 0. Clear. $n \to n + 1$.

$$\begin{aligned}(m + 1) + (n + 1) &= [(m + 1) + n] + 1\\ &= [(m + n) + 1] + 1 && \text{by induction hypothesis}\\ &= [m + (n + 1)] + 1.\end{aligned}$$

Now the claim $m + n = n + m$ can be proved by induction on m. 0. By (i). Step $m \to m + 1$.

$$\begin{aligned}(m + 1) + n &= (m + n) + 1 && \text{by (ii)}\\ &= (n + m) + 1 && \text{by induction hypothesis}\\ &= n + (m + 1).\end{aligned}$$

□

We define

$$p_m(0) = 0, \quad p_m(n+1) = p_m(n) + m.$$

By ω-recursion, for every m there is a unique such function. Here we need

$$\mathcal{G}\colon V \to V,$$

$$\mathcal{G}(x) = \begin{cases} x + m, & \text{if } x \in \omega; \\ \emptyset, & \text{otherwise.} \end{cases}$$

We finally define $m \cdot n := p_m(n)$. Observe that this implies $m \cdot 0 = 0$, $m \cdot (n+1) = m \cdot n + m$.

LEMMA. (a) $m \cdot n \in \omega$.
(b) $m \cdot (n+p) = m \cdot n + m \cdot p$.
(c) $(n+p) \cdot m = n \cdot m + p \cdot m$.
(d) $(m \cdot n) \cdot p = m \cdot (n \cdot p)$.
(e) $0 \cdot n = 0$, $1 \cdot n = n$, $m \cdot n = n \cdot m$.

PROOF. Exercise. □

REMARK. n^m, $m - n$ can be treated similarly; later (when we deal with ordinal arithmetic) this will be done more generally. - We could now introduce the integers, rationals, reals and complex numbers in the well-known way, and prove their elementary properties.

5.3.4. Transitive closure. We define the $\mathcal{R}$-transitive closure of a set a, w.r.t. a relation $\mathcal{R}$ with the property that the $\mathcal{R}$-predecessors of an arbitrary element of its domain form a set.

THEOREM. *Let $\mathcal{R}$ be a relation on $\mathcal{A}$ such that $\hat{x}^{\mathcal{R}}$ $(:= \{ y \mid y\mathcal{R}x \})$ is a set, for every $x \in \mathcal{A}$. Then for every subset $a \subseteq \mathcal{A}$ there is a uniquely determined set b such that*

(a) *$a \subseteq b \subseteq \mathcal{A}$;*
(b) *b is $\mathcal{R}$-transitive;*
(c) *$\forall_c(a \subseteq c \subseteq \mathcal{A} \to c$ $\mathcal{R}$-transitive $\to b \subseteq c)$.*

b is called the $\mathcal{R}$-transitive closure *of a.*

PROOF. Uniqueness. Clear by (c). Existence. We shall define $f\colon \omega \to V$ by recursion on ω, such that

$$f(0) = a,$$
$$f(n+1) = \{ y \mid \exists_{x \in f(n)}(y\mathcal{R}x) \}.$$

In order to apply the recursion theorem for ω, we must define $f(n+1)$ in the form $\mathcal{G}(f(n))$. To this end choose $\mathcal{G}\colon V \to V$, $z \mapsto \bigcup \mathrm{rng}(\mathcal{H}\restriction z)$ such that $\mathcal{H}\colon V \to V$, $x \mapsto \hat{x}$; by assumption $\mathcal{H}$ is a function. Then

$$\begin{aligned} y \in \mathcal{G}(f(n)) &\leftrightarrow y \in \bigcup \mathrm{rng}(\mathcal{H}\restriction f(n)) \\ &\leftrightarrow \exists_z(z \in \mathrm{rng}(\mathcal{H}\restriction f(n)) \wedge y \in z) \\ &\leftrightarrow \exists_{z,x}(x \in f(n) \wedge z = \hat{x} \wedge y \in z) \\ &\leftrightarrow \exists_{x \in f(n)}(y \in \hat{x}) \end{aligned}$$

$$\leftrightarrow \exists_{x\in f(n)}(y\mathcal{R}x).$$

By induction on n one can see easily that $f(n)$ is a set: for 0 this is clear, and in the step $n \to n+1$ it follows – using $f(n+1) = \bigcup\{\, \hat{x} \mid x \in f(n)\,\}$ – from the induction hypothesis, replacement and the union axiom. – We now define $b := \bigcup \mathrm{rng}(f) = \bigcup\{\, f(n) \mid n \in \omega\,\}$. Then

(a). $a = f(0) \subseteq b \subseteq \mathcal{A}$.

(b).

$$\begin{aligned} &y\mathcal{R}x \in b\\ &y\mathcal{R}x \in f(n)\\ &y \in f(n+1)\\ &y \in b. \end{aligned}$$

(c). Let $a \subseteq c \subseteq \mathcal{A}$ and c be $\mathcal{R}$-transitive. We show $f(n) \subseteq c$ by induction on n. 0. $a \subseteq c$. $n \to n+1$.

$$\begin{aligned} &y \in f(n+1)\\ &y\mathcal{R}x \in f(n)\\ &y\mathcal{R}x \in c\\ &y \in c. \end{aligned}$$ □

In the special case of the element relation $\in$ on V, the condition $\forall_x(\hat{x} = \{\, y \mid y \in x\,\}$ is a set) clearly holds. Hence for every set a there is a uniquely determined $\in$-transitive closure of a. It is called the *transitive closure* of a.

By means of the notion of the $\mathcal{R}$-transitive closure we can now show that the transitively well-founded relations on $\mathcal{A}$ coincide with the well-founded relations on $\mathcal{A}$.

Let $\mathcal{R}$ be a relation on $\mathcal{A}$. $\mathcal{R}$ is a *well-founded relation* on $\mathcal{A}$ if

(a)
$$\forall_a(a \subseteq \mathcal{A} \to a \neq \emptyset \to \exists_{x\in a}(\hat{x} \cap a = \emptyset)),$$
i.e., every nonempty subset of $\mathcal{A}$ has an $\mathcal{R}$-minimal element, and

(b) for every $x \in \mathcal{A}$, $\hat{x}$ is a set.

THEOREM. *The transitively well-founded relations on $\mathcal{A}$ are the same as the well-founded relations on $\mathcal{A}$.*

PROOF. Every transitively well-founded relation on $\mathcal{A}$ is well-founded by part (b) of the lemma in 5.3.1. Conversely, every well-founded relation on $\mathcal{A}$ is transitively well-founded, since for every $x \in \mathcal{A}$, the $\mathcal{R}$-transitive closure of $\hat{x}$ is an $\mathcal{R}$-transitive $b \subseteq \mathcal{A}$ such that $\hat{x} \subseteq b$. □

Therefore, the induction theorem and the recursion theorem in 5.3.2 also hold for well-founded relations. Moreover, by part (a) of the lemma in 5.3.1,

every nonempty sub*class* of a well-founded relation $\mathcal{R}$ has an $\mathcal{R}$-minimal element.

Later we will require the so-called regularity axiom, which says that the relation $\in$ on V is well-founded, i.e.,

$$\forall_a(a \neq \emptyset \to \exists_{x\in a}(x \cap a = \emptyset)).$$

This will provide us with an important example of a well-founded relation.

We now consider extensional well-founded relations. From the regularity axiom it will follow that the $\in$-relation on an arbitrary class $\mathcal{A}$ is a well-founded extensional relation. Here we show – even without the regularity axiom – the converse, namely that every well-founded extensional relation is isomorphic to the $\in$-relation on a transitive class. This is Mostowski's isomorphy theorem. Then we consider linear well-founded orderings, well-orderings for short. They are always extensional, and hence isomorphic to the $\in$-relation on certain classes, which will be called ordinal classes. Ordinals will then be defined as ordinal sets.

A relation $\mathcal{R}$ on $\mathcal{A}$ is *extensional* if for all $x, y \in \mathcal{A}$

$$\forall_{z\in\mathcal{A}}(z\mathcal{R}x \leftrightarrow z\mathcal{R}y) \to x = y.$$

For example, for a transitive class $\mathcal{A}$ the relation $\in \cap(\mathcal{A} \times \mathcal{A})$ is extensional on $\mathcal{A}$. This can be seen as follows. Let $x, y \in \mathcal{A}$. For $\mathcal{R} :=\in \cap(\mathcal{A} \times \mathcal{A})$ we have $z\mathcal{R}x \leftrightarrow z \in x$, since $\mathcal{A}$ is transitive. Now assume $\forall_{z\in\mathcal{A}}(z\mathcal{R}x \leftrightarrow z\mathcal{R}y)$. Then $\forall_z(z \in x \leftrightarrow z \in y)$ and hence $x = y$ by extensionality.

From the regularity axiom it will follow that all these relations are well-founded. But even without the regularity axiom these relations have a distinguished meaning; cf. the corollary below.

THEOREM (Isomorphy theorem of Mostowski). *Let $\mathcal{R}$ be a well-founded extensional relation on $\mathcal{A}$. Then there is a unique isomorphism $\mathcal{F}$ of $\mathcal{A}$ onto a transitive class $\mathcal{B}$, i.e.,*

$$\exists^{=1}_{\mathcal{F}}(\mathcal{F}\colon \mathcal{A} \leftrightarrow \mathrm{rng}(\mathcal{F}) \wedge \mathrm{rng}(\mathcal{F}) \textit{ transitive } \wedge \forall_{x,y\in\mathcal{A}}(y\mathcal{R}x \leftrightarrow \mathcal{F}(y) \in \mathcal{F}(x))).$$

PROOF. Existence. We define by the recursion theorem

$$\mathcal{F}\colon \mathcal{A} \to V,$$
$$\mathcal{F}(x) = \mathrm{rng}(\mathcal{F}{\restriction}\hat{x}) \quad (= \{\, \mathcal{F}(y) \mid y\mathcal{R}x \,\}).$$

$\mathcal{F}$ injective: We show $\forall_{x,y\in\mathcal{A}}(\mathcal{F}(x) = \mathcal{F}(y) \to x = y)$ by $\mathcal{R}$-induction on x. Let $x, y \in \mathcal{A}$ be given such that $\mathcal{F}(x) = \mathcal{F}(y)$. By induction hypothesis

$$\forall_{z\in\mathcal{A}}(z\mathcal{R}x \to \forall_{u\in\mathcal{A}}(\mathcal{F}(z) = \mathcal{F}(u) \to z = u)).$$

Since $\mathcal{R}$ is extensional on $\mathcal{A}$ it suffices to show $\forall_{x\in\mathcal{A}}(z\mathcal{R}x \leftrightarrow z\mathcal{R}y)$. Let $x \in \mathcal{A}$. $\to$.

$$z\mathcal{R}x$$

$\mathcal{F}(z) \in \mathcal{F}(x) = \mathcal{F}(y) = \{\, \mathcal{F}(u) \mid u\mathcal{R}y \,\}$
$\mathcal{F}(z) = \mathcal{F}(u)$ for some $u\mathcal{R}y$
$z = u$ by induction hypothesis, since $z\mathcal{R}x$
$z\mathcal{R}y$

$\leftarrow$.

$z\mathcal{R}y$
$\mathcal{F}(z) \in \mathcal{F}(y) = \mathcal{F}(x) = \{\, \mathcal{F}(u) \mid u\mathcal{R}x \,\}$
$\mathcal{F}(z) = \mathcal{F}(u)$ for some $u\mathcal{R}x$
$z = u$ by induction hypothesis, since $u\mathcal{R}x$
$z\mathcal{R}x$.

$\mathrm{rng}(\mathcal{F})$ is transitive: Assume $u \in v \in \mathrm{rng}(\mathcal{F})$. Then $v = \mathcal{F}(x)$ for some $x \in \mathcal{A}$, hence $u = \mathcal{F}(y)$ for some $y\mathcal{R}x$.

$y\mathcal{R}x \leftrightarrow \mathcal{F}(y) \in \mathcal{F}(x)$: $\rightarrow$. Assume $y\mathcal{R}x$. Then $\mathcal{F}(y) \in \mathcal{F}(x)$ by definition of $\mathcal{F}$. $\leftarrow$.

$\mathcal{F}(y) \in \mathcal{F}(x) = \{\, \mathcal{F}(z) \mid z\mathcal{R}x \,\}$
$\mathcal{F}(y) = \mathcal{F}(z)$ for some $z\mathcal{R}x$
$y = z$ since $\mathcal{F}$ is injective
$y\mathcal{R}x$.

Uniqueness. Let $\mathcal{F}_i$ $(i = 1, 2)$ be two isomorphisms as described in the theorem. We show $\forall_{x\in\mathcal{A}}(\mathcal{F}_1(x) = \mathcal{F}_2(x))$, by $\mathcal{R}$-induction on x. By symmetry it suffices to prove $u \in \mathcal{F}_1(x) \rightarrow u \in \mathcal{F}_2(x)$.

$u \in \mathcal{F}_1(x) \in \mathrm{rng}(\mathcal{F}_1)$
$u \in \mathrm{rng}(\mathcal{F}_1)$ since $\mathrm{rng}(\mathcal{F}_1)$ is transitive
$u = \mathcal{F}_1(y)$ for some $y \in \mathcal{A}$
$y\mathcal{R}x$ by the isomorphy condition for $\mathcal{F}_1$
$u = \mathcal{F}_2(y)$ by induction hypothesis
$\mathcal{F}_2(y) \in \mathcal{F}_2(x)$ by the isomorphy condition for $\mathcal{F}_2$
$u \in \mathcal{F}_2(x)$.

This concludes the proof. □

A relation $\mathcal{R}$ on $\mathcal{A}$ is a *linear ordering* if for all $x, y, z \in \mathcal{A}$

$\neg x\mathcal{R}x$ irreflexivity,
$x\mathcal{R}y \rightarrow y\mathcal{R}z \rightarrow x\mathcal{R}z$ transitivity,
$x\mathcal{R}y \vee x = y \vee y\mathcal{R}x$ trichotomy (or comparability).

$\mathcal{R}$ is a *well-ordering* if $\mathcal{R}$ is a well-founded linear ordering.

REMARK. Every well-ordering $\mathcal{R}$ on $\mathcal{A}$ is extensional. To see this, assume

$$\forall_{z\in\mathcal{A}}(z\mathcal{R}x \leftrightarrow z\mathcal{R}y).$$

Then $x = y$ by trichotomy, since from $x\mathcal{R}y$ we obtain by assumption $x\mathcal{R}x$, contradicting irreflexivity, and similarly $y\mathcal{R}x$ entails a contradiction.

COROLLARY. *For every well-ordering $\mathcal{R}$ on $\mathcal{A}$ there is a unique isomorphism $\mathcal{F}$ of $\mathcal{A}$ onto a transitive class $\mathcal{B}$.*

5.3.5. Ordinal classes and ordinals. We now study more closely the transitive classes that appear as images of well-orderings.

$\mathcal{A}$ is an *ordinal class* if $\mathcal{A}$ is transitive and $\in \cap(\mathcal{A}\times\mathcal{A})$ is a well-ordering on $\mathcal{A}$. Ordinal classes that happen to be sets are called *ordinals*. Define

$$\mathrm{On} := \{\, x \mid x \text{ is an ordinal} \,\}.$$

First we give a convenient characterization of ordinal classes. $\mathcal{A}$ is called *connex* if for all $x, y \in \mathcal{A}$

$$x \in y \vee x = y \vee y \in x.$$

For instance, ω is connex (i.e., $n \in m \vee n = m \vee m \in n$) by a lemma in 5.3.3. Also every n is connex, since by the transitivity lemma, ω is transitive.

A class $\mathcal{A}$ is *well-founded* if $\in \cap(\mathcal{A}\times\mathcal{A})$ is a well-founded relation on $\mathcal{A}$, i.e., if

$$\forall_{a\subseteq\mathcal{A}}(a \neq \emptyset \to \exists_{x\in a}(x \cap a = \emptyset)).$$

We now show that in well-founded classes there can be no finite $\in$-cycles.

LEMMA ($\in$-cycles). *Let $\mathcal{A}$ be well-founded. Then for arbitrary elements $x_1, \ldots, x_n \in \mathcal{A}$ we can never have*

$$x_1 \in x_2 \in \cdots \in x_n \in x_1.$$

PROOF. Assume $x_1 \in x_2 \in \cdots \in x_n \in x_1$. Consider $\{x_1, \ldots, x_n\}$. Since $\mathcal{A}$ is well-founded we may assume $x_1 \cap \{x_1, \ldots, x_n\} = \emptyset$. But this contradicts $x_n \in x_1$. □

COROLLARY. *$\mathcal{A}$ is an ordinal class iff $\mathcal{A}$ is transitive, connex and well-founded.*

PROOF. $\to$ is clear; $\mathcal{A}$ is connex because of trichotomy. $\leftarrow$: We must show, for all $x, y, z \in \mathcal{A}$,

$$x \notin x,$$
$$x \in y \to y \in z \to x \in z.$$

Since $\mathcal{A}$ is connex, both propositions follow from the lemma on $\in$-cycles. □

Here are some examples of ordinals. ω is transitive by the transitivity lemma, connex as noted above and well-founded by the principle of least element. So, ω is an ordinal class. Since ω by the infinity axiom is a set, ω is even an ordinal. Also, n is transitive by the transitivity lemma, connex (see above) and well-founded; the latter follows with transitivity of ω by the principle of least element.

Let us write $\mathrm{Ord}(\mathcal{A})$ for "$\mathcal{A}$ is an ordinal class". We show that ordinal classes have properties similar to those of natural numbers: the relation $\in$ has the properties of $<$, and the relation $\subseteq$ has the properties of $\leq$.

LEMMA (Ordinal classes). (a) $\mathrm{Ord}(\mathcal{A}) \to \mathrm{Ord}(\mathcal{B}) \to \mathrm{Ord}(\mathcal{A} \cap \mathcal{B})$.
(b) $\mathrm{Ord}(\mathcal{A}) \to x \in \mathcal{A} \to \mathrm{Ord}(x)$.
(c) $\mathrm{Ord}(\mathcal{A}) \to \mathrm{Ord}(\mathcal{B}) \to (\mathcal{A} \subseteq \mathcal{B} \leftrightarrow \mathcal{A} \in \mathcal{B} \vee \mathcal{A} = \mathcal{B})$.
(d) $\mathrm{Ord}(\mathcal{A}) \to \mathrm{Ord}(\mathcal{B}) \to (\mathcal{A} \in \mathcal{B} \vee \mathcal{A} = \mathcal{B} \vee \mathcal{B} \in \mathcal{A})$.

PROOF. (a). $\mathcal{A} \cap \mathcal{B}$ transitive:

$$\begin{aligned} &x \in y \in \mathcal{A} \cap \mathcal{B} \\ &x \in y \in \mathcal{A} \quad \text{and} \quad x \in y \in \mathcal{B} \\ &x \in \mathcal{A} \quad \text{and} \quad x \in \mathcal{B} \\ &x \in \mathcal{A} \cap \mathcal{B}. \end{aligned}$$

$\mathcal{A} \cap \mathcal{B}$ connex, well-founded: Clear.
(b). x transitive:

$$\begin{aligned} &u \in v \in x \in \mathcal{A} \\ &u \in v \in \mathcal{A} \\ &u \in \mathcal{A} \\ &u \in x \vee u = x \vee x \in u. \end{aligned}$$

From $u = x$ it follows that $u \in v \in u$ contradicting the lemma on $\in$-cycles, and from $x \in u$ it follows that $u \in v \in x \in u$, again contradicting the lemma on $\in$-cycles. Hence $u \in x$.

x connex, well-founded. Clear, since $x \subseteq \mathcal{A}$.

(c). $\leftarrow$. Clear, for $\mathcal{B}$ is transitive. $\rightarrow$. Let $\mathcal{A} \subseteq \mathcal{B}$. Without loss of generality $\mathcal{A} \subsetneq \mathcal{B}$. Choose $x \in \mathcal{B} \setminus \mathcal{A}$ such that $x \cap (\mathcal{B} \setminus \mathcal{A}) = \emptyset$ (this is possible, since $\mathcal{B}$ is well-founded); it suffices to show that $x = \mathcal{A}$.

$x \subseteq \mathcal{A}$. Assume $y \in x$, hence $y \in x \in \mathcal{B}$. Then $y \in \mathcal{A}$, for $x \cap (\mathcal{B} \setminus \mathcal{A}) = \emptyset$.

$\mathcal{A} \subseteq x$. Assume $y \in \mathcal{A}$. Then also $y \in \mathcal{B}$. It follows that $x \in y \vee x = y \vee y \in x$. But the first two cases are impossible, for in both of them we obtain $x \in \mathcal{A}$.

(d). Assume $\mathrm{Ord}(\mathcal{A})$ and $\mathrm{Ord}(\mathcal{B})$. Then by (a), $\mathrm{Ord}(\mathcal{A} \cap \mathcal{B})$. Using (c) we obtain

$$[(\mathcal{A} \cap \mathcal{B} \in \mathcal{A}) \vee (\mathcal{A} \cap \mathcal{B} = \mathcal{A})] \wedge [(\mathcal{A} \cap \mathcal{B} \in \mathcal{B}) \vee (\mathcal{A} \cap \mathcal{B} = \mathcal{B})].$$

Distributing yields

$$(\mathcal{A} \cap \mathcal{B} \in \mathcal{A} \cap \mathcal{B}) \vee (\mathcal{A} \in \mathcal{B}) \vee (\mathcal{B} \in \mathcal{A}) \vee (\mathcal{A} = \mathcal{B}).$$

But the first case $\mathcal{A} \cap \mathcal{B} \in \mathcal{A} \cap \mathcal{B}$ is impossible by the lemma on $\in$-cycles. □

LEMMA. (a) $\mathrm{Ord}(\mathrm{On})$.
(b) On *is not a set.*
(c) On *is the only proper ordinal class.*

PROOF. (a). On is transitive by part (b) of the lemma on ordinal classes and it is connex by part (d). On is well-founded: Let $a \subseteq \mathrm{On}$, $a \neq \emptyset$. Choose $x \in a$. Without loss of generality $x \cap a \neq \emptyset$. Since x is well-founded, there is a $y \in x \cap a$ such that $y \cap x \cap a = \emptyset$. It follows that $y \in a$ and $y \cap a = \emptyset$; the latter holds since $y \subseteq x$ because of $y \in x$, x transitive.

(b). Assume On is a set. Then $\mathrm{On} \in \mathrm{On}$, contradicting the lemma on $\in$-cycles.

(c). Let $\mathrm{Ord}(\mathcal{A})$, $\mathcal{A}$ not a set. By part (d) of the lemma on ordinal classes

$$\mathcal{A} \in \mathrm{On} \vee \mathcal{A} = \mathrm{On} \vee \mathrm{On} \in \mathcal{A}.$$

The first and the last case are excluded, for then $\mathcal{A}$ (or On, resp.) would be a set. □

LEMMA. (a) On *is inductive.*
(b) $n, \omega \in \mathrm{On}$.

PROOF. (a). $0 \in \mathrm{On}$ is clear. Let $x \in \mathrm{On}$. We must show $x + 1 \in \mathrm{On}$, that is $x \cup \{x\} \in \mathrm{On}$.

$x \cup \{x\}$ transitive: Assume $u \in v \in x \cup \{x\}$, so $u \in v \in x$ or $u \in v = x$. In both cases it follows that $u \in x$.

$x \cup \{x\}$ is connex: Assume $u, v \in x \cup \{x\}$. Then

$$u, v \in x \vee (u \in x \wedge v = x) \vee (u = x \wedge v \in x) \vee (u = v = x)$$
$$u \in v \vee u = v \vee v \in u.$$

$x \cup \{x\}$ is well-founded: Let $a \subseteq x \cup \{x\}$, $a \neq \emptyset$. We must show $\exists_{y \in a}(y \cap a = \emptyset)$. *Case* $a \cap x \neq \emptyset$. Then the claim follows from the well-foundedness of x. *Case* $a \cap x = \emptyset$. Then $a = \{x\}$, and we have $x \cap \{x\} = \emptyset$.

(b). This has been proved above, in 5.3.5 and 5.3.3 ("Principle of least element for ω"). □

LEMMA. $x, y \in \mathrm{On} \to x + 1 = y + 1 \to x = y$.

PROOF. The proof is similar to the proof of the second Peano axiom in the theorem on Peano axioms above.

$$x + 1 = y + 1$$
$$x \in y + 1 \wedge y \in x + 1$$
$$(x \in y \wedge y \in x) \vee x = y.$$

Since the first case is impossible by the lemma on ordinal classes, we have $x = y$. □

LEMMA. $\mathcal{A} \subseteq \mathrm{On} \to \bigcup \mathcal{A} \in \mathrm{On} \vee \bigcup \mathcal{A} = \mathrm{On}$.

PROOF. It suffices to show $\mathrm{Ord}(\bigcup \mathcal{A})$. $\bigcup \mathcal{A}$ is transitive: Let $x \in y \in \bigcup \mathcal{A}$, so $x \in y \in z \in \mathcal{A}$ for some z. Then we have $x \in z \in \mathcal{A}$, since $\mathcal{A} \subseteq \mathrm{On}$. Hence $x \in \bigcup \mathcal{A}$.

$\bigcup \mathcal{A}$ is connex and well-founded: It suffices to prove $\bigcup \mathcal{A} \subseteq \mathrm{On}$. Let $x \in \bigcup \mathcal{A}$, hence $x \in y \in \mathcal{A}$ for some y. Then $x \in y$ and $y \in \mathrm{On}$, so $x \in \mathrm{On}$. □

REMARK. If $\mathcal{A} \subseteq \mathrm{On}$, then $\bigcup \mathcal{A}$ is the least upper bound of $\mathcal{A}$ w.r.t. the well-ordering $\in \cap(\mathrm{On} \times \mathrm{On})$ of On, for by definition of $\bigcup \mathcal{A}$ we have

$$x \in \mathcal{A} \to x \subseteq \bigcup \mathcal{A},$$
$$\forall_{x \in \mathcal{A}}(x \subseteq y) \to \bigcup \mathcal{A} \subseteq y.$$

We therefore also write $\sup \mathcal{A}$ for $\bigcup \mathcal{A}$.

Here are some examples of ordinals:

$$0$$
$$1 = 0 + 1$$
$$2 = 1 + 1$$
$$\vdots$$
$$\omega \quad \text{set by the infinity axiom}$$
$$\omega + 1$$
$$\omega + 2$$
$$\vdots$$
$$\omega \cdot 2 := \bigcup \{\, \omega + n \mid n \in \omega \,\} \qquad \text{by recursion on } \omega$$
$$\omega \cdot 2 + 1$$
$$\omega \cdot 2 + 2$$

$$\vdots$$

$$\omega \cdot 3 := \bigcup \{ \omega \cdot 2 + n \mid n \in \omega \}$$

$$\vdots$$

$$\omega \cdot 4$$

$$\vdots$$

$$\omega \cdot \omega := \omega^2 := \bigcup \{ \omega \cdot n \mid n \in \omega \}$$

$$\omega^2 + 1$$

$$\omega^2 + 2$$

$$\vdots$$

$$\omega^2 + \omega$$

$$\omega^2 + \omega + 1$$

$$\omega^2 + \omega + 2$$

$$\vdots$$

$$\omega^2 + \omega \cdot 2$$

$$\vdots$$

$$\omega^2 + \omega \cdot 3$$

$$\vdots$$

$$\omega^3$$

$$\vdots$$

$$\omega^4$$

$$\vdots$$

$$\omega^\omega$$

$$\omega^\omega + 1$$

$$\vdots$$

and so on.

α, β, γ will denote ordinals. α is a *successor number* if $\exists_\beta(\alpha = \beta + 1)$. α is a *limit* if α is neither 0 nor a successor number. We write

$$\mathrm{Lim}(\alpha) \text{ for } \alpha \neq 0 \wedge \neg\exists_\beta(\alpha = \beta + 1).$$

Clearly for arbitrary α either $\alpha = 0$ or α is successor number or α is a limit.

LEMMA (Characterization of limits).

(a) $\mathrm{Lim}(\alpha) \leftrightarrow \alpha \neq 0 \wedge \forall_{\beta\in\alpha}(\beta + 1 \in \alpha)$.
(b) $\mathrm{Lim}(\omega)$.
(c) $\mathrm{Lim}(\alpha) \to \omega \subseteq \alpha$.

PROOF. (a). $\to$: Let $\beta \in \alpha$. Then $\beta+1 \in \alpha \vee \beta+1 = \alpha \vee \alpha \in \beta+1$. The second case $\beta + 1 = \alpha$ is excluded assumption. In the third case it follows that $\alpha \in \beta \vee \alpha = \beta$; but because of $\beta \in \alpha$ both are impossible by the lemma on $\in$-cycles. $\leftarrow$. Let $\alpha \neq 0$ and assume $\forall_{\beta\in\alpha}(\beta + 1 \in \alpha)$. Then if α is not a limit, we must have $\alpha = \beta + 1$. Then we obtain $\beta \in \alpha$, hence by assumption also $\beta + 1 \in \alpha$ and hence $\alpha \in \alpha$, which is impossible.

(b). Follows from (a), since ω is inductive.

(c). Assume $\mathrm{Lim}(\alpha)$. We show $n \in \alpha$ by induction on n. 0. We have $0 \in \alpha \vee 0 = \alpha \vee \alpha \in 0$, where the cases two and three clearly are impossible. $n + 1$. We have $n \in \alpha$ by induction hypothesis, hence $n + 1 \in \alpha$ by (a). □

LEMMA (Properties of limit ordinals). (a) $\alpha = \bigcup_{\beta\in\alpha}(\beta + 1)$.
(b) *For limits α we have* $\alpha = \bigcup_{\beta\in\alpha} \beta$.

PROOF. (a). $\subseteq$. Let $\beta \in \alpha$. The claim follows from $\beta \in \beta + 1$. $\supseteq$. Let $\beta \in \alpha$. Then $\beta + 1 \subseteq \alpha$.

(b). $\subseteq$. Let $\gamma \in \alpha$. Then $\gamma \in \gamma + 1 \in \alpha$. $\supseteq$. Let $\gamma \in \beta \in \alpha$. We obtain $\gamma \in \alpha$. □

THEOREM (Transfinite induction on On; class form).

$$\forall_{\alpha\subseteq\mathcal{B}}(\alpha \in \mathcal{B}) \to \mathrm{On} \subseteq \mathcal{B}.$$

PROOF. This is a special case of the induction theorem in 5.3.2. □

COROLLARY (Different forms of transfinite induction on On). *First form:*

$$\begin{aligned} A(0) &\to \forall_\alpha(A(\alpha) \to A(\alpha + 1)) \\ &\to \forall_\alpha(\mathrm{Lim}(\alpha) \to \forall_{\beta\in\alpha} A(\beta) \to A(\alpha)) \\ &\to \forall_\alpha A(\alpha). \end{aligned}$$

Second form: (Transfinite induction on On*, using all predecessors).*

$$\forall_\alpha(\forall_{\beta\in\alpha} A(\beta) \to A(\alpha)) \to \forall_\alpha A(\alpha).$$

Third form: (Principle of least element for On).

$$\exists_\alpha A(\alpha) \to \exists_\alpha(A(\alpha) \wedge \neg\exists_{\beta\in\alpha} A(\beta)).$$

PROOF. The third form follows from the second by contraposition. Also, the first form follows easily from the second. The second form follows from the theorem using $\mathcal{B} := \{\, \alpha \mid A(\alpha) \,\}$. □

THEOREM (Transfinite recursion on On). *Let $\mathcal{G}\colon V \to V$. Then there is exactly one function $\mathcal{F}\colon \mathrm{On} \to V$ such that for all α*

$$\mathcal{F}(\alpha) = \mathcal{G}(\mathcal{F}{\restriction}\alpha).$$

PROOF. This is a special case of the recursion theorem. □

COROLLARY. *Assume $\mathcal{G}\colon V \to V$, $\mathcal{H}\colon V \to V$ and a is a set. Then there is a unique function $\mathcal{F}\colon \mathrm{On} \to V$ such that*

$$\begin{aligned}
&\mathcal{F}(0) = a,\\
&\mathcal{F}(\alpha+1) = \mathcal{G}(\mathcal{F}(\alpha)),\\
&\mathcal{F}(\alpha) = \mathcal{H}(\mathcal{F}{\restriction}\alpha) \qquad \textit{for } \alpha \textit{ limit.}
\end{aligned}$$

PROOF. First observe that $\bigcup(\alpha+1) = \alpha$, because of

$$\begin{aligned}
\gamma \in \bigcup(\alpha+1) &\leftrightarrow \exists_\beta(\gamma \in \beta \in \alpha+1)\\
&\leftrightarrow \exists_\beta(\gamma \in \beta \subseteq \alpha)\\
&\leftrightarrow \gamma \in \alpha.
\end{aligned}$$

For given a, $\mathcal{G}$ and $\mathcal{H}$ we shall find a $\mathcal{G}'$ such that

$$\begin{aligned}
&\mathcal{G}'(0) = a,\\
&\mathcal{G}'(\mathcal{F}{\restriction}\alpha+1) = \mathcal{G}(\mathcal{F}(\alpha)),\\
&\mathcal{G}'(\mathcal{F}{\restriction}\alpha) = \mathcal{H}(\mathcal{F}{\restriction}\alpha) \qquad \text{for } \alpha \text{ limit.}
\end{aligned}$$

We define a function $\mathcal{G}'\colon V \to V$ by

$$\mathcal{G}'(x) = \begin{cases} a, & \text{otherwise;}\\ \mathcal{G}(x(\bigcup \mathrm{dom}(x))), & \text{if } \exists_\beta(\mathrm{dom}(x) = \beta+1);\\ \mathcal{H}(x), & \text{if } \mathrm{Lim}(\mathrm{dom}(x)). \end{cases}$$

By the recursion theorem there is a unique $\mathcal{F}\colon \mathrm{On} \to V$ such that, for all α,

$$\mathcal{F}(\alpha) = \mathcal{G}'(\mathcal{F}{\restriction}\alpha).$$

Clearly this property of $\mathcal{F}$ is equivalent to the equations above. □

5.3.6. Regularity axiom, von Neumann levels, rank. Recall the cumulative type structure:

$$\begin{aligned}
&\text{Level 0:} \quad -\\
&\text{Level 1:} \quad \emptyset\\
&\text{Level 2:} \quad \emptyset, \{\emptyset\}
\end{aligned}$$

$$\text{Level 3:}\quad \emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \{\emptyset, \{\emptyset\}\}$$
$$\text{and so on.}$$

Using ordinals we can now consider transfinite levels as well. The level ω consists of all sets whose elements are formed on finite levels, and the level $\omega+1$ consists of all sets whose elements are formed on finite levels or at level ω, and so on. Generally we define the *von Neumann levels* V_α as follows, by transfinite recursion on On.

$$\begin{aligned} V_0 &= \emptyset, \\ V_{\alpha+1} &= \mathcal{P}(V_\alpha), \\ V_\alpha &= \bigcup_{\beta\in\alpha} V_\beta \qquad \text{for } \alpha \text{ limit.} \end{aligned}$$

REMARK. More precisely, $V_\alpha := \mathcal{F}(\alpha)$, where $\mathcal{F}\colon \mathrm{On} \to V$ is defined as follows, by transfinite recursion on On:

$$\begin{aligned} \mathcal{F}(0) &= \emptyset, \\ \mathcal{F}(\alpha+1) &= \mathcal{P}(\mathcal{F}(\alpha)), \\ \mathcal{F}(\alpha) &= \bigcup \mathrm{rng}(\mathcal{F}\restriction\alpha) \qquad \text{for } \alpha \text{ limit.} \end{aligned}$$

LEMMA. (a) *V_α is transitive.*
(b) $\alpha \in \beta \to V_\alpha \in V_\beta$.
(c) $\alpha \subseteq \beta \to V_\alpha \subseteq V_\beta$.
(d) $V_\alpha \cap \mathrm{On} = \alpha$.

PROOF. (a). (Transfinite) induction on α. 0. $\emptyset$ is transitive. $\alpha+1$.

$$\begin{aligned} &x \in y \in V_{\alpha+1} = \mathcal{P}(V_\alpha) \\ &x \in y \subseteq V_\alpha \\ &x \in V_\alpha \\ &x \subseteq V_\alpha && \text{by induction hypothesis} \\ &x \in V_{\alpha+1}. \end{aligned}$$

α limit.

$$\begin{aligned} &x \in y \in V_\alpha = \bigcup_{\beta\in\alpha} V_\beta \\ &x \in y \in V_\beta && \text{for some } \beta \in \alpha \\ &x \in V_\beta && \text{by induction hypothesis} \\ &x \in V_\alpha. \end{aligned}$$

(b). Induction on β. 0. Clear. $\beta+1$.

$$\alpha \in \beta + 1$$

$$\begin{aligned}
&\alpha \in \beta \text{ or } \alpha = \beta \\
&V_\alpha \in V_\beta \text{ or } V_\alpha = V_\beta \quad \text{by induction hypothesis} \\
&V_\alpha \subseteq V_\beta \qquad\qquad\quad \text{by (a)} \\
&V_\alpha \in V_{\beta+1}.
\end{aligned}$$

β limit.

$$\begin{aligned}
&\alpha \in \beta \\
&\alpha + 1 \in \beta \\
&V_\alpha \in V_{\alpha+1} \subseteq \bigcup_{\gamma\in\beta} V_\gamma = V_\beta.
\end{aligned}$$

(c). Using $\alpha \subseteq \beta \leftrightarrow \alpha \in \beta \vee \alpha = \beta$ the claim follows from (a) and (b).
(d). Induction on α. 0. Clear. $\alpha + 1$.

$$\begin{aligned}
\beta \in V_{\alpha+1} &\leftrightarrow \beta \subseteq V_\alpha \\
&\leftrightarrow \beta \subseteq V_\alpha \cap \mathrm{On} = \alpha \quad \text{by induction hypothesis} \\
&\leftrightarrow \beta \in \alpha + 1.
\end{aligned}$$

α limit.

$$\begin{aligned}
V_\alpha \cap \mathrm{On} &= (\bigcup_{\beta\in\alpha} V_\beta) \cap \mathrm{On} \\
&= \bigcup_{\beta\in\alpha} (V_\beta \cap \mathrm{On}) \\
&= \bigcup_{\beta\in\alpha} \beta \quad \text{by induction hypothesis} \\
&= \alpha.
\end{aligned}$$

□

We now show that the von Neumann levels exhaust the universe, which means that $V = \bigcup_{\alpha\in\mathrm{On}} V_\alpha$. However, this requires another axiom, the *regularity axiom*, which says that the relation $\in$ on V is well-founded, i.e.,

Axiom (Regularity).

$$\forall_a(a \neq \emptyset \to \exists_{x\in a}(x \cap a = \emptyset)).$$

We want to assign to every set x an ordinal α, namely the least α such that $x \subseteq V_\alpha$. To this end we need the notion of the *rank* $\mathrm{rn}(x)$ of a set x, which is defined recursively by

$$\mathrm{rn}(x) := \bigcup\{\, \mathrm{rn}(y) + 1 \mid y \in x \,\}.$$

More precisely we define $\mathrm{rn}(x) := \mathcal{F}(x)$, where $\mathcal{F}\colon V \to V$ is defined as follows (using the recursion theorem for well-founded relations):

$$\mathcal{F}(x) := \bigcup \mathrm{rng}(\mathcal{H}(\mathcal{F}\restriction x))$$

with

$$\mathcal{H}(z) := \{\, (u, v+1) \mid (u,v) \in z \,\}.$$

We first show that $\mathrm{rn}(x)$ has the property formulated above.

LEMMA. (a) $\mathrm{rn}(x) \in \mathrm{On}$.
(b) $x \subseteq V_{\mathrm{rn}(x)}$.
(c) $x \subseteq V_\alpha \to \mathrm{rn}(x) \subseteq \alpha$.

PROOF. (a). $\in$-induction on x. We have $\mathrm{rn}(x) = \bigcup\{\, \mathrm{rn}(y) + 1 \mid y \in x \,\} \in \mathrm{On}$, for by induction hypothesis $\mathrm{rn}(y) \in \mathrm{On}$ for every $y \in x$.

(b). $\in$-induction on x. Let $y \in x$. Then $y \subseteq V_{\mathrm{rn}(y)}$ by induction hypothesis, hence $y \in \mathcal{P}(V_{\mathrm{rn}(y)}) = V_{\mathrm{rn}(y)+1} \subseteq V_{\mathrm{rn}(x)}$ because of $\mathrm{rn}(y) + 1 \subseteq \mathrm{rn}(x)$.

(c). Induction on α. Let $x \subseteq V_\alpha$. We must show $\mathrm{rn}(x) = \bigcup\{\, \mathrm{rn}(y) + 1 \mid y \in x \,\} \subseteq \alpha$. Let $y \in x$. We must show $\mathrm{rn}(y) + 1 \subseteq \alpha$. Because of $x \subseteq V_\alpha$ we have $y \in V_\alpha$. This implies $y \subseteq V_\beta$ for some $\beta \in \alpha$, for in case $\alpha = \alpha' + 1$ we have $y \in V_{\alpha'+1} = \mathcal{P}(V_{\alpha'})$ and hence $y \subseteq V_{\alpha'}$, and in case α limit we have $y \in V_\alpha = \bigcup_{\beta\in\alpha} V_\beta$, hence $y \in V_\beta$ and therefore $y \subseteq V_\beta$ for some $\beta \in \alpha$. – By induction hypothesis it follows that $\mathrm{rn}(y) \subseteq \beta$, whence $\mathrm{rn}(y) \in \alpha$. □

Now we obtain easily the proposition formulated above as our goal.

COROLLARY. $V = \bigcup_{\alpha\in\mathrm{On}} V_\alpha$.

PROOF. $\supseteq$ is clear. $\subseteq$. For every x we have $x \subseteq V_{\mathrm{rn}(x)}$ by part (b) of the lemma, hence $x \in V_{\mathrm{rn}(x)+1}$. □

Now V_α can be characterized as the set of all sets of rank less than α.

LEMMA. $V_\alpha = \{\, x \mid \mathrm{rn}(x) \in \alpha \,\}$.

PROOF. $\supseteq$. Let $\mathrm{rn}(x) \in \alpha$. Then $x \subseteq V_{\mathrm{rn}(x)}$ implies $x \in V_{\mathrm{rn}(x)+1} \subseteq V_\alpha$.

$\subseteq$. Induction on α. *Case* 0. Clear. *Case* $\alpha + 1$. Let $x \in V_{\alpha+1}$. Then $x \in \mathcal{P}(V_\alpha)$, hence $x \subseteq V_\alpha$. For every $y \in x$ we have $y \in V_\alpha$ and hence $\mathrm{rn}(y) \in \alpha$ by induction hypothesis, so $\mathrm{rn}(y) + 1 \subseteq \alpha$. Therefore $\mathrm{rn}(x) = \bigcup\{\, \mathrm{rn}(y) + 1 \mid y \in x \,\} \subseteq \alpha$. *Case* α limit. Let $x \in V_\alpha$. Then $x \in V_\beta$ for some $\beta \in \alpha$, hence $\mathrm{rn}(x) \in \beta$ by induction hypothesis, hence $\mathrm{rn}(x) \in \alpha$. □

From $x \in y$ and $x \subseteq y$, resp., we can infer the corresponding relations between the ranks.

LEMMA. (a) $x \in y \to \mathrm{rn}(x) \in \mathrm{rn}(y)$.
(b) $x \subseteq y \to \mathrm{rn}(x) \subseteq \mathrm{rn}(y)$.

PROOF. (a). Because of $\mathrm{rn}(y) = \bigcup\{\, \mathrm{rn}(x)+1 \mid x \in y \,\}$ this is clear. (b). For every $z \in x$ we have $\mathrm{rn}(z) \in \mathrm{rn}(y)$ by (a), hence $\mathrm{rn}(x) = \bigcup\{\, \mathrm{rn}(z)+1 \mid z \in x \,\} \subseteq \mathrm{rn}(y)$. □

Moreover we can show that the sets α and V_α both have rank α.

LEMMA. (a) $\mathrm{rn}(\alpha) = \alpha$.
(b) $\mathrm{rn}(V_\alpha) = \alpha$.

PROOF. (a). Induction on α. We have $\mathrm{rn}(\alpha) = \bigcup\{\, \mathrm{rn}(\beta)+1 \mid \beta \in \alpha \,\}$, hence by induction hypothesis $\mathrm{rn}(\alpha) = \bigcup\{\, \beta+1 \mid \beta \in \alpha \,\} = \alpha$ by the lemma on properties of limit ordinals.

(b). We have

$$\begin{aligned}\mathrm{rn}(V_\alpha) &= \bigcup\{\, \mathrm{rn}(x)+1 \mid x \in V_\alpha \,\} \\ &= \bigcup\{\, \mathrm{rn}(x)+1 \mid \mathrm{rn}(x) \in \alpha \,\} \quad \text{by the next to last lemma} \\ &\subseteq \alpha.\end{aligned}$$

Conversely, let $\beta \in \alpha$. Then $\beta \in V_\alpha$ since $\alpha \subseteq V_\alpha$, hence by (a) $\beta = \mathrm{rn}(\beta) \in \mathrm{rn}(V_\alpha)$. □

We finally show that a class $\mathcal{A}$ is a set if and only if the ranks of their elements can be bounded by an ordinal.

LEMMA. *$\mathcal{A}$ is set iff there is an α such that $\forall_{y \in \mathcal{A}}(\mathrm{rn}(y) \in \alpha)$.*

PROOF. $\to$. Let $\mathcal{A} = x$. From part (a) of the next to last lemma we obtain that $\mathrm{rn}(x)$ is the α we need.
$\leftarrow$. Assume $\forall_{y \in \mathcal{A}}(\mathrm{rn}(y) \in \alpha)$. Then $\mathcal{A} \subseteq \{\, y \mid \mathrm{rn}(y) \in \alpha \,\} = V_\alpha$. □

5.4. Cardinals

We now introduce cardinals and develop their basic properties.

5.4.1. Size comparison between sets. Define

$$\begin{aligned}|a| \le |b| &:\leftrightarrow \exists_f(f\colon a \to b \text{ and } f \text{ injective}),\\ |a| = |b| &:\leftrightarrow \exists_f(f\colon a \leftrightarrow b),\\ |a| < |b| &:\leftrightarrow |a| \le |b| \wedge |a| \ne |b|,\\ {}^b a &:= \{\, f \mid f\colon b \to a \,\}.\end{aligned}$$

Two sets a and b are *equinumerous* if $|a| = |b|$. Notice that we did not define $|a|$, but only the relations $|a| \le |b|$, $|a| = |b|$ and $|a| < |b|$.

The following properties are clear:

$$|a \times b| = |b \times a|;$$
$$|{}^a({}^b c)| = |{}^{a\times b} c|;$$
$$|\mathcal{P}(a)| = |{}^a\{0,1\}|.$$

THEOREM (Cantor). $|a| < |\mathcal{P}(a)|$.

PROOF. Clearly $f\colon a \to \mathcal{P}(a)$, $x \mapsto \{x\}$ is injective. Assume that we have $g\colon a \leftrightarrow \mathcal{P}(a)$. Consider

$$b := \{\, x \mid x \in a \wedge x \notin g(x) \,\}.$$

Then $b \subseteq a$, hence $b = g(x_0)$ for some $x_0 \in a$. It follows that $x_0 \in g(x_0) \leftrightarrow x_0 \notin g(x_0)$ and hence a contradiction. □

THEOREM (Cantor, Bernstein). *If $a \subseteq b \subseteq c$ and $|a| = |c|$, then $|b| = |c|$.*

PROOF. Let $f\colon c \to a$ be bijective and $r := c \setminus b$. We recursively define $g\colon \omega \to V$ by

$$g(0) = r,$$
$$g(n+1) = f[g(n)].$$

Let

$$\overline{r} := \bigcup_n g(n)$$

and define $i\colon c \to b$ by

$$i(x) := \begin{cases} f(x), & \text{if } x \in \overline{r}, \\ x, & \text{if } x \notin \overline{r}. \end{cases}$$

It suffices to show that (a) $\mathrm{rng}(i) = b$ and (b) i is injective. Ad (a). Let $x \in b$. We must show $x \in \mathrm{rng}(i)$. Without loss of generality let $x \in \overline{r}$. Because of $x \in b$ we then have $x \notin g(0)$. Hence there is an n such that $x \in g(n+1) = f[g(n)]$, so $x = f(y) = i(y)$ for some $y \in \overline{r}$. Ad (b). Let $x \neq y$. Without loss of generality $x \in \overline{r}$, $y \notin \overline{r}$. But then $i(x) \in \overline{r}$, $i(y) \notin \overline{r}$, hence $i(x) \neq i(y)$. □

REMARK. The theorem of Cantor and Bernstein can be seen as an application of the fixed point theorem of Knaster-Tarski.

COROLLARY. $|a| \leq |b| \to |b| \leq |a| \to |a| = |b|$.

PROOF. Let $f\colon a \to b$ and $g\colon b \to a$ injective. Then $(g \circ f)[a] \subseteq g[b] \subseteq a$ and $|(g \circ f)[a]| = |a|$. By the theorem of Cantor and Bernstein $|b| = |g[b]| = |a|$. □

5.4.2. Cardinals, aleph function. A cardinal is defined to be an ordinal that is not equinumerous to a smaller ordinal:

$$\alpha \text{ is a } \textit{cardinal} \text{ if } \forall_{\beta<\alpha}(|\beta| \neq |\alpha|).$$

Here and later we write – because of the lemma on ordinal classes - $\alpha < \beta$ for $\alpha \in \beta$ and $\alpha \leq \beta$ for $\alpha \subseteq \beta$.

LEMMA. $|n| = |m| \to n = m$.

PROOF. Induction on n. 0. Clear. $n+1$. Let $f\colon n+1 \leftrightarrow m+1$. We may assume $f(n) = m$. Hence $f{\restriction}n\colon n \leftrightarrow m$ and therefore $n = m$ by induction hypothesis, hence also $n + 1 = m + 1$. □

COROLLARY. *n is a cardinal.*

LEMMA. $|n| \neq |\omega|$.

PROOF. Assume $|n| = |\omega|$. Because of $n \subseteq n + 1 \subseteq \omega$ the theorem of Cantor and Bernstein implies $|n| = |n + 1|$, a contradiction. □

COROLLARY. *ω is a cardinal.*

LEMMA. $\omega \leq \alpha \to |\alpha + 1| = |\alpha|$.

PROOF. Define $f\colon \alpha \to \alpha + 1$ by

$$f(x) := \begin{cases} \alpha, & \text{if } x = 0; \\ n, & \text{if } x = n + 1; \\ x, & \text{otherwise.} \end{cases}$$

Then $f\colon \alpha \leftrightarrow \alpha + 1$. □

COROLLARY. *If $\omega \leq \alpha$ and α is a cardinal, then α is a limit.*

PROOF. Assume $\alpha = \beta + 1$. Then $\omega \leq \beta < \alpha$, hence $|\beta| = |\beta + 1|$, contradicting the assumption that α is a cardinal. □

LEMMA. *If a is a set of cardinals, then* $\sup(a)$ *(*$:= \bigcup a$*) is a cardinal.*

PROOF. Otherwise there would be an $\alpha < \sup(a)$ such that $|\alpha| = |\sup(a)|$. Hence $\alpha \in \bigcup a$ and therefore $\alpha \in \beta \in a$ for some cardinal β. By the theorem of Cantor and Bernstein from $\alpha \subseteq \beta \subseteq \bigcup a$ and $|\alpha| = |\bigcup a|$ it follows that $|\alpha| = |\beta|$. Because of $\alpha \in \beta$ and β a cardinal this is impossible. □

We now show that for every ordinal there is a strictly bigger cardinal. More generally, even the following holds:

THEOREM (Hartogs).

$$\forall_a \exists^{=1}_\alpha (\forall_{\beta<\alpha}(|\beta| \leq |a|) \wedge |\alpha| \not\leq |a|).$$

α is the Hartogs number *of a, denoted $\mathcal{H}(a)$.*

PROOF. Uniqueness. Clear. Existence. Let $w := \{\, (b,r) \mid b \subset a \wedge r \text{ well-ordering on } b \,\}$ and $\gamma_{(b,r)}$ the uniquely determined ordinal isomorphic to (b,r). Then $\{\, \gamma_{(b,r)} \mid (b,r) \in w \,\}$ is a transitive subset of On, hence an ordinal α. We must show

(a) $\beta < \alpha \to |\beta| \leq |a|$,
(b) $|\alpha| \not\leq |a|$.

(a). Let $\beta < \alpha$. Then β is isomorphic to a $\gamma_{(b,r)}$ with $(b,r) \in w$, hence there exists an $f\colon \beta \leftrightarrow b$.

(b). Assume $f\colon \alpha \to a$ is injective. Then $\alpha = \gamma_{(b,r)}$ for some $b \subseteq a$ ($b := \mathrm{rng}(f)$), hence $\alpha \in \alpha$, a contradiction. □

REMARK. (a). The Hartogs number of a is a cardinal. For let α be the Hartogs number of a, $\beta < \alpha$. If $|\beta| = |\alpha|$, we would have $|\alpha| = |\beta| \leq |a|$, a contradiction.

(b). The Hartogs number of β is the least cardinal α such that $\alpha > \beta$.

The aleph function $\aleph\colon \mathrm{On} \to V$ is defined recursively by

$$\begin{aligned}
&\aleph_0 := \omega,\\
&\aleph_{\alpha+1} := \mathcal{H}(\aleph_\alpha),\\
&\aleph_\alpha := \sup\{\, \aleph_\beta \mid \beta < \alpha \,\} \text{ for } \alpha \text{ limit.}
\end{aligned}$$

LEMMA (Properties of $\aleph$). (a) *$\aleph_\alpha$ is a cardinal.*
(b) $\alpha < \beta \to \aleph_\alpha < \aleph_\beta$.
(c) $\forall_\beta(\beta \text{ cardinal } \to \omega \leq \beta \to \exists_\alpha(\beta = \aleph_\alpha))$.

PROOF. (a). Induction on α; clear. (b). Induction on β. 0. Clear. $\beta + 1$.

$$\begin{aligned}
&\alpha < \beta + 1\\
&\alpha < \beta \vee \alpha = \beta\\
&\aleph_\alpha < \aleph_\beta \vee \aleph_\alpha = \aleph_\beta\\
&\aleph_\alpha < \aleph_{\beta+1}.
\end{aligned}$$

β limit.

$$\begin{aligned}
&\alpha < \beta\\
&\alpha < \gamma \text{ for some } \gamma < \beta\\
&\aleph_\alpha < \aleph_\gamma \leq \aleph_\beta.
\end{aligned}$$

(c). Let α be minimal such that $\beta \leq \aleph_\alpha$. Such an α exists, for otherwise $\aleph\colon \mathrm{On} \to \beta$ would be injective. We show $\aleph_\alpha \leq \beta$ by cases on α. 0. Clear. $\alpha = \alpha' + 1$. By the choice of α we have $\aleph_{\alpha'} < \beta$, hence $\aleph_\alpha \leq \beta$ (since $\aleph_\alpha = \mathcal{H}(\aleph_{\alpha'})$ is the least cardinal $> \alpha$). α limit. By the choice of α we have $\aleph_\gamma < \beta$ for all $\gamma < \alpha$, hence $\aleph_\alpha = \sup\{\, \aleph_\gamma \mid \gamma < \alpha \,\} \leq \beta$. □

We show that every infinite ordinal is equinumerous to a cardinal.

LEMMA. $\forall_{\beta \geq \omega} \exists_\alpha (|\beta| = |\aleph_\alpha|)$.

PROOF. Consider $\delta := \min\{\, \gamma \mid \gamma < \beta \wedge |\gamma| = |\beta| \,\}$. Clearly δ is a cardinal. Moreover $\delta \geq \omega$, for otherwise

$$\begin{aligned} &\delta = n \\ &|n| = |\beta| \\ &n \subseteq n+1 \subseteq \beta \\ &|n| = |n+1|, \end{aligned}$$

a contradiction. Hence $\delta = \aleph_\alpha$ for some α, and therefore $|\delta| = |\beta| = |\aleph_\alpha|$. □

5.4.3. Products of cardinals. We now show that $|\aleph_\alpha \times \aleph_\alpha| = |\aleph_\alpha|$. On the set $\mathrm{On} \times \mathrm{On}$ we define a relation $\prec$ by

$$\begin{aligned} (\alpha, \beta) \prec (\gamma, \delta) :\leftrightarrow &\max\{\alpha, \beta\} < \max\{\gamma, \delta\} \vee \\ &(\max\{\alpha, \beta\} = \max\{\gamma, \delta\} \wedge \alpha < \gamma) \vee \\ &(\max\{\alpha, \beta\} = \max\{\gamma, \delta\} \wedge \alpha = \gamma \wedge \beta < \delta). \end{aligned}$$

LEMMA. *$\prec$ is a well-ordering on* $\mathrm{On} \times \mathrm{On}$.

PROOF. Clearly $\prec$ is a linear ordering. To see the well-foundedness of $\prec$ consider an $a \subseteq \mathrm{On} \times \mathrm{On}$ such that $a \neq \emptyset$. Then

$$\emptyset \neq \mathcal{A} := \{\, \alpha \mid \exists_{\rho,\mu}((\rho, \mu) \in a \wedge \max\{\rho, \mu\} = \alpha) \,\} \subseteq \mathrm{On}.$$

Let $\alpha_0 := \min(\mathcal{A})$. Then

$$\emptyset \neq \mathcal{A}_1 := \{\, \rho \mid \exists \mu((\rho, \mu) \in a \wedge \max\{\rho, \mu\} = \alpha_0) \,\} \subseteq \mathrm{On}.$$

Let $\rho_0 := \min(\mathcal{A}_1)$. Then

$$\emptyset \neq \mathcal{A}_2 := \{\, \mu \mid (\rho_0, \mu) \in a \wedge \max\{\rho_0, \mu\} = \alpha_0) \,\} \subseteq \mathrm{On}.$$

Let $\mu_0 := \min(\mathcal{A}_2)$. Then clearly $(\rho_0, \mu_0) = \min_\prec(a)$. Finally notice that $\widehat{(\alpha, \beta)}$ must be a set, for $\widehat{(\alpha, \beta)} \subseteq \gamma \times \gamma$ with $\gamma := \max\{\alpha, \beta\} + 1$. □

COROLLARY. $\mathrm{On} \times \mathrm{On}$ *is isomorphic to* On *(w.r.t.* $\prec$ *and* $\in \restriction \mathrm{On}$*).*

PROOF. By the lemma $\prec$ is a well-ordering on On × On. Hence by a corollary to Mostowski's isomorphy theorem there is an isomorphism onto a transitive and hence also ordinal class. This class cannot possibly be a set, for then On × On would be a set as well. But by a lemma above On is the only proper ordinal class. □

THEOREM. *$\aleph_\alpha \times \aleph_\alpha$ is isomorphic to $\aleph_\alpha$ (w.r.t. the relations $\prec$ on $\aleph_\alpha \times \aleph_\alpha$ and $\in$ on $\aleph_\alpha$).*

PROOF. Assume: $\exists_\alpha(\aleph_\alpha \times \aleph_\alpha$ not isomorphic to $\aleph_\alpha)$. Let

$$\alpha_0 := \min\{\, \alpha \mid \aleph_\alpha \times \aleph_\alpha \text{ not isomorphic to } \aleph_\alpha \,\}.$$

Clearly $\alpha_0 \neq 0$. Since $\aleph_{\alpha_0} \times \aleph_{\alpha_0}$ and $\aleph_{\alpha_0}$ are well-ordered sets, one of them must be isomorphic to a proper initial segment of the other. Therefore we distinguish two cases.

Case (a). $\aleph_{\alpha_0}$ is isomorphic to $\widehat{(\beta,\gamma)}$ with $\beta, \gamma < \aleph_{\alpha_0}$. Choose $\delta < \aleph_{\alpha_0}$ with $\beta, \gamma < \delta$. Then $\widehat{(\beta,\gamma)} \subseteq \delta \times \delta$, and

$$\begin{aligned} |\aleph_{\alpha_0}| = |\widehat{(\beta,\gamma)}| \leq |\delta \times \delta| = |\aleph_\tau \times \aleph_\tau| \quad & \text{for some } \tau < \alpha_0 \\ = |\aleph_\tau| \quad & \text{by choice of } \alpha_0, \end{aligned}$$

hence a contradiction to the strict monotonicity of $\aleph$.

Case (b). $\aleph_{\alpha_0} \times \aleph_{\alpha_0}$ is isomorphic to $\beta < \aleph_{\alpha_0}$. Then

$$\begin{aligned} &|\aleph_{\alpha_0}| \leq |\aleph_{\alpha_0} \times \aleph_{\alpha_0}| = |\beta| \leq |\aleph_{\alpha_0}| \\ &|\aleph_{\alpha_0}| = |\beta|, \end{aligned}$$

hence a contradiction to the fact that $\aleph_{\alpha_0}$ is a cardinal, $\beta < \aleph_{\alpha_0}$. □

COROLLARY. (a) $|\aleph_\alpha \times \aleph_\beta| = |\max\{\aleph_\alpha, \aleph_\beta\}|$.
(b) $n \neq 0 \to |{}^n\aleph_\alpha| = |\aleph_\alpha|$.

PROOF. (a). We may assume $\alpha \leq \beta$. Then

$$|\aleph_\beta| \leq |\aleph_\alpha \times \aleph_\beta| \leq |\aleph_\beta \times \aleph_\beta| = |\aleph_\beta|.$$

(b). This follows easily from the theorem, by induction on n. □

5.5. The Axiom of Choice

5.5.1. Axiom of choice, well ordering theorem, Zorn's lemma. A relation $\mathcal{R}$ on $\mathcal{A}$ is a (strict) *partial ordering* if for all $x, y, z \in \mathcal{A}$

$$\begin{aligned} &\neg x\mathcal{R}x, && \text{irreflexivity} \\ &x\mathcal{R}y \to y\mathcal{R}z \to x\mathcal{R}z, && \text{transitivity.} \end{aligned}$$

An element $x \in \mathcal{A}$ is *maximal* if there is no $y \in \mathcal{A}$ such that $x\mathcal{R}y$. Let $\mathcal{B} \subseteq \mathcal{A}$. An element $x \in \mathcal{A}$ is an *upper bound* of $\mathcal{B}$ if

$$\forall_{y\in\mathcal{B}}(y\mathcal{R}x \vee y = x).$$

THEOREM. *The following are equivalent.*

(a) *The* axiom of choice (AC)

$$\forall_x(\emptyset \notin x \to \exists_f(f\colon x \to \bigcup x \wedge \forall_{y\in x}(f(y) \in y))).$$

(b) *The* well ordering theorem (WO)

$$\forall_a\exists_r(r \textit{ is a well ordering on } a).$$

(c) Zorn's Lemma (ZL)*: Let $(P, <)$ be a non empty partial ordering, with the property that every (by $<$) linearly ordered subset $L \subseteq P$ has an upper bound in P. Then P has a maximal element.*

PROOF. (ZL) $\to$ (WO). Let a be given, and define

$$P := \{\, f \mid \exists_\alpha(f\colon \alpha \to a \text{ injective}) \,\} \subseteq \mathcal{P}(\mathcal{H}(a) \times a).$$

P is partially ordered by proper inclusion $\subsetneq$. Let $L \subseteq P$ be linearly ordered. Then $\bigcup L \in P$, hence $\bigcup L$ is an upper bound of L. Zorn's Lemma then gives a maximal element $f_0 \in P$. Clearly f_0 is a bijection of an ordinal α_0 onto a, hence f_0 induces a well ordering on a.

(WO) $\to$ (AC). Let $\emptyset \notin x$. By (WO) there is a well ordering $<$ on $\bigcup x$. Clearly $<$ induces a well ordering on every $y \in x$. Define

$$\begin{aligned} f\colon x &\to \bigcup x, \\ y &\mapsto \min{}_<(y) \in y. \end{aligned}$$

(AC) $\to$ (ZL). Let $<$ be a partial ordering on $P \neq \emptyset$. Assume that every subset $L \subseteq P$ linearly ordered by $<$ has an upper bound in P. By (AC) there is a choice function f on $\mathcal{P}(P) \setminus \{\emptyset\}$. Let $z \notin P$ be arbitrary, and define

$$\mathcal{F}\colon \mathrm{On} \to V$$

$$\mathcal{F}(\alpha) = \begin{cases} f(\{\, y \mid y \in P \setminus \mathcal{F}[\alpha] \wedge y \text{ upper bound of } \mathcal{F}[\alpha] \,\}), & \text{if } \{\ldots\} \neq \emptyset; \\ z, & \text{otherwise.} \end{cases}$$

Then there is a ρ such that $\mathcal{F}(\rho) = z$, for otherwise $\mathcal{F}\colon \mathrm{On} \to P$ would be injective, contradicting our assumption that P is a set. Let $\rho_0 := \min\{\, \rho \mid \mathcal{F}(\rho) = z \,\}$. $\mathcal{F}[\rho_0]$ is linearly ordered, and we have $\mathcal{F}[\rho_0] \subseteq P$. By assumption there is an upper bound $y_0 \in P$ of $\mathcal{F}[\rho_0]$. We show that y_0 is a maximal element in P. So assume $y_0 < y$ for some $y \in P$. Then y is an upper bound of $\mathcal{F}[\rho_0]$ and $y \notin \mathcal{F}[\rho_0]$. But this contradicts the definition of ρ_0. $\square$

From now on we will assume the axiom of choice; however, we will mark every theorem and every definition depending on it by (AC).

(AC) clearly is equivalent to its special case where every two elements $y_1, y_2 \in x$ are disjoint. We hence note the following equivalent to the axiom of choice:

LEMMA. *The following are equivalent*

(a) *The axiom of choice (AC).*
(b) *For every surjective* $g\colon a \to b$ *there is an injective* $f\colon b \to a$ *such that* $\forall_{x\in b}(g(fx) = x)$.

PROOF. (a) $\Rightarrow$ (b). Assume $g\colon a \to b$ is surjective. By (AC) there is a well-ordering $<$ of a. Define $f\colon b \to a$ by $f(x) := \min_<\{\, x \mid x \in a \wedge g(x) = y \,\}$.

(b) $\Rightarrow$ (a). We may assume $x \neq \emptyset$ and $\forall_{y_1,y_2\in x}(y_1 \cap y_2 = \emptyset)$. Define $g\colon \bigcup x \to x$ by $g(z) :=$ the unique $y \in x$ such that $z \in y$. Then g is surjective. By our assumption there is an injective $f\colon x \to \bigcup x$ such that $g(f(y)) = y$ for all $y \in x$, hence $f(y) \in y$. □

5.5.2. Cardinality. α is the *cardinality* of a if α is a cardinal and there is a bijection $f\colon a \to \alpha$.

THEOREM (AC). *Every set has a unique cardinality.*

PROOF. Uniqueness. Clear. Existence. Let $<$ be a well–ordering on a. Then there is a γ such that a is isomorphic to γ. Hence $\{\, \tau \mid |\tau| = |a| \,\} \neq \emptyset$ and therefore $\min\{\, \tau \mid |\tau| = |a| \,\}$ is a cardinal. □

Clearly $|a| = |b|$ iff the cardinality of a equals the cardinality of b, and $|a| \leq |b|$ iff the cardinality of a is less than or equal to the cardinality of b. Therefore we can use $|a|$ as a notation for the cardinality of a.

A set a is defined to be *finite* if a can be mapped bijectively onto a natural number, and *infinite* otherwise. Using (AC) it follows that a is finite iff $|a| < \omega$.

LEMMA (AC). *If* $a, b \neq \emptyset$ *and* a *or* b *is infinite, then*

$$|a \times b| = \max\{|a|, |b|\}.$$

PROOF. Let $|a| = \max\{|a|, |b|\}$. Then

$$|a| \leq |a \times b| = ||a| \times |b|| \leq ||a| \times |a|| = |a|. \qquad \square$$

THEOREM (AC; Cardinality of unions). *Let* I *be infinite or* $\sup_{i\in I} |A_i|$ *be infinite. Then*

(a) $|\bigcup_{i\in I} A_i| \leq \max\{|I|, \sup_{i\in I} |A_i|\}$.

(b) *If in addition* $\forall_{i\in I}(A_i \neq \emptyset)$ *and* $\forall_{i,j\in I}(i \neq j \to A_i \cap A_j = \emptyset)$, *then equality holds.*

PROOF. (a). We may assume $\kappa := \sup_{i\in I} |A_i| \neq 0$. Choose a well-ordering $<$ of I and define w.r.t. this well-ordering

$$f\colon \bigcup_{i\in I} A_i \to \bigcup_{i\in I}(\{i\} \times A_i),$$
$$f(x) = (\min\{\, i \in I \mid x \in A_i \,\}, x).$$

Clearly f is injective. Hence

$$\begin{aligned} |\bigcup_{i\in I} A_i| &\leq |\bigcup_{i\in I}(\{i\} \times A_i)| \\ &\leq |\bigcup_{i\in I}(\{i\} \times \kappa)| \\ &= |I \times \kappa| \\ &= \max\{|I|, \kappa\}. \end{aligned}$$

(b). Because of (a) it suffices to show that $|I|, |A_i| \leq |\bigcup_{i\in I} A_i|$. The second estimate is clear. For the first one choose a well-ordering $<$ of $\bigcup_{i\in I} A_i$ and define $f\colon I \to \bigcup_{i\in I} A_i$ by $f(i) := \min_<\{\, x \mid x \in A_i \,\}$. By our assumption f is injective. □

A set a is *Dedekind-finite* if a cannot be mapped bijectively onto a proper subset b of a, otherwise *Dedekind-infinite*.

THEOREM (AC). *A set a is Dedekind-infinite iff a is infinite.*

PROOF. $\to$. Let $b \subsetneq a$ and $f\colon a \leftrightarrow b$. Assume $|a| < \omega$, say $|a| = n$. Then there is a $c \subsetneq n$ and some $g\colon n \leftrightarrow c$. We show by induction on n that this is impossible:

$$\forall_n \neg\exists_{c\subsetneq n}\exists_g(g\colon n \leftrightarrow c).$$

0. Clear. $n + 1$. Let $g\colon n + 1 \leftrightarrow c$ and $c \subsetneq n + 1$. We may assume $n \notin \mathrm{rng}(g{\restriction}n)$. It follows that $g{\restriction}n\colon n \leftrightarrow c\setminus\{n\} \subsetneq n$ and hence a contradiction to the induction hypothesis.

$\leftarrow$. Let $g\colon \omega \to a$ be injective and $h\colon g[\omega] \leftrightarrow g[\omega \setminus 1]$ defined by

$$h = \{\, (g(n), g(n + 1)) \mid n \in \omega \,\}.$$

Define $f\colon a \leftrightarrow (a \setminus \{g(0)\})$ by

$$f(x) = \begin{cases} x, & \text{if } x \in a \setminus g[\omega]; \\ h(x), & \text{otherwise.} \end{cases}$$

□

5.5.3. Regular and singular cardinals. Let κ, λ denote cardinals $\geq \omega$. In this section we shall always assume the Axiom of Choice (AC).

DEFINITION (AC). (a) $x \subseteq \kappa$ is *confinal* in κ if $\sup(x) = \kappa$.
(b) $\mathrm{cf}(\kappa) := \min\{\, |x| \mid x \subseteq \kappa \text{ and } x \text{ confinal with } \kappa \,\}$ is the *confinality* of κ.
(c) κ is *regular* if $\mathrm{cf}(\kappa) = \kappa$.
(d) κ is *singular* if $\mathrm{cf}(\kappa) < \kappa$.

THEOREM (AC). (a) $\omega = \aleph_0$ *is regular.*
(b) $\aleph_{\alpha+1}$ *is regular.*
(c) *If* β *is a limit and* $\beta < \aleph_\beta$*, then* $\aleph_\beta$ *is singular.*

PROOF. (a). Assume ω is singular, that is $\mathrm{cf}(\omega) < \omega$. Then there is an $x \subseteq \omega$ such that $|x| = n$ and $\sup(x) = \omega$. But this is impossible (proof by induction on n).

(b). Assume $\aleph_{\alpha+1}$ is singular. Then $\mathrm{cf}(\aleph_{a+1}) \leq \aleph_a$. Hence there is an $x \subseteq \aleph_{\alpha+1}$ such that $|x| \leq \aleph_\alpha$ and $\sup(x) = \aleph_{\alpha+1}$. But then

$$\begin{aligned} \aleph_{\alpha+1} &= |\bigcup x| \\ &\leq \max\{|x|, \sup\{\, |y| \mid y \in x \,\}\} \quad \text{(cardinality of unions)} \\ &\leq \aleph_\alpha, \end{aligned}$$

a contradiction.

(c). Let β be a limit such that $\beta < \aleph_\beta$. Then we have $\aleph_\beta = \sup\{\, \aleph_\gamma \mid \gamma < \beta \,\}$ and moreover $|\{\, \aleph_\gamma \mid \gamma < \beta \,\}| = |\beta| < \aleph_\beta$. Hence $\aleph_\beta$ is singular. □

By definition for every infinite cardinal κ there is a subset $x \subseteq \kappa$ whose cardinality equals $\mathrm{cf}(\kappa)$, hence which can be mapped bijectively onto $\mathrm{cf}(\kappa)$. We now show that one can even assume that this bijection is an isomorphism.

LEMMA (AC). *Let* κ *be an infinite cardinal. Then there exists a subset* $x \subseteq \kappa$ *confinal in* κ *that is isomorphic to* $\mathrm{cf}(\kappa)$.

PROOF. Let $y \subseteq \kappa$, $\sup(y) = \kappa$, $|y| = \mathrm{cf}(\kappa)$ and $g\colon \mathrm{cf}(\kappa) \leftrightarrow y$. By transfinite recursion we define

$$\begin{aligned} &\mathcal{F}\colon \mathrm{On} \to V, \\ &\mathcal{F}(\alpha) := \sup(\mathcal{F}[\alpha] \cup g[\alpha]) + 1. \end{aligned}$$

Let $f := \mathcal{F} \restriction \mathrm{cf}(\kappa)$. One can see easily
(a) $\alpha < \beta < \mathrm{cf}(\kappa) \to f(\alpha) < f(\beta) \wedge g(\alpha) < f(\beta)$.
(b) $\mathrm{rng}(f) \subseteq \kappa$.
(c) $\mathrm{rng}(f)$ is confinal with κ.
$\mathrm{rng}(f)$ is the x we are looking for. □

COROLLARY (AC). *If* κ *is an infinite cardinal, then* $\mathrm{cf}(\kappa)$ *is a regular cardinal.*

PROOF. $\mathrm{cf}(\mathrm{cf}(\kappa)) \leq \mathrm{cf}(\kappa)$ is clear. We must show $\mathrm{cf}(\kappa) \leq \mathrm{cf}(\mathrm{cf}(\kappa))$. By the lemma above there are x, f such that $x \subseteq \kappa$, $\sup(x) = \kappa$ and $f\colon \mathrm{cf}(\kappa) \leftrightarrow x$ isomorphism. Moreover there is $y \subseteq \mathrm{cf}(\kappa)$ such that $\sup(y) = \mathrm{cf}(\kappa)$ and $|y| = \mathrm{cf}(\mathrm{cf}(\kappa))$. One can see easily that $\{\, f(\alpha) \mid \alpha \in y \,\}$ is confinal with κ. Hence

$$\mathrm{cf}(\kappa) \leq |\{\, f(\alpha) \mid \alpha \in y \,\}| = |y| = \mathrm{cf}(\mathrm{cf}(\kappa)). \qquad \square$$

THEOREM (König). *Let κ be an infinite cardinal. Then $\kappa < |^{\mathrm{cf}(\kappa)}\kappa|$.*

PROOF. $\kappa = |^1\kappa| \leq |^{\mathrm{cf}(\kappa)}\kappa|$ is clear. Hence it suffices to derive a contradiction from the assumption that there is a bijection $f\colon \kappa \leftrightarrow {}^{\mathrm{cf}(\kappa)}\kappa$. According to the lemma there exists $x \subseteq \kappa$ such that $\sup(x) = \kappa$ and moreover an isomorphism $g\colon \mathrm{cf}(\kappa) \leftrightarrow x$. For every $\alpha < \mathrm{cf}(\kappa)$ we therefore have $g(\alpha) < \kappa$ and hence

$$|\{\, f(\gamma)(\alpha) \mid \gamma < g(\alpha) \,\}| \leq |g(\alpha)| < \kappa,$$

hence $\{\, f(\gamma)(\alpha) \mid \gamma < g(\alpha) \,\} \subsetneq \kappa$. Let

$$h\colon \mathrm{cf}(\kappa) \to \kappa,$$
$$h(\alpha) := \min(\kappa \setminus \{\, f(\gamma)(\alpha) \mid \gamma < g(\alpha) \,\}).$$

We obtain the desired contradiction by showing that $f(\gamma) \neq h$ for all $\gamma < \kappa$. Let $\gamma < \kappa$. Choose $\alpha < \mathrm{cf}(\kappa)$ such that $\gamma < g(\alpha)$. Then $h(\alpha) \neq f(\gamma)(\alpha)$ by construction of h. $\square$

5.5.4. Cardinal powers, continuum hypothesis. In this section we again assume (AC). We define

$$\aleph_\alpha^{\aleph_\beta} := |^{\aleph_\beta}\aleph_\alpha|.$$

Later we will introduce powers of ordinals as well. It should always be clear from the context whether we mean ordinal or cardinal power.

THEOREM (AC). (a) $\aleph_\beta < \mathrm{cf}(\aleph_\alpha) \to \aleph_\alpha \leq \aleph_\alpha^{\aleph_\beta} \leq |\mathcal{P}(\aleph_\alpha)|$
(b) $\mathrm{cf}(\aleph_\alpha) \leq \aleph_\beta \leq \aleph_\alpha \to \aleph_\alpha < \aleph_\alpha^{\aleph_\beta} \leq |\mathcal{P}(\aleph_\alpha)|$
(c) $\aleph_\alpha \leq \aleph_\beta \to \aleph_\alpha^{\aleph_\beta} = |\mathcal{P}(\aleph_\beta)|$.

PROOF. (a).

$$\begin{aligned}
\aleph_\alpha &\leq |^{\aleph_\beta}\aleph_\alpha| \\
&\leq |^{\aleph_\beta}({}^{\aleph_\alpha}\{0,1\})| \\
&= |^{\aleph_\beta \times \aleph_\alpha}\{0,1\}| \\
&= |^{\aleph_\alpha}\{0,1\}| \quad \text{because } \aleph_\beta \leq \aleph_\alpha \\
&= |\mathcal{P}(\aleph_\alpha)|.
\end{aligned}$$

(b).

$$\begin{aligned}\aleph_\alpha &< |{}^{\mathrm{cf}(\aleph_\alpha)}\aleph_\alpha| \quad \text{König's Theorem}\\ &\leq |{}^{\aleph_\beta}\aleph_\alpha|\\ &\leq |\mathcal{P}(\aleph_\alpha)| \quad \text{as in (a).}\end{aligned}$$

(c).

$$\begin{aligned}|\mathcal{P}(\aleph_\beta)| &= |{}^{\aleph_\beta}\{0,1\}|\\ &\leq |{}^{\aleph_\beta}\aleph_\alpha|\\ &\leq |{}^{\aleph_\beta\times\aleph_\alpha}\{0,1\}|\\ &= |{}^{\aleph_\beta}\{0,1\}|\\ &= |\mathcal{P}(\aleph_\beta)|. \qquad \square\end{aligned}$$

One can say much more about cardinal powers if one assumes the so-called *continuum hypothesis*:

$$|\mathcal{P}(\aleph_0)| = \aleph_1. \tag{CH}$$

An obvious generalization to all cardinals is the *generalized continuum hypothesis*:

$$|\mathcal{P}(\aleph_\alpha)| = \aleph_{\alpha+1}. \tag{GCH}$$

It is an open problem whether the continuum hypothesis holds in the cumulative type structure (No. 1 in Hilbert's list of mathematical problems, posed in a lecture at the international congress of mathematicians in Paris 1900). However, it is known that the continuum hypothesis is independent from the other axioms of set theory. We shall always indicate use of (CH) or (GCH).

THEOREM (GCH). (a) $\aleph_\beta < \mathrm{cf}(\aleph_\alpha) \to \aleph_\alpha = \aleph_\alpha^{\aleph_\beta}$.

(b) $\mathrm{cf}(\aleph_\alpha) \leq \aleph_\beta \leq \aleph_\alpha \to \aleph_\alpha^{\aleph_\beta} = \aleph_{\alpha+1}$.

(c) $\aleph_\alpha \leq \aleph_\beta \to \aleph_\alpha^{\aleph_\beta} = \aleph_{\beta+1}$.

PROOF. (b) and (c) follow with (GCH) from the previous theorem.

(a). Let $\aleph_\beta < \mathrm{cf}(\aleph_\alpha)$. First note that

$${}^{\aleph_\beta}\aleph_\alpha = \bigcup\{\, {}^{\aleph_\beta}\gamma \mid \gamma < \aleph_\alpha \,\}$$

This can be seen as follows. $\supseteq$ is clear. $\subseteq$. Let $f\colon \aleph_\beta \to \aleph_\alpha$. Because of $|f[\aleph_\beta]| \leq \aleph_\beta < \mathrm{cf}(\aleph_\alpha)$ we have $\sup(f[\aleph_\beta]) < \gamma < \aleph_\alpha$ for some γ, hence $f\colon \aleph_\beta \to \gamma$.

This gives

$\aleph_\alpha \leq |{}^{\aleph_\beta}\aleph_\alpha|$ previous theorem

$$= |\bigcup\{\, {}^{\aleph_\beta}\gamma \mid \gamma < \aleph_\alpha \,\}| \qquad \text{by the note above}$$
$$\leq \max\{|\aleph_\alpha|, \sup_{\gamma<\aleph_\alpha} |{}^{\aleph_\beta}\gamma| \,\} \quad \text{by the theorem on cardinality of unions.}$$

Hence it suffices to show that $|{}^{\aleph_\beta}\gamma| \leq \aleph_\alpha$ for $\gamma < \aleph_\alpha$. Let $\gamma < \aleph_\alpha$.

$$\begin{aligned}
|{}^{\aleph_\beta}\gamma| &\leq |{}^{\aleph_\beta \times \gamma}\{0,1\}| \\
&\leq |{}^{\aleph_\beta \times \aleph_\delta}\{0,1\}| \quad \text{for some } \delta \text{ with } |\gamma| \leq \aleph_\delta < \aleph_\alpha \\
&\leq \begin{cases} |\mathcal{P}(\aleph_\delta)| & \text{if } \beta < \delta \\ |\mathcal{P}(\aleph_\beta)| & \text{if } \delta \leq \beta \end{cases} \\
&= \begin{cases} \aleph_{\delta+1} & \text{if } \beta < \delta \\ \aleph_{\beta+1} & \text{if } \delta \leq \beta \end{cases} \\
&\leq \aleph_\alpha.
\end{aligned}$$

□

5.6. Ordinal Arithmetic

We define addition, multiplication and exponentiation for ordinals and prove their basic properties. We also treat Cantor's normal form.

5.6.1. Ordinal addition. Let

$$\begin{aligned}
&\alpha + 0 := \alpha, \\
&\alpha + (\beta + 1) := (\alpha + \beta) + 1, \\
&\alpha + \beta := \sup\{\, \alpha + \gamma \mid \gamma < \beta \,\} \quad \text{if } \beta \text{ limit.}
\end{aligned}$$

More precisely, define $s_\alpha \colon \mathrm{On} \to V$ by

$$\begin{aligned}
&s_\alpha(0) := \alpha, \\
&s_\alpha(\beta + 1) := s_\alpha(\beta) + 1, \\
&s_\alpha(\beta) := \bigcup \mathrm{rng}(s_\alpha \restriction \beta) \quad \text{if } \beta \text{ limit}
\end{aligned}$$

and then let $\alpha + \beta := s_\alpha(\beta)$.

Lemma (Properties of ordinal addition). (a) $\alpha + \beta \in \mathrm{On}$.
(b) $0 + \beta = \beta$.
(c) $\exists \alpha, \beta (\alpha + \beta \neq \beta + \alpha)$.
(d) $\beta < \gamma \to \alpha + \beta < \alpha + \gamma$.
(e) *There are* α, β, γ *such that* $\alpha < \beta$, *but* $\alpha + \gamma \not< \beta + \gamma$.
(f) $\alpha \leq \beta \to \alpha + \gamma \leq \beta + \gamma$.
(g) *For* $\alpha \leq \beta$ *there is a unique* γ *such that* $\alpha + \gamma = \beta$.
(h) *If* β *is a limit, then so is* $\alpha + \beta$.
(i) $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$.

PROOF. (a). Induction on β. *Case* 0. Clear. *Case* $\beta+1$. Then $\alpha+(\beta+1)=(\alpha+\beta)+1\in\mathrm{On}$, for by induction hypothesis $\alpha+\beta\in\mathrm{On}$. *Case* β limit. Then $\alpha+\beta=\sup\{\,\alpha+\gamma\mid\gamma<\beta\,\}\in\mathrm{On}$, for by induction hypothesis $\alpha+\gamma\in\mathrm{On}$ for all $\gamma<\beta$.

(b). Induction on β. *Case* 0. Clear. *Case* $\beta+1$. Then $0+(\beta+1)=(0+\beta)+1=\beta+1$, for by induction hypothesis $0+\beta=\beta$. *Case* β limit. Then

$$\begin{aligned}0+\beta&=\sup\{\,0+\gamma\mid\gamma<\beta\,\}\\&=\sup\{\,\gamma\mid\gamma<\beta\,\}&&\text{by induction hypothesis}\\&=\bigcup\beta\\&=\beta,&&\text{because }\beta\text{ is a limit.}\end{aligned}$$

(c). $1+\omega=\sup\{\,1+n\mid n\in\omega\,\}=\omega\neq\omega+1$.

(d). Induction on γ. *Case* 0. Clear. *Case* $\gamma+1$. Then

$$\begin{aligned}&\beta<\gamma+1,\\&\beta<\gamma\vee\beta=\gamma,\\&\alpha+\beta<\alpha+\gamma\vee\alpha+\beta=\alpha+\gamma&&\text{by induction hypothesis,}\\&\alpha+\beta\leq\alpha+\gamma<(\alpha+\gamma)+1=\alpha+(\gamma+1).\end{aligned}$$

Case γ limit. Let $\beta<\gamma$, hence $\beta<\delta$ for some $\delta<\gamma$. Then $\alpha+\beta<\alpha+\delta$ by induction hypothesis, hence $\alpha+\beta<\sup\{\,\alpha+\delta\mid\delta<\gamma\,\}=\alpha+\gamma$.

(e). $0<1$, but $0+\omega=\omega=1+\omega$

(f). We first remark that there can be no β such that $\alpha<\beta<\alpha+1$, for otherwise we would have in case $\beta\in\alpha$ the contradiction $\beta\in\alpha\in\beta$ and in case $\beta=\alpha$ the contradiction $\alpha\in\alpha$. As a second preliminary remark we note that

$$\alpha\leq\beta\to\alpha+1\leq\beta+1,$$

for in case $\beta+1<\alpha+1$ we would have $\alpha<\beta+1<\alpha+1$, which cannot be the case (as we have just seen). – We now show the claim $\alpha\leq\beta\to\alpha+\gamma\leq\beta+\gamma$ by induction on γ. *Case* 0. Clear. *Case* $\gamma+1$. Then

$$\begin{aligned}&\alpha+\gamma\leq\beta+\gamma&&\text{by induction hypothesis,}\\&(\alpha+\gamma)+1\leq(\beta+\gamma)+1&&\text{by the second preliminary remark,}\\&\alpha+(\gamma+1)\leq\beta+(\gamma+1)&&\text{by definition.}\end{aligned}$$

Case γ limit. Then

$$\begin{aligned}&\alpha+\delta\leq\beta+\delta\quad\text{for all }\delta<\gamma,\text{ by induction hypothesis,}\\&\alpha+\delta\leq\sup\{\,\beta+\delta\mid\delta<\gamma\,\}\\&\sup\{\,\alpha+\delta\mid\delta<\gamma\,\}\leq\sup\{\,\beta+\delta\mid\delta<\gamma\,\}\end{aligned}$$

$$\alpha + \gamma \leq \beta + \gamma \quad \text{by definition.}$$

(g). Uniqueness of γ follows from (d). Existence: Let $\alpha \leq \beta$. By (b) and (f) $\beta = 0 + \beta \leq \alpha + \beta$. Let γ be the least ordinal such that $\beta \leq \alpha + \gamma$. We show that $\beta = \alpha + \gamma$. *Case* $\gamma = 0$. Then $\beta \leq \alpha + \gamma = \alpha + 0 = \alpha \leq \beta$, hence $\beta = \alpha + \gamma$. *Case* $\gamma = \gamma' + 1$. Then $\alpha + \gamma' < \beta$, hence $(\alpha + \gamma') + 1 \leq \beta$ by the first preliminary remark for (f) and hence $\alpha + \gamma = \beta$. *Case* γ limit. Then $\alpha + \delta < \beta$ for all $\delta < \gamma$, hence $\alpha + \gamma = \sup\{\, \alpha + \delta \mid \delta < \gamma \,\} \leq \beta$ and hence $\alpha + \gamma = \beta$.

(h). Let β limit. We use the characterization of limits. $\alpha + \beta \neq 0$: Because of $0 \leq \alpha$ we have $0 < \beta = 0 + \beta \leq \alpha + \beta$ by (f). $\gamma < \alpha + \beta \to \gamma + 1 < \alpha + \beta$: Let $\gamma < \alpha + \beta = \sup\{\, \alpha + \delta \mid \delta < \beta \,\}$, hence $\gamma < \alpha + \delta$ for some $\delta < \beta$, hence $\gamma + 1 < \alpha + (\delta + 1)$ with $\delta + 1 < \beta$, hence $\gamma + 1 < \sup\{\, \alpha + \delta \mid \delta < \beta \,\}$.

(i). Induction on γ. *Case* 0. Clear. *Case* $\gamma + 1$. Then

$$\begin{aligned}(\alpha + \beta) + (\gamma + 1) &= [(\alpha + \beta) + \gamma] + 1 \\ &= [\alpha + (\beta + \gamma)] + 1 \quad \text{by induction hypothesis} \\ &= \alpha + [(\beta + \gamma) + 1] \\ &= \alpha + [\beta + (\gamma + 1)]\end{aligned}$$

Case γ limit. By (h) also $\beta + \gamma$ is a limit. Hence

$$\begin{aligned}(\alpha + \beta) + \gamma &= \sup\{\, (\alpha + \beta) + \delta \mid \delta < \gamma \,\} \\ &= \sup\{\, \alpha + (\beta + \delta) \mid \delta < \gamma \,\} \quad \text{by induction hypothesis} \\ &= \sup\{\, \alpha + \varepsilon \mid \varepsilon < \beta + \gamma \,\} \quad \text{see below} \\ &= \alpha + (\beta + \gamma).\end{aligned}$$

The equality of both suprema can be seen as follows. If $\varepsilon < \beta + \gamma$, then $\varepsilon < \beta + \delta$ for some $\delta < \gamma$ (by definition of $\beta + \gamma$) and hence $\alpha + \varepsilon < \alpha + (\beta + \delta)$. If conversely $\delta < \gamma$, then $\beta + \delta < \beta + \gamma$, hence $\alpha + (\beta + \delta) = \alpha + \varepsilon$ for some $\varepsilon < \beta + \gamma$ (take $\varepsilon := \beta + \delta$). □

5.6.2. Ordinal multiplication. Ordinal multiplication is defined by

$$\begin{aligned}&\alpha \cdot 0 := 0, \\ &\alpha \cdot (\beta + 1) := (\alpha \cdot \beta) + \alpha, \\ &\alpha \cdot \beta := \sup\{\, \alpha \cdot \gamma \mid \gamma < \beta \,\} \quad \text{if } \beta \text{ limit.}\end{aligned}$$

We write $\alpha\beta$ for $\alpha \cdot \beta$.

Lemma (Properties of ordinal multiplication). (a) $\alpha\beta \in \mathrm{On}$.
(b) $0\beta = 0$, $1\beta = \beta$.
(c) $\exists \alpha, \beta (\alpha\beta \neq \beta\alpha)$.
(d) $0 < \alpha \to \beta < \gamma \to \alpha\beta < \alpha\gamma$.

(e) *There are* α, β, γ *such that* $0 < \gamma$ *and* $\alpha < \beta$*, but* $\alpha\gamma \not< \beta\gamma$.
(f) $\alpha \leq \beta \to \alpha\gamma \leq \beta\gamma$.
(g) *If* $0 < \alpha$ *and* β *is a limit, then so is* $\alpha\beta$.
(h) $\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma$.
(i) *There are* α, β, γ *such that* $(\alpha + \beta)\gamma \neq \alpha\gamma + \beta\gamma$.
(j) $\alpha\beta = 0 \to \alpha = 0 \vee \beta = 0$.
(k) $(\alpha\beta)\gamma = \alpha(\beta\gamma)$.
(l) *If* $0 < \beta$*, then there are unique* γ, ρ *such that* $\alpha = \beta\gamma + \rho$ *and* $\rho < \beta$.

PROOF. (a). Induction on β. *Case* 0. Clear. *Case* $\beta + 1$. Then $\alpha(\beta + 1) = (\alpha\beta) + \alpha \in \mathrm{On}$, for by induction hypothesis $\alpha\beta \in \mathrm{On}$. *Case* β limit. Then $\alpha\beta = \sup\{\, \alpha\gamma \mid \gamma < \beta \,\} \in \mathrm{On}$, for by induction hypothesis $\alpha\gamma \in \mathrm{On}$ for all $\gamma < \beta$.

(b). $0\beta = 0$: Induction on β. *Case* 0. Clear. *Case* $\beta + 1$. Then $0(\beta + 1) = (0\beta) + 0 = 0$ by induction hypothesis. *Case* β limit. $0\beta = \sup\{\, 0\gamma \mid \gamma < \beta \,\} = 0$ by induction hypothesis. – $1\beta = \beta$: Induction on β. *Case* 0. Clear. *Case* $\beta + 1$. Then $1(\beta + 1) = (1\beta) + 1 = \beta + 1$ by induction hypothesis. *Case* β limit. $1\beta = \sup\{\, 1\gamma \mid \gamma < \beta \,\} = \sup\{\, \gamma \mid \gamma < \beta \,\} = \beta$ by induction hypothesis.

(c). Fist note that for all $n \in \omega$ we have $n\omega = \sup\{\, nm \mid m < \omega \,\} = \omega$. This implies $2\omega = \omega$, but $\omega 2 = \omega(1 + 1) = \omega 1 + \omega = \omega + \omega > \omega$.

(d). Let $0 < \alpha$. We show $\beta < \gamma \to \alpha\beta < \alpha\gamma$ by induction on γ. *Case* 0. Clear. *Case* $\gamma + 1$. Then

$$\begin{aligned} &\beta < \gamma + 1, \\ &\beta < \gamma \vee \beta = \gamma, \\ &\alpha\beta < \alpha\gamma \vee \alpha\beta = \alpha\gamma && \text{by induction hypothesis,} \\ &\alpha\beta \leq \alpha\gamma < (\alpha\gamma) + \alpha = \alpha(\gamma + 1). \end{aligned}$$

Case γ limit. Let $\beta < \gamma$, hence $\beta < \delta$ for some $\delta < \gamma$. Then $\alpha\beta < \alpha\delta$ by induction hypothesis, hence $\alpha\beta < \sup\{\, \alpha\delta \mid \delta < \gamma \,\} = \alpha\gamma$.

(e). We have $0 < \omega$ and $1 < 2$, but $1\omega = \omega = 2\omega$.

(f). We show the claim $\alpha \leq \beta \to \alpha\gamma \leq \beta\gamma$ by induction on γ. *Case* 0. Clear. *Case* $\gamma + 1$. Then

$$\begin{aligned} &\alpha\gamma \leq \beta\gamma && \text{by induction hypothesis,} \\ &(\alpha\gamma) + \alpha \leq (\beta\gamma) + \alpha \leq (\beta\gamma) + \beta && \text{by properties (f), (d) of ord. addition} \\ &\alpha(\gamma + 1) \leq \beta(\gamma + 1) && \text{by definition.} \end{aligned}$$

Case γ limit. Then

$$\begin{aligned} &\alpha\delta \leq \beta\delta \quad \text{for all } \delta < \gamma, \text{ by induction hypothesis,} \\ &\alpha\delta \leq \sup\{\, \beta\delta \mid \delta < \gamma \,\}, \end{aligned}$$

$$\sup\{\,\alpha\delta \mid \delta < \gamma\,\} \leq \sup\{\,\beta\delta \mid \delta < \gamma\,\},$$
$$\alpha\gamma \leq \beta\gamma \quad \text{by definition.}$$

(g). Let $0 < \alpha$ and β limit. For the proof of $\alpha\beta$ limit we again use the characterization of limits. $\alpha\beta \neq 0$: Because of $1 \leq \alpha$ and $\omega \leq \beta$ we have $0 < \omega = 1\omega \leq \alpha\beta$ by (f). $\gamma < \alpha\beta \to \gamma+1 < \alpha\beta$: Let $\gamma < \alpha\beta = \sup\{\,\alpha\delta \mid \delta < \beta\,\}$, hence $\gamma < \alpha\delta$ for some $\delta < \beta$, hence $\gamma + 1 < \alpha\delta + 1 \leq \alpha\delta + \alpha = \alpha(\delta + 1)$ with $\delta + 1 < \beta$, hence $\gamma + 1 < \sup\{\,\alpha\delta \mid \delta < \beta\,\}$.

(h). We must show $\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma$. We may assume let $0 < \alpha$. We employ induction on γ. *Case* 0. Clear. *Case* $\gamma + 1$. Then

$$\begin{aligned}
\alpha[\beta + (\gamma + 1)] &= \alpha[(\beta + \gamma) + 1] \\
&= \alpha(\beta + \gamma) + \alpha \\
&= (\alpha\beta + \alpha\gamma) + \alpha \quad \text{by induction hypothesis} \\
&= \alpha\beta + (\alpha\gamma + \alpha) \\
&= \alpha\beta + \alpha(\gamma + 1).
\end{aligned}$$

Case γ limit. By (g) $\alpha\gamma$ is a limit as well. We obtain

$$\begin{aligned}
\alpha(\beta + \gamma) &= \sup\{\,\alpha\delta \mid \delta < \beta + \gamma\,\} \\
&= \sup\{\,\alpha(\beta + \varepsilon) \mid \varepsilon < \gamma\,\} \\
&= \sup\{\,\alpha\beta + \alpha\varepsilon \mid \varepsilon < \gamma\,\} \quad \text{by induction hypothesis} \\
&= \sup\{\,\alpha\beta + \delta \mid \delta < \alpha\gamma\,\} \\
&= \alpha\beta + \alpha\gamma.
\end{aligned}$$

(i). $(1 + 1)\omega = 2\omega = \omega$, but $1\omega + 1\omega = \omega + \omega$.

(j). If $0 < \alpha, \beta$, hence $1 \leq \alpha, \beta$, then $0 < 1 \cdot 1 \leq \alpha\beta$.

(k). Induction on γ. We may assume $\beta \neq 0$. *Case* 0. Clear. *Case* $\gamma + 1$. Then

$$\begin{aligned}
(\alpha\beta)(\gamma + 1) &= (\alpha\beta)\gamma + \alpha\beta \\
&= \alpha(\beta\gamma) + \alpha\beta \quad \text{by induction hypothesis} \\
&= \alpha(\beta\gamma + \beta) \quad \text{by (h)} \\
&= \alpha[\beta(\gamma + 1)]
\end{aligned}$$

Case γ limit. By (g) $\beta\gamma$ is a limit as well. We obtain

$$\begin{aligned}
(\alpha\beta)\gamma &= \sup\{\,(\alpha\beta)\delta \mid \delta < \gamma\,\} \\
&= \sup\{\,\alpha(\beta\delta) \mid \delta < \gamma\,\} \quad \text{by induction hypothesis} \\
&= \sup\{\,\alpha\varepsilon \mid \varepsilon < \beta\gamma\,\} \\
&= \alpha(\beta\gamma).
\end{aligned}$$

(l). Existence: Let $0 < \beta$, hence $1 \leq \beta$ and hence $\alpha = 1\alpha \leq \beta\alpha$. Let γ be the least ordinal such that $\alpha \leq \beta\gamma$. *Case* $\alpha = \beta\gamma$. Let $\rho = 0$. *Case* $\alpha < \beta\gamma$. If $\gamma = \gamma' + 1$, then $\beta\gamma' < \alpha$. Hence there is a ρ such that $\beta\gamma' + \rho = \alpha$. Moreover, $\rho < \beta$, because from $\rho \geq \beta$ is follows that $\alpha = \beta\gamma' + \rho \geq \beta\gamma' + \beta = \beta(\gamma' + 1) = \beta\gamma$, contradicting our assumption. If γ is a limit, then $\alpha < \beta\gamma = \sup\{\, \beta\delta \mid \delta < \gamma \,\}$, hence $\alpha < \beta\delta$ for some $\delta < \gamma$, a contradiction.

Uniqueness: Assume $\beta\gamma_1 + \rho_1 = \beta\gamma_2 + \rho_2$ with $\rho_1, \rho_2 < \beta$. If say $\gamma_1 < \gamma_2$, then

$$\begin{aligned} \beta\gamma_1 + \rho_1 &< \beta\gamma_1 + \beta \\ &= \beta(\gamma_1 + 1) \\ &\leq \beta\gamma_2 \\ &\leq \beta\gamma_2 + \rho_2 \end{aligned}$$

hence we have a contradiction. Therefore $\gamma_1 = \gamma_2$, and hence $\rho_1 = \rho_2$. □

Corollary. *Every ordinal α can be written uniquely in the form $\alpha = \omega\gamma + n$. Here $n = 0$ iff $\alpha = 0$ or α is a limit.*

Proof. It remains to be shown that for every γ either $\omega\gamma = 0$ or $\omega\gamma$ is a limit. In case $\gamma = 0$ this is clear. In case $\gamma + 1$, the ordinal $\omega(\gamma + 1) = \omega\gamma + \omega$ is a limit by property (h) of ordinal addition. If γ is a limit, then so is $\omega\gamma$ (by property (g) of ordinal multiplication). □

5.6.3. Ordinal exponentiation. Ordinal exponentiation is defined by

$$\begin{aligned} \alpha^0 &:= \begin{cases} 0, & \text{if } \alpha = 0; \\ 1, & \text{otherwise,} \end{cases} \\ \alpha^{\beta+1} &:= \alpha^\beta \alpha, \\ \alpha^\beta &:= \sup\{\, \alpha^\gamma \mid \gamma < \beta \,\} \quad \text{if } \beta \text{ limit.} \end{aligned}$$

Lemma (Properties of ordinal exponentiation). (a) $\alpha^\beta \in \mathrm{On}$.
(b) $0^\beta = 0$, $1^\beta = 1$.
(c) $1 < \alpha \to \beta < \gamma \to \alpha^\beta < \alpha^\gamma$.
(d) *There are α, β, γ such that $1 < \gamma$ and $1 < \alpha < \beta$, but $\alpha^\gamma \not< \beta^\gamma$.*
(e) $\alpha \leq \beta \to \alpha^\gamma \leq \beta^\gamma$.
(f) *If $1 < \alpha$ and β is a limit, then so is α^β.*
(g) $\alpha^{\beta+\gamma} = \alpha^\beta \alpha^\gamma$.
(h) $\alpha^{\beta\gamma} = (\alpha^\beta)^\gamma$.
(i) $1 < \alpha \to \beta \leq \alpha^\beta$.

Proof. (a). Induction on β. *Case* 0. Clear. *Case* $\beta + 1$. Then $\alpha^{\beta+1} = (\alpha^\beta)\alpha \in \mathrm{On}$, for by induction hypothesis $\alpha^\beta \in \mathrm{On}$. *Case* β limit. Then

$\alpha^\beta = \sup\{\, \alpha^\gamma \mid \gamma < \beta \,\} \in \mathrm{On}$, for by induction hypothesis $\alpha^\gamma \in \mathrm{On}$ for all $\gamma < \beta$.

(b). $0^\beta = 0$: Induction on β. *Case* 0. $0^0 = 0$ holds by definition. *Case* $\beta + 1$. Then $0^{\beta+1} = (0^\beta)0 = 0$. *Case* β limit. $0^\beta = \sup\{\, 0^\gamma \mid \gamma < \beta \,\} = 0$ by induction hypothesis. – $1^\beta = 1$: Induction on β. *Case* 0. Clear. *Case* $\beta + 1$. Then $1^{\beta+1} = (1^\beta)1 = 1$ by induction hypothesis. *Case* β limit. $1^\beta = \sup\{\, 1^\gamma \mid \gamma < \beta \,\} = \sup\{\, 1 \mid \gamma < \beta \,\} = 1$ by induction hypothesis.

(c). Let $1 < \alpha$. We show $\beta < \gamma \to \alpha^\beta < \alpha^\gamma$ by induction on γ. *Case* 0. Clear. *Case* $\gamma + 1$. Then

$$\begin{aligned}
&\beta < \gamma + 1,\\
&\beta < \gamma \vee \beta = \gamma,\\
&\alpha^\beta < \alpha^\gamma \vee \alpha^\beta = \alpha^\gamma && \text{by induction hypothesis,}\\
&\alpha^\beta \leq \alpha^\gamma = \alpha^\gamma \cdot 1 < \alpha^\gamma \cdot \alpha = \alpha^{\gamma+1}.
\end{aligned}$$

Case γ limit. Let $\beta < \gamma$, hence $\beta < \delta$ for some $\delta < \gamma$. Then $\alpha^\beta < \alpha^\delta$ by induction hypothesis, hence $\alpha^\beta < \sup\{\, \alpha^\delta \mid \delta < \gamma \,\} = \alpha^\gamma$.

(d). For $1 < n$ we have $n^\omega = \sup\{\, n^m \mid m < \omega \,\} = \omega$ and hence $2^\omega = \omega = 3^\omega$.

(e). We show the claim $\alpha \leq \beta \to \alpha^\gamma \leq \beta^\gamma$ by induction on γ. *Case* 0. Clear. *Case* $\gamma + 1$. Let $\alpha \leq \beta$. Then

$$\begin{aligned}
\alpha^\gamma &\leq \beta^\gamma \quad \text{by induction hypothesis,}\\
\alpha^{\gamma+1} &= \alpha^\gamma \alpha\\
&\leq \beta^\gamma \alpha\\
&\leq \beta^\gamma \beta\\
&= \beta^{\gamma+1}.
\end{aligned}$$

Case γ limit. Let again $\alpha \leq \beta$. Then

$$\begin{aligned}
&\alpha^\delta \leq \beta^\delta \quad \text{for all } \delta < \gamma, \text{ by induction hypothesis,}\\
&\alpha^\delta \leq \sup\{\, \beta^\delta \mid \delta < \gamma \,\}\\
&\sup\{\, \alpha^\delta \mid \delta < \gamma \,\} \leq \sup\{\, \beta^\delta \mid \delta < \gamma \,\}\\
&\alpha^\gamma \leq \beta^\gamma \quad \text{by definition.}
\end{aligned}$$

(f). Let $1 < \alpha$ and β limit. For the proof of α^β limit we again use the characterization of limits. $\alpha^\beta \neq 0$: Because of $1 \leq \alpha$ we have $1 = 1^\beta \leq \alpha^\beta$. $\gamma < \alpha^\beta \to \gamma + 1 < \alpha^\beta$: Let $\gamma < \alpha^\beta = \sup\{\, \alpha^\delta \mid \delta < \beta \,\}$, hence $\gamma < \alpha^\delta$ for some $\delta < \beta$, hence $\gamma + 1 < \alpha^\delta + 1 \leq \alpha^\delta 2 \leq \alpha^{\delta+1}$ with $\delta + 1 < \beta$, hence $\gamma + 1 < \sup\{\, \alpha^\delta \mid \delta < \beta \,\}$.

(g). We must show $\alpha^{\beta+\gamma} = \alpha^\beta \alpha^\gamma$. We may assume $\alpha \neq 0, 1$. The proof is by induction on γ. *Case* 0. Clear. *Case* $\gamma + 1$. Then

$$\begin{aligned} \alpha^{\beta+\gamma+1} &= \alpha^\beta \alpha^\gamma \alpha \qquad \text{by induction hypothesis} \\ &= \alpha^\beta \alpha^{\gamma+1}. \end{aligned}$$

Case γ limit.

$$\begin{aligned} \alpha^{\beta+\gamma} &= \sup\{\, \alpha^\delta \mid \delta < \beta + \gamma \,\} \\ &= \sup\{\, \alpha^{\beta+\varepsilon} \mid \varepsilon < \gamma \,\} \\ &= \sup\{\, \alpha^\beta \alpha^\varepsilon \mid \varepsilon < \gamma \,\} \quad \text{by induction hypothesis} \\ &= \sup\{\, \alpha^\beta \delta \mid \delta < \alpha^\gamma \,\} \\ &= \alpha^\beta \alpha^\gamma. \end{aligned}$$

(h). We must show $\alpha^{\beta\gamma} = (\alpha^\beta)^\gamma$. We may assume $\alpha \neq 0, 1$ and $\beta \neq 0$. The proof is by induction on γ. *Case* 0. Clear. *Case* $\gamma + 1$. Then

$$\begin{aligned} \alpha^{\beta(\gamma+1)} &= \alpha^{\beta\gamma} \alpha^\beta \\ &= (\alpha^\beta)^\gamma \alpha^\beta \quad \text{by induction hypothesis} \\ &= (\alpha^\beta)^{\gamma+1}. \end{aligned}$$

Case γ limit. Because of $\alpha \neq 0, 1$ and $\beta \neq 0$ we know that $\alpha^{\beta\gamma}$ and $(\alpha^\beta)^\gamma$ are limits. Hence

$$\begin{aligned} \alpha^{\beta\gamma} &= \sup\{\, \alpha^\delta \mid \delta < \beta\gamma \,\} \\ &= \sup\{\, \alpha^{\beta\varepsilon} \mid \varepsilon < \gamma \,\} \\ &= \sup\{\, (\alpha^\beta)^\varepsilon \mid \varepsilon < \gamma \,\} \quad \text{by induction hypothesis} \\ &= (\alpha^\beta)^\gamma. \end{aligned}$$

(i). Let $1 < \alpha$. We show $\beta \leq \alpha^\beta$ by induction on β. *Case* 0. Clear. *Case* $\beta + 1$. Then $\beta \leq \alpha^\beta$ by induction hypothesis, hence

$$\begin{aligned} \beta + 1 &\leq \alpha^\beta + 1 \\ &\leq \alpha^\beta + \alpha^\beta \\ &\leq \alpha^{\beta+1}. \end{aligned}$$

Case β limit.

$$\begin{aligned} \beta &= \sup\{\, \gamma \mid \gamma < \beta \,\} \\ &\leq \sup\{\, \alpha^\gamma \mid \gamma < \beta \,\} \quad \text{by induction hypothesis} \\ &= \alpha^\beta. \end{aligned}$$

□

This concludes the proof.

5.6.4. Cantor normal form.

THEOREM (Cantor normal form). *Let $\gamma \geq 2$. Every α can be written uniquely in the form*

$$\alpha = \gamma^{\alpha_1}\beta_1 + \cdots + \gamma^{\alpha_n}\beta_n \quad \textit{where } \alpha \geq \alpha_1 > \cdots > \alpha_n \textit{ and } 0 < \beta_i < \gamma.$$

PROOF. Existence. Induction on α. Let δ be minimal such that $\alpha < \gamma^\delta$; such a δ exists since $\alpha \leq \gamma^\alpha$. But δ cannot be a limit, for otherwise $\alpha < \gamma^\varepsilon$ for some $\varepsilon < \delta$. If $\delta = 0$, then $\alpha = 0$ and the claim is trivial. Let $\delta = \alpha_1 + 1$, hence

$$\gamma^{\alpha_1} \leq \alpha < \gamma^{\alpha_1+1}.$$

Division with remainder gives

$$\alpha = \gamma^{\alpha_1}\beta_1 + \rho \quad \text{with } \rho < \gamma^{\alpha_1}.$$

Clearly $0 < \beta_1 < \gamma$. Now if $\rho = 0$ we are done. Otherwise we have

$$\rho = \gamma^{\alpha_2}\beta_2 + \cdots + \gamma^{\alpha_n}\beta_n \quad \text{by induction hypothesis.}$$

We still must show $\alpha_1 > \alpha_2$. But this holds, because $\alpha_2 \geq \alpha_1$ entails $\rho \geq \gamma^{\alpha_2} \geq \gamma^{\alpha_1}$, a contradiction.

Uniqueness. Let

$$\gamma^{\alpha_1}\beta_1 + \cdots + \gamma^{\alpha_n}\beta_n = \gamma^{\alpha'_1}\beta'_1 + \cdots + \gamma^{\alpha'_m}\beta'_m.$$

and assume that both representations are different. Since no such sum can extend the other, we must have $i \leq n, m$ such that $(\alpha_i, \beta_i) \neq (\alpha'_i, \beta'_i)$. By property (d) of ordinal addition we can assume $i = 1$. First we have

$$\begin{aligned}
&\gamma^{\alpha_1}\beta_1 + \cdots + \gamma^{\alpha_{n-1}}\beta_{n-1} + \gamma^{\alpha_n}\beta_n \\
&< \gamma^{\alpha_1}\beta_1 + \cdots + \gamma^{\alpha_{n-1}}\beta_{n-1} + \gamma^{\alpha_n+1} && \text{since } \beta_n < \gamma \\
&\leq \gamma^{\alpha_1}\beta_1 + \cdots + \gamma^{\alpha_{n-1}}(\beta_{n-1} + 1) && \text{for } \alpha_n < \alpha_{n-1} \\
&\leq \gamma^{\alpha_1}\beta_1 + \cdots + \gamma^{\alpha_{n-1}+1} \\
&\cdots \\
&\leq \gamma^{\alpha_1}(\beta_1 + 1).
\end{aligned}$$

Now if e.g., $\alpha_1 < \alpha'_1$, then we would have $\gamma^{\alpha_1}\beta_1 + \cdots + \gamma^{\alpha_n}\beta_n < \gamma^{\alpha_1}(\beta_1+1) \leq \gamma^{\alpha_1+1} \leq \gamma^{\alpha'_1}$, which cannot be. Hence $\alpha_1 = \alpha'_1$. If e.g., $\beta_1 < \beta'_1$, then we would have $\gamma^{\alpha_1}\beta_1 + \cdots + \gamma^{\alpha_n}\beta_n < \gamma^{\alpha_1}(\beta_1 + 1) \leq \gamma^{\alpha_1}\beta'_1$, which again cannot be the case. Hence $\beta_1 = \beta'_1$. □

COROLLARY (Cantor normal form with base ω). *Every α can be written uniquely in the form*

$$\alpha = \omega^{\alpha_1} + \cdots + \omega^{\alpha_n} \quad \textit{with } \alpha \geq \alpha_1 \geq \cdots \geq \alpha_n.$$

An ordinal α is an *additive principal number* when $\alpha \neq 0$ and $\beta + \gamma < \alpha$ for $\beta, \gamma < \alpha$.

COROLLARY. *Additive principal numbers are exactly the ordinals of the form* ω^ξ.

PROOF. This follows from Cantor's normal form with base ω. □

COROLLARY (Cantor normal form with base 2). *Every* α *can be written uniquely in the form*

$$\alpha = 2^{\alpha_1} + \cdots + 2^{\alpha_n} \quad \textit{with } \alpha \geq \alpha_1 > \cdots > \alpha_n.$$

Let $\omega_0 := 1$, $\omega_{k+1} := \omega^{\omega_k}$ and $\varepsilon_0 := \sup_{k<\omega} \omega_k$. Notice that ε_0 is the least ordinal α such that $\omega^\alpha = \alpha$.

5.7. Notes

Set theory as presented in these notes is commonly called ZFC (Zermelo-Fraenkel set theory with the axiom of choice; C for Choice). Zermelo wrote these axioms in 1908, with the exceptions of the regularity axioms (of von Neumann, 1925) and the replacement scheme (Fraenkel, 1922). Also Skolem considered principles related to these additional axioms. In ZFC the only objects are sets, classes are only a convenient way to speak of formulas.

CHAPTER 6

Proof Theory

This chapter presents an example of the type of proof theory inspired by Hilbert's programme and Gödel's incompleteness theorems. The principal goal will be to offer an example of a true mathematically meaningful principle not derivable in first-order arithmetic.

The main tool for proving theorems in arithmetic is clearly the induction schema

$$A(0) \to \forall_x(A(x) \to A(S(x))) \to \forall_x A(x).$$

Here $A(x)$ is an arbitrary formula. An equivalent form of this schema is "cumulative" or course-of-values induction

$$\forall_x(\forall_{y<x} A(y) \to A(x)) \to \forall_x A(x).$$

Both schemes refer to the standard ordering of the natural numbers. Now it is tempting to try to strengthen arithmetic by allowing more general induction schemas, e.g., with respect to the lexicographical ordering of $\mathbb{N}\times\mathbb{N}$. More generally, we might pick an arbitrary well-ordering $\prec$ over $\mathbb{N}$ and use the schema of *transfinite induction*:

$$\forall_x(\forall_{y\prec x} A(y) \to A(x)) \to \forall_x A(x).$$

This can be read as follows. Suppose the property $A(x)$ is "progressive", i.e., from the validity of $A(y)$ for all $y \prec x$ we can always conclude that $A(x)$ holds. Then $A(x)$ holds for all x.

One might wonder whether this schema of transfinite induction actually strengthens arithmetic. We will prove here a classic result of Gentzen (1943) which in a sense answers this question completely. However, in order to state the result we have to be more explicit about the well-orderings used. This is done in the next section.

6.1. Ordinals Below ε_0

In order to be able to speak in arithmetical theories about ordinals, we use a Gödelization of ordinals. This clearly is possible for countable ordinals only. Here we restrict ourselves to a countable set of relatively small ordinals, the ordinals below ε_0. Moreover, we equip these ordinals with an extra structure (a kind of algebra). It is then customary to speak

of *ordinal notations*. These ordinal notations could be introduced without any set theory in a purely formal, combinatorial way, based on the Cantor normal form for ordinals. However, we take advantage of the fact that we have just dealt with ordinals within set theory. We also introduce some elementary relations and operations for such ordinal notations, which will be used later. For brevity we from now on use the word "ordinal" instead of "ordinal notation".

6.1.1. Comparison of ordinals; natural sum.

LEMMA. *Let* $\omega^{\alpha_m}+\cdots+\omega^{\alpha_0}$ *and* $\omega^{\beta_n}+\cdots+\omega^{\beta_0}$ *be Cantor normal forms (with* $m,n\geq -1$*). Then*

$$\omega^{\alpha_m}+\cdots+\omega^{\alpha_0}<\omega^{\beta_n}+\cdots+\omega^{\beta_0}$$

iff there is an $i\geq 0$ *such that* $\alpha_{m-i}<\beta_{n-i}$, $\alpha_{m-i+1}=\beta_{n-i+1},\ldots,\alpha_m=\beta_n$, *or* $m<n$ *and* $\alpha_m=\beta_n,\ldots,\alpha_0=\beta_{n-m}$.

PROOF. Exercise. □

We use the notations 1 for ω^0, a for $\omega^0+\cdots+\omega^0$ with a copies of ω^0 and $\omega^\alpha a$ for $\omega^\alpha+\cdots+\omega^\alpha$ again with a copies of ω^α.

LEMMA. *Let* $\omega^{\alpha_m}+\cdots+\omega^{\alpha_0}$ *and* $\omega^{\beta_n}+\cdots+\omega^{\beta_0}$ *be Cantor normal forms. Then*

$$\omega^{\alpha_m}+\cdots+\omega^{\alpha_0}+\omega^{\beta_n}+\cdots+\omega^{\beta_0}=\omega^{\alpha_m}+\cdots+\omega^{\alpha_i}+\omega^{\beta_n}+\cdots+\omega^{\beta_0},$$

where i *is minimal such that* $\alpha_i\geq\beta_n$*; if there is no such* i*, let* $i=m+1$ *(so* $\omega^{\beta_n}+\cdots+\omega^{\beta_0}$*).*

PROOF. Exercise. □

One can also define a commutative variant of addition. This is the so-called *natural sum* or *Hessenberg sum* of two ordinals. For Cantor normal forms $\omega^{\alpha_m}+\cdots+\omega^{\alpha_0}$ and $\omega^{\beta_n}+\cdots+\omega^{\beta_0}$ it is defined by

$$(\omega^{\alpha_m}+\cdots+\omega^{\alpha_0})\#(\omega^{\beta_n}+\cdots+\omega^{\beta_0}):=\omega^{\gamma_{m+n+1}}+\cdots+\omega^{\gamma_0},$$

where $\gamma_{m+n+1},\ldots,\gamma_0$ is a decreasing permutation of $\alpha_m,\ldots,\alpha_0,\beta_n,\ldots,\beta_0$.

LEMMA. *#* *is associative, commutative and strongly monotonic in both arguments.*

PROOF. Exercise. □

6.1.2. Enumerating ordinals. In order to work with ordinals in a purely arithmetical system we set up some effective bijection between our ordinals $< \varepsilon_0$ and non-negative integers (i.e., a Gödel numbering). For its definition it is useful to refer to ordinals in the form

$$\omega^{\alpha_m} k_m + \cdots + \omega^{\alpha_0} k_0 \quad \text{with } \alpha_m > \cdots > \alpha_0 \text{ and } k_i \neq 0 \ (m \geq -1).$$

(By convention, $m = -1$ corresponds to the empty sum.)

For every ordinal α we define its Gödel number $\ulcorner \alpha \urcorner$ inductively by

$$\ulcorner \omega^{\alpha_m} k_m + \cdots + \omega^{\alpha_0} k_0 \urcorner := \Big(\prod_{i \leq m} p_{\ulcorner \alpha_i \urcorner}^{k_i} \Big) - 1,$$

where p_n is the n-th prime number starting with $p_0 := 2$. For every non-negative integer x we define its corresponding ordinal notation $\mathrm{o}(x)$ inductively by

$$\mathrm{o}\Big((\prod_{i \leq l} p_i^{q_i}) - 1 \Big) := \sum_{i \leq l} \omega^{\mathrm{o}(i)} q_i,$$

where the sum is to be understood as the natural sum.

LEMMA. (a) $\mathrm{o}(\ulcorner \alpha \urcorner) = \alpha$,
(b) $\ulcorner \mathrm{o}(x) \urcorner = x$.

PROOF. This can be proved easily by induction. □

Hence we have a simple bijection between ordinals and non-negative integers. Using this bijection we can transfer our relations and operations on ordinals to computable relations and operations on non-negative integers. We use the following abbreviations.

$$\begin{aligned} x \prec y &:= \mathrm{o}(x) < \mathrm{o}(y), \\ \omega^x &:= \ulcorner \omega^{\mathrm{o}(x)} \urcorner, \\ x \oplus y &:= \ulcorner \mathrm{o}(x) + \mathrm{o}(y) \urcorner, \\ xk &:= \ulcorner \mathrm{o}(x) k \urcorner, \\ \omega_k &:= \ulcorner \omega_k \urcorner, \end{aligned}$$

where $\omega_0 := 1$, $\omega_{k+1} := \omega^{\omega_k}$.

We leave it to the reader to verify that $\prec$, $\lambda_x \omega^x$, $\lambda_{x,y}(x \oplus y)$, $\lambda_{x,k}(xk)$ and $\lambda_k \ulcorner \omega_k \urcorner$ are all elementary.

6.2. Provability of Initial Cases of Transfinite Induction

We now derive initial cases of the principle of transfinite induction in arithmetic, i.e., of

$$\forall_x (\forall_{y \prec x} Py \to Px) \to \forall_{x \prec a} Px$$

for some number a and a predicate symbol P, where $\prec$ is the standard order of order type ε_0 defined in the preceding section. In a later section we will see that our results here are optimal in the sense that for the full system of ordinals $< \varepsilon_0$ the principle

$$\forall_x(\forall_{y\prec x}Py \to Px) \to \forall_x Px$$

of transfinite induction is underivable. All these results are due to Gentzen (1943).

6.2.1. Arithmetical systems. By an *arithmetical system* **Z** we mean a theory based on minimal logic in the $\forall{\to}\bot$-language (including equality axioms), with the following properties. The language of **Z** consists of a fixed (possibly countably infinite) supply of function and relation constants which are assumed to denote fixed functions and relations on the non-negative integers for which a computation procedure is known. Among the function constants there must be a constant S for the successor function and 0 for (the 0-place function) zero. Among the relation constants there must be a constant $=$ for equality and $\prec$ for the ordering of type ε_0 of the natural numbers, as introduced in section 6.1. In order to formulate the general principle of transfinite induction we also assume that a unary relation symbol P is present, which acts like a free set variable.

Terms are built up from object variables x, y, z by means of $f(t_1, \dots, t_m)$, where f is a function constant. We identify closed terms which have the same value; this is a convenient way to express in our formal systems the assumption that for each function constant a computation procedure is known. Terms of the form $\mathrm{S}(\mathrm{S}(\dots \mathrm{S}0 \dots))$ are called *numerals*. We use the notation S^n0 or $\overline{n}$ or (only in this chapter) even n for them. *Formulas* are built up from $\bot$ and atomic formulas $R(t_1, \dots, t_m)$, with R a relation constant or a relation symbol, by means of $A \to B$ and $\forall_x A$. Recall that we abbreviate $A \to \bot$ by $\neg A$.

The *axioms* of **Z** include the *Peano axioms*, i.e., the universal closures of

$$\mathrm{S}x = \mathrm{S}y \to x = y, \tag{6.1}$$

$$\mathrm{S}x = 0 \to A, \tag{6.2}$$

$$A(0) \to \forall_x(A(x) \to A(\mathrm{S}x)) \to \forall_x A(x), \tag{6.3}$$

with $A(x)$ an arbitrary formula. We express our assumption that for every relation constant R a decision procedure is known by adding the axiom $R\vec{n}$ whenever $R\vec{n}$ is true, and $\neg R\vec{n}$ whenever $R\vec{n}$ is false. Concerning $\prec$ we require irreflexivity and transitivity for $\prec$ as axioms, and also – following

Schütte – the universal closures of

$$x \prec 0 \to A, \tag{6.4}$$

$$z \prec y \oplus \omega^0 \to (z \prec y \to A) \to (z = y \to A) \to A, \tag{6.5}$$

$$x \oplus 0 = x, \tag{6.6}$$

$$x \oplus (y \oplus z) = (x \oplus y) \oplus z, \tag{6.7}$$

$$0 \oplus x = x, \tag{6.8}$$

$$\omega^x 0 = 0, \tag{6.9}$$

$$\omega^x(\mathrm{S}y) = \omega^x y \oplus \omega^x, \tag{6.10}$$

$$z \prec y \oplus \omega^{\mathrm{S}x} \to z \prec y \oplus \omega^{e(x,y,z)} m(x,y,z), \tag{6.11}$$

$$z \prec y \oplus \omega^{\mathrm{S}x} \to e(x,y,z) \prec \mathrm{S}x, \tag{6.12}$$

where $\oplus$, $\lambda_{x,y}(\omega^x y)$, e and m denote the appropriate function constants and A is any formula. (The reader should check that e, m can be taken to be elementary.) These axioms are formal counterparts to the properties of the ordinal notations observed in the preceding section. We also allow an arbitrary supply of true formulas $\forall_{\vec{x}} A$ with A quantifier-free and without P as axioms. Such formulas are called Π_1-*formulas* (in the literature also Π^0_1-formulas).

Moreover, we may also add an *ex-falso-quodlibet schema* Efq or even a *stability schema* Stab for A:

$$\bot \to A,$$
$$\neg\neg A \to A.$$

Addition of Efq leads to an intuitionistic arithmetical system (the $\forall{\to}\bot$-fragment of of Heyting arithmetic HA), and addition of Stab to a classical arithmetical system (a version of Peano arithmetic PA). Note that in our $\forall{\to}\bot$-fragment of minimal logic these schemas are derivable from their instances

$$\bot \to R\vec{x},$$
$$\neg\neg R\vec{x} \to R\vec{x},$$

with R a relation constant or the special relation symbol P. Note also that when the stability schema is present, we can replace (6.2), (6.4) and (6.5) by their more familiar classical versions

$$\mathrm{S}x \neq 0, \tag{6.13}$$

$$x \not\prec 0, \tag{6.14}$$

$$z \prec y \oplus \omega^0 \to z \neq y \to z \prec y. \tag{6.15}$$

We will also consider *restricted* arithmetical systems $\mathbf{Z}_k$. They are defined like $\mathbf{Z}$, but with the induction schema (6.3) restricted to formulas A of level $\mathrm{lev}(A) \leq k$. The *level* of a formula A is defined by

$$\begin{aligned}\mathrm{lev}(R\vec{t}\,) &:= \mathrm{lev}(\bot) := 0,\\ \mathrm{lev}(A \to B) &:= \max(\mathrm{lev}(A)+1, \mathrm{lev}(B)),\\ \mathrm{lev}(\forall_x A) &:= \max(1, \mathrm{lev}(A)).\end{aligned}$$

However, the trivial special case of induction $A(0) \to \forall_x A(\mathrm{S}x) \to \forall_x A$, which amounts to case distinction, is allowed for arbitrary A. (This is needed in the proof of the theorem below).

6.2.2. Gentzen's Proof.

THEOREM (Provable initial cases of transfinite induction in $\mathbf{Z}$). *Transfinite induction up to ω_n, i.e., for arbitrary $A(x)$ the formula*

$$\forall_x(\forall_{y\prec x} A(y) \to A(x)) \to \forall_{x\prec\omega_n} A(x),$$

is derivable in $\mathbf{Z}$.

PROOF. To every formula $A(x)$ we assign a formula $A^+(x)$ (with respect to a fixed variable x) by

$$A^+(x) := \forall_y(\forall_{z\prec y} A(z) \to \forall_{z\prec y\oplus\omega^x} A(z)).$$

We first show

$$\text{If } A(x) \text{ is progressive, then } A^+(x) \text{ is progressive,}$$

where "$B(x)$ is *progressive*" means $\forall_x(\forall_{y\prec x} B(y) \to B(x))$. So assume that $A(x)$ is progressive and

$$\forall_{y\prec x} A^+(y). \tag{6.16}$$

We have to show $A^+(x)$. So assume further

$$\forall_{z\prec y} A(z) \tag{6.17}$$

and $z \prec y \oplus \omega^x$. We have to show $A(z)$.

Case $x = 0$. Then $z \prec y \oplus \omega^0$. By (6.5) it suffices to derive $A(z)$ from $z \prec y$ as well as from $z = y$. If $z \prec y$, then $A(z)$ follows from (6.17), and if $z = y$, then $A(z)$ follows from (6.17) and the progressiveness of $A(x)$.

Case $\mathrm{S}x$. From $z \prec y \oplus \omega^{\mathrm{S}x}$ we obtain $z \prec y \oplus \omega^{e(x,y,z)} m(x,y,z)$ by (6.11) and $e(x,y,z) \prec \mathrm{S}x$ by (6.12). From (6.16) we obtain $A^+(e(x,y,z))$. By the definition of $A^+(x)$ we get

$$\forall_{u\prec y\oplus\omega^{e(x,y,z)}v} A(u) \to \forall_{u\prec(y\oplus\omega^{e(x,y,z)}v)\oplus\omega^{e(x,y,z)}} A(u)$$

and hence, using (6.7) and (6.10)

$$\forall_{u\prec y\oplus\omega^{e(x,y,z)}v} A(u) \to \forall_{u\prec y\oplus\omega^{e(x,y,z)}(\mathrm{S}v)} A(u).$$

Also from (6.17) and (6.9), (6.6) we obtain

$$\forall_{u \prec y \oplus \omega^{e(x,y,z)}0} A(u).$$

Using an appropriate instance of the induction schema we can conclude

$$\forall_{u \prec y \oplus \omega^{e(x,y,z)} m(x,y,z)} A(u)$$

and hence $A(z)$.

We now show, by induction on n, how for an arbitrary formula $A(x)$ we can obtain a derivation of

$$\forall_x(\forall_{y \prec x} A(y) \to A(x)) \to \forall_{x \prec \omega_n} A(x).$$

So assume the left hand side, i.e., assume that $A(x)$ is progressive.

Case 0. Then $x \prec \omega^0$ and hence $x \prec 0 \oplus \omega^0$ by (6.8). By (6.5) it suffices to derive $A(x)$ from $x \prec 0$ as well as from $x = 0$. Now $x \prec 0 \to A(x)$ holds by (6.4), and $A(0)$ then follows from the progressiveness of $A(x)$.

Case $n+1$. Since $A(x)$ is progressive, by what we have shown above $A^+(x)$ is also progressive. Applying the induction hypothesis to $A^+(x)$ yields $\forall_{x \prec \omega_n} A^+(x)$, and hence $A^+(\omega_n)$ by the progressiveness of $A^+(x)$. Now the definition of $A^+(x)$ (together with (6.4) and (6.8)) yields $\forall_{z \prec \omega^{\omega_n}} A(z)$. □

Note that in the induction step of this proof we have derived transfinite induction up to ω_{n+1} for $A(x)$ from transfinite induction up to ω_n for a formula of level higher than the level of $A(x)$.

We now want to refine the preceding theorem to a corresponding result for the subsystems $\mathbf{Z}_k$ of $\mathbf{Z}$.

THEOREM (Provable initial cases of transfinite induction in $\mathbf{Z}_k$). *Let $1 \le l \le k$. Then in $\mathbf{Z}_k$ we can derive transfinite induction for any formula $A(x)$ of level $\le l$ up to $\omega_{k-l+2}[m]$ for arbitrary m, i.e.*

$$\forall_x(\forall_{y \prec x} A(y) \to A(x)) \to \forall_{x \prec \omega_{k-l+2}[m]} A(x),$$

where $\omega_1[m] := m$, $\omega_{i+1}[m] := \omega^{\omega_i[m]}$.

PROOF. Note first that if $A(x)$ is a formula of level $l \ge 1$, then the formula $A^+(x)$ constructed in the proof of the preceding theorem has level $l+1$, and for the proof of

$$\text{If } A(x) \text{ is progressive, then } A^+(x) \text{ is progressive,}$$

we have used induction with an induction formula of level l.

Now let $A(x)$ be a fixed formula of level $\le l$, and assume that $A(x)$ is progressive. Define $A^0 := A$, $A^{i+1} := (A^i)^+$. Then $\mathrm{lev}(A^i) \le l+i$, and hence in $\mathbf{Z}_k$ we can derive that $A^1, A^2, \ldots A^{k-l+1}$ are all progressive. Now from

the progressiveness of $A^{k-l+1}(x)$ we obtain $A^{k-l+1}(0)$, $A^{k-l+1}(1)$, $A^{k-l+1}(2)$ and generally $A^{k-l+1}(m)$ for any m, i.e., $A^{k-l+1}(\omega_1[m])$. But since

$$A^{k-l+1}(x) = (A^{k-l})^+(x) = \forall_y(\forall_{z\prec y} A^{k-l}(z) \to \forall_{z\prec y\oplus\omega^x} A^{k-l}(z))$$

we first get (with $y = 0$) $\forall_{z\prec\omega_2[m]} A^{k-l}(z)$ and then $A^{k-l}(\omega_2[m])$ by the progressiveness of A^{k-l}. Repeating this argument we finally obtain

$$\forall_{z\prec\omega_{k-l+2}[m]} A^0(z). \qquad \square$$

Our next aim is to prove that these bounds are sharp. More precisely, we will show that in $\mathbf{Z}$ (no matter how many true Π_1-formulas we have added as axioms) one cannot derive "purely schematic" transfinite induction up to ε_0, i.e., one cannot derive the formula

$$\forall_x(\forall_{y\prec x} Py \to Px) \to \forall_x Px$$

with a relation symbol P, and that in $\mathbf{Z}_k$ one cannot derive transfinite induction up to ω_{k+1}, i.e., the formula

$$\forall_x(\forall_{y\prec x} Py \to Px) \to \forall_{x\prec\omega_{k+1}} Px.$$

This will follow from the method of normalization applied to arithmetical systems, which we have to develop first.

6.3. Normalization with the Omega Rule

We will show below that a normalization theorem does not hold for arithmetical systems $\mathbf{Z}$, in the sense that for any formula A derivable in $\mathbf{Z}$ there is a derivation of the same formula A in $\mathbf{Z}$ which only uses formulas of a level bounded by the level of A. The reason for this failure is the presence of induction axioms, which can be of arbitrary level.

Here we remove that obstacle against normalization in a somewhat drastic way: we leave the realm of proofs as finite combinatory objects and replace the induction axiom by a rule with infinitely many premises, the so-called ω-rule (suggested by Hilbert and studied by Lorenzen, Novikov and Schütte), which allows us to conclude $\forall_x A(x)$ from $A(0), A(1), A(2), \ldots$, i.e.

$$\frac{\begin{array}{cccccc} d_0 & d_1 & & d_i & \\ A(0) & A(1) & \ldots & A(i) & \ldots \end{array}}{\forall_x A(x)}\,\omega$$

So derivations can be viewed as labelled infinite (countably branching) trees. As in the finitary case a label consists of the derived formula and the name of the rule applied. Since we define derivations inductively, any such derivation tree must be well-founded, i.e., must not contain an infinite descending path.

Clearly this ω-rule can also be used to replace the rule $\forall^+ x$. As a consequence we do not need to consider free individual variables.

It is plain that every derivation in an arithmetical system **Z** can be translated into an infinitary derivation with the ω-rule; this will be carried out in 6.3.3 below. The resulting infinitary derivation has a noteworthy property: in any application of the ω-rule the cutranks of the infinitely many immediate subderivations d_n are bounded, and also their sets of free assumption variables are bounded by a finite set. Here the cutrank of a derivation is as usual the least number $\geq$ the level of any subderivation obtained by $\rightarrow^+$ as the main premise of $\rightarrow^-$ or by the ω-rule as the main premise of $\forall^-$, where the *level of a derivation* is the level of its type as a term, i.e., of the formula it derives. Clearly a derivation is called normal iff its cutrank is zero, and we will prove below that any (possibly infinite) derivation of finite cutrank can be transformed into a derivation of cutrank zero. The resulting normal derivation will continue to be infinite, so the result may seem useless at first sight. However, we will be able to bound the depth of the resulting derivation in an informative way, and this will enable us in Section 6.4 to obtain the desired results on unprovable initial cases of transfinite induction. Let us now carry out this programme.

N.B. The standard definition of cutrank in predicate logic measures the depth of formulas; here one uses the level.

6.3.1. Infinitary derivations. The systems $\mathbf{Z}^\infty$ of ω-arithmetic are defined as follows. $\mathbf{Z}^\infty$ has the same language and – apart from the induction axioms – the same axioms as **Z**. Derivations in $\mathbf{Z}^\infty$ are infinite objects. It is useful to employ a term notation for these, and we temporarily use d, e, f to denote such (infinitary) derivation terms. For the term corresponding to the deduction obtained by applying the ω-rule to d_i, $i \in \mathbb{N}$ we write $\langle d_i \rangle_{i<\omega}$. However, for our purposes here it suffices to only consider derivations whose depth is bounded below ε_0.

We define the notion "d is a *derivation of depth* $\leq \alpha$" (written $|d| \leq \alpha$) inductively as follows (i ranges over numerals).

(A) Any assumption variable u^A with A a closed formula and any axiom Ax^A is a derivation of depth $\leq \alpha$, for any α.

($\rightarrow^+$) If d^B is a derivation of depth $\leq \alpha_0 < \alpha$, then $(\lambda_{u^A} d^B)^{A\rightarrow B}$ is a derivation of depth $\leq \alpha$.

($\rightarrow^-$) If $d^{A\rightarrow B}$ and e^A are derivations of depths $\leq \alpha_i < \alpha$ (i=1,2), then $(d^{A\rightarrow B} e^A)^B$ is a derivation of depth $\leq \alpha$.

(ω) For all $A(x)$, if $d_i^{A(i)}$ are derivations of depths $\leq \alpha_i < \alpha$ ($i < \omega$), then $(\langle d_i^{A(i)} \rangle_{i<\omega})^{\forall_x A}$ is a derivation of depth $\leq \alpha$.

($\forall^-$) For all $A(x)$, if $d^{\forall_x A(x)}$ is a derivation of depth $\leq \alpha_0 < \alpha$, then, for all i, $(d^{\forall x A(x)} i)^{A(i)}$ is a derivation of depth $\leq \alpha$.

We will use $|d|$ to denote the least α such that $|d| \leq \alpha$.

Note that in $(\forall^-)$ it suffices to use numerals as minor premises. The reason is that we only need to consider closed terms, and any such term is in our setup identified with a numeral.

The *cutrank* $\mathrm{cr}(d)$ of a derivation d is defined by

$$\begin{aligned}
\mathrm{cr}(u^A) &:= \mathrm{cr}(\mathrm{Ax}^A) := 0,\\
\mathrm{cr}(\lambda_u d) &:= \mathrm{cr}(d),\\
\mathrm{cr}(d^{A\to B}e^A) &:= \begin{cases} \max(\mathrm{lev}(A\to B), \mathrm{cr}(d), \mathrm{cr}(e)), & \text{if } d = \lambda_u d',\\ \max(\mathrm{cr}(d), \mathrm{cr}(e)), & \text{otherwise,}\end{cases}\\
\mathrm{cr}(\langle d_i\rangle_{i<\omega}) &:= \sup_{i<\omega} \mathrm{cr}(d_i),\\
\mathrm{cr}(d^{\forall_x A(x)} j) &:= \begin{cases} \max(\mathrm{lev}(\forall_x A(x)), \mathrm{cr}(d)), & \text{if } d = \langle d_i\rangle_{i<\omega},\\ \mathrm{cr}(d), & \text{otherwise.}\end{cases}
\end{aligned}$$

Clearly $\mathrm{cr}(d) \in \mathbb{N} \cup \{\omega\}$ for all d. For our purposes it will suffice to consider only derivations with finite cutranks (i.e., with $\mathrm{cr}(d) \in \mathbb{N}$) and with finitely many free assumption variables.

Lemma (Substitution). *If d is a derivation of depth $\le \alpha$, with free assumption variables among $u, \vec{u}$ and of cutrank $\mathrm{cr}(d) = k$, and e is a derivation of depth $\le \beta$, with free assumption variables among $\vec{u}$ and of cutrank $\mathrm{cr}(e) = l$, then $d[u := e]$ is a derivation with free assumption variables among $\vec{u}$, of depth $|d[u := e]| \le \beta + \alpha$ and of cutrank $\mathrm{cr}(d[u := e]) \le \max(\mathrm{lev}(e), k, l)$.*

Proof. Straightforward induction on the depth of d. □

Using this lemma we can now embed our systems $\mathbf{Z}_k$ (i.e., arithmetic with induction restricted to formulas of level $\le k$) and hence $\mathbf{Z}$ into $\mathbf{Z}^\infty$. In this embedding we refer to the number $n_I(d)$ of nested applications of the induction schema within a $\mathbf{Z}_k$-derivation d.

The *nesting* of applications of induction in d, $n_I(d)$, is defined by induction on d, as follows.

$$\begin{aligned}
n_I(u) &:= n_I(\mathrm{Ax}) := 0,\\
n_I(\mathrm{Ind}) &:= 1,\\
n_I(\mathrm{Ind}\,\vec{t}de) &:= \max(n_I(d), n_I(e) + 1),\\
n_I(de) &:= \max(n_I(d), n_I(e)), \quad \text{if } d \text{ is not of the form } \mathrm{Ind}\,\vec{t}d_0,\\
n_I(\lambda_u d) &:= n_I(\lambda x d) := n_I(dt) := n_I(d).
\end{aligned}$$

6.3.2. Long normal form. For the next lemma we need the notion of the *long normal form* of a derivation. In 1.2.6 we have studied the form of normal derivations in minimal logic. We considered the notion of a *track* and observed, that in every track all elimination rules precede all introduction rules, and that in a uniquely determined *minimal node* we encounter a *minimal formula*, that is a subformula of any formula in the elimination part as well as in the introduction part of the track. In the notion of a long normal form we additionally require that every minimal formula is atomic.

For simplicity we restrict ourselves to the $\to$-fragment of minimal propositional logic; however, our considerations are valid for the full language as well.

For terms of the typed λ-calculus we define the *η-expansion of a variable* by

$$\eta_V(x^{\vec{\tau}\to\iota}) := \lambda_{\vec{z}^{\vec{\tau}}}(x\eta_V(\vec{z})),$$

so by induction on the type of the variable. The *η-expansion of a term* in normal form can then be defined by induction on terms:

$$\eta\big(\lambda_{\vec{y}}(x\vec{M})^{\vec{\tau}\to\iota}\big) := \lambda_{\vec{y},\vec{z}^{\vec{\tau}}}(x\eta(\vec{M})\eta_V(\vec{z})).$$

Note that we always have $\eta(x) = \eta_V(x)$. – Hence clearly:

Lemma. *Every term can be transformed into long normal form, by first normalizing and then η-expanding it.*

6.3.3. Embedding of $\mathbf{Z}_k$.

Lemma. *Let a $\mathbf{Z}_k$-derivation in long normal form be given with $\leq m$ nested applications of the induction schema, i.e., of*

$$A(0) \to \forall_x(A(x) \to A(\mathrm{S}x)) \to \forall_x A(x),$$

all with $\mathrm{lev}(A) \leq k$. *We consider subderivations d^B not of the form* $\mathrm{Ind}\,\vec{t}$ *or* $\mathrm{Ind}\,\vec{t}d_0$. *For every such subderivation and closed substitution instance $B\sigma$ of B we construct $(d^\infty_\sigma)^{B\sigma}$ in $\mathbf{Z}^\infty$ with free assumption variables $u^{C\sigma}$ for u^C free assumption of d, such that $|d^\infty_\sigma| < \omega^{m+1}$ and $\mathrm{cr}(d^\infty_\sigma) \leq k$, and moreover such that d is obtained by $\to^+$ iff d^∞_σ is, and d is obtained by $\forall^+$ or of the form* $\mathrm{Ind}\,\vec{t}d_0e$ *iff d^∞_σ is obtained by the ω-rule.*

Proof. By recursion on such subderivations d.

Case u^C or Ax. Take $u^{C\sigma}$ or Ax.

Case $\mathrm{Ind}\,\vec{t}de'$. Since the deduction is in long normal form, $e' = \lambda_{x,v}e$. By induction hypothesis we have d^∞_σ and e^∞_σ. (Note that neither d nor e can have one of the forbidden forms $\mathrm{Ind}\,\vec{t}$ and $\mathrm{Ind}\,\vec{t}d_0$, since both are in long normal form). Write $e^\infty_\sigma(t,f)$ for $e^\infty_\sigma[x,v := t,f]$, and let

$$(\mathrm{Ind}\,\vec{t}d(\lambda_{x,v}e))^\infty_\sigma := \langle d^\infty_\sigma, e^\infty_\sigma(0,d^\infty_\sigma), e^\infty_\sigma(1,e^\infty_\sigma(0,d^\infty_\sigma)),\dots\rangle.$$

By induction hypothesis $|e^\infty_\sigma| \le \omega^{m-1}\cdot p$ and $|d^\infty_\sigma| \le \omega^m\cdot q$ for some $p, q < \omega$. By the substitution lemma in 6.3.1 we obtain

$$|e^\infty_\sigma(0, d^\infty_\sigma)| \le \omega^m\cdot q + \omega^{m-1}\cdot p,$$
$$|e^\infty_\sigma(1, e^\infty_\sigma(0, d^\infty_\sigma))| \le \omega^m\cdot q + \omega^{m-1}\cdot 2p$$

and so on, and hence

$$|(\mathrm{Ind}\, d(\lambda_{x,v}e))^\infty_\sigma| \le \omega^m\cdot(q+1).$$

Concerning the cutrank we have by induction hypothesis $\mathrm{cr}(d^\infty_\sigma), \mathrm{cr}(e^\infty_\sigma) \le k$. Therefore

$$\mathrm{cr}(e^\infty_\sigma(0, d^\infty_\sigma)) \le \max(\mathrm{lev}(A(0)), \mathrm{cr}(d^\infty_\sigma), \mathrm{cr}(e^\infty_\sigma)) \le k,$$
$$\mathrm{cr}(e^\infty_\sigma(1, e^\infty_\sigma(0, d^\infty_\sigma))) \le \max(\mathrm{lev}(A(1)), k, \mathrm{cr}(e^\infty_\sigma)) = k,$$

and so on, and hence

$$\mathrm{cr}((\mathrm{Ind}\, d(\lambda_{x,v}e))^\infty_\sigma) \le k.$$

Case $\lambda_{u^C} d^B$. By induction hypothesis, we have $(d^\infty_\sigma)^{B\sigma}$ with possibly free assumptions $u^{C\sigma}$. Take $(\lambda_u d)^\infty_\sigma := \lambda_{u^{C\sigma}} d^\infty_\sigma$.

Case de, with d not of the form $\mathrm{Ind}\,\vec{t}$ or $\mathrm{Ind}\,\vec{t}d_0$. By induction hypothesis we have d^∞_σ and e^∞_σ. Since de is subderivation of a normal derivation we know that d and hence also d^∞_σ is not obtained by $\to^+$. Therefore $(de)^\infty_\sigma := d^\infty_\sigma e^\infty_\sigma$ is normal and $\mathrm{cr}(d^\infty_\sigma e^\infty_\sigma) = \max(\mathrm{cr}(d^\infty_\sigma), \mathrm{cr}(e^\infty_\sigma)) \le k$. Also we clearly have $|d^\infty_\sigma e^\infty_\sigma| < \omega^{m+1}$.

Case $(\lambda_x d)^{\forall_x B(x)}$. By induction hypothesis for every i and substitution instance $B(i)\sigma$ we have $d^\infty_{\sigma,i}$. Take $(\lambda_x d)^\infty_\sigma := \langle d^\infty_{\sigma,i}\rangle_{i<\omega}$.

Case $(dt)^{B(t)}$. By induction hypothesis, we have $(d^\infty_\sigma)^{(\forall_x B(x))\sigma}$. Let j be the numeral with the same value as $t\sigma$. If $d^\infty_\sigma = \langle d_i\rangle_{i<\omega}$ (which can only be the case if $d = \mathrm{Ind}\,\vec{t}d_0e_0$, for dt is a subderivation of a normal derivation), take $(dt)^\infty_\sigma := d_j$. Otherwise take $(dt)^\infty_\sigma := d^\infty_\sigma j$ □

6.3.4. Normalization for $\mathbf{Z}^\infty$. A derivation is called *convertible* or a *redex* if it is of the form $(\lambda_u d(u))e$ or else $\langle d_i\rangle_{i<\omega} j$, which can be converted into $d(e)$ or d_j, respectively. A derivation is called *normal* if it does not contain a convertible subderivation. Note that a derivation is normal iff it is of cutrank 0.

Call a derivation a *simple application* if it is of the form $d_0 d_1 \dots d_m$ with d_0 an assumption variable or an axiom.

We want to define an operation which by repeated conversions transforms a given derivation into a normal one with the same end formula and no additional free assumption variables. The usual methods to achieve such a task have to be adapted properly in order to deal with the new situation

of infinitary derivations. Here we give a particularly simple argument due to Tait (1965).

LEMMA. *For any derivation d^A of depth $\leq \alpha$ and cutrank $k+1$ we can find a derivation $(d^k)^A$ with free assumption variables contained in those of d, which has depth $\leq 2^\alpha$ and cutrank $\leq k$.*

PROOF. By induction on α. The only case which requires some argument is when the derivation is of the form de with $|d| \leq \alpha_1 < \alpha$ and $|e| \leq \alpha_2 < \alpha$, but is not a simple application. We first consider the subcase where $d^k = \lambda_u d_1(u)$ and $\mathrm{lev}(d) = k+1$. Then $\mathrm{lev}(e) \leq k$ by the definition of level (recall that the level of a derivation was defined to be the level of the formula it derives), and hence $d_1(e^k)$ has cutrank $\leq k$ by the substitution lemma in 6.3.1. Furthermore, by the same lemma, $d_1(e^k)$ has depth $\leq 2^{\alpha_2} + 2^{\alpha_1} \leq 2^{\max(\alpha_2,\alpha_1)+1} \leq 2^\alpha$. Hence we can take $(de)^k$ to be $d_1(e^k)$.

In the subcase where $d^k = \langle d_i \rangle_{i<\omega}$, $\mathrm{lev}(d) = k+1$ and $e^k = j$ we can take $(de)^k$ to be d_j, since clearly d_j has cutrank $\leq k$ and depth $\leq 2^\alpha$. If we are not in the above subcases, we can simply take $(de)^k$ to be $d^k e^k$. This derivation clearly has depth $\leq 2^\alpha$. Also it has cutrank $\leq k$, which can be seen as follows. If $\mathrm{lev}(d) \leq k+1$ we are done. But $\mathrm{lev}(d) \geq k+2$ is impossible, since we have assumed that de is not a simple application. In order to see this, note that if de is not a simple application, it must be of the form $d_0 d_1 \ldots d_n e$ with d_0 not an assumption variable or axiom and d_0 not itself of the form $d'd''$; then d_0 must end with an introduction $\to^+$ or ω, hence there is a cut of a degree exceeding $k+1$, which is excluded by assumption. □

As an immediate consequence we obtain:

THEOREM (Normalization for $\mathbf{Z}^\infty$). *For any derivation d^A of depth $\leq \alpha$ and cutrank $\leq k$ we can find a normal derivation $(d^*)^A$ with free assumption variables contained in those of d, which has depth $\leq 2^\alpha_k$, where $2^\alpha_0 := \alpha$, $2^\alpha_{m+1} := 2^{2^\alpha_m}$.*

As in 1.2.6 we can now analyze the structure of normal derivations in $\mathbf{Z}^\infty$. In particular we obtain:

THEOREM (Subformula property for $\mathbf{Z}^\infty$). *Let d be a normal deduction in $\mathbf{Z}^\infty$ for $\Gamma \vdash A$. Then each formula in d is a subformula of a formula in $\Gamma \cup \{A\}$.*

PROOF. We prove this for tracks of order n, by induction on n. □

6.4. Unprovable Initial Cases of Transfinite Induction

We now apply the technique of normalization for arithmetic with the ω-rule to obtain a proof that transfinite induction up to ε_0 is underivable in

$\mathbf{Z}$, i.e., a proof of

$$\mathbf{Z} \not\vdash \forall_x(\forall_{y\prec x} Py \to Px) \to \forall_x Px$$

with a relation symbol P, and that transfinite induction up to ω_{k+1} is underivable in $\mathbf{Z}_k$, i.e., a proof of

$$\mathbf{Z}_k \not\vdash \forall_x(\forall_{y\prec x} Py \to Px) \to \forall_{x\prec\omega_{k+1}} Px.$$

It clearly suffices to prove this for arithmetical systems based on classical logic. Hence we may assume that we have used only the classical versions (6.13), (6.14) and (6.15) of the axioms from 6.2.1.

Our proof is based on an idea of Schütte, which consists in adding a so-called *progression rule* to the infinitary systems. This rule allows us to conclude Pj (where j is any numeral) from all Pi for $i \prec j$.

6.4.1. Progression rule. More precisely, we define the notion of a derivation in $\mathbf{Z}^\infty + \mathrm{Prog}(P)$ of depth $\leq \alpha$ by the inductive clauses above and the additional clause $\mathrm{Prog}(P)$:

(Prog) If for all $i \prec j$ we have derivations d_i^{Pi} of depths $\leq \alpha_i < \alpha$, then $\langle d_i^{Pi}\rangle_{i\prec j}^{Pj}$ is a derivation of depth $\leq \alpha$.

We also define $\mathrm{cr}(\langle d_i\rangle_{i\prec j}) := \sup_{i\prec j} \mathrm{cr}(d_i)$.

Since this progression rule only deals with derivations of atomic formulas, it does not affect the cutranks of derivations. Hence the proof of normalization for $\mathbf{Z}^\infty$ carries over unchanged to $\mathbf{Z}^\infty + \mathrm{Prog}(P)$. In particular we have

Lemma (Rank reduction). *For any derivation d^A in $\mathbf{Z}^\infty + \mathrm{Prog}(P)$ of depth $\leq \alpha$ and cutrank $\leq k+1$ we can find a derivation $(d^k)^A$ in $\mathbf{Z}^\infty + \mathrm{Prog}(P)$ with free assumption variables contained in those of d, which has depth $\leq 2^\alpha$ and cutrank $\leq k$.*

We now show that from the progression rule for P we can easily derive the progressiveness of P.

Lemma. *We have a normal derivation of $\forall_x(\forall_{y\prec x} Py \to Px)$ in $\mathbf{Z}^\infty + \mathrm{Prog}(P)$ with depth 5.*

Proof.

$$\dfrac{\dots \quad \dfrac{\dfrac{\forall_{y\prec j} Py}{i\prec j \to Pi}\,\forall^- \qquad i\prec j}{Pi}\,{\to}^- \quad \dots \quad (\text{all } i\prec j)}{\dfrac{Pj}{\forall_{y\prec j} Py \to Pj}\,{\to}^+}\,Prog \qquad \dots \quad (\text{all } j)}{\forall_x(\forall_{y\prec x} Py \to Px)}\,\omega$$

□

6.4.2. Quasi-normal derivations. The crucial observation is that a normal derivation of $P^{\ulcorner}\beta^{\urcorner}$ must essentially have a depth of at least β. However, to obtain the right estimates for the subsystems $\mathbf{Z}_k$ we cannot apply rank reduction lemma down to cutrank 0 (i.e., to normal form) but must stop at cutrank 1. Such derivations, i.e., those of cutrank ≤ 1, will be called *quasi-normal*; they can also be analyzed easily.

We begin by showing that a quasi-normal derivation of a quantifier-free formula can always be transformed without increasing its cutrank or its depth into a quasi-normal derivation of the same formula which

(i) does not use the ω-rule, *and*
(ii) contains $\forall^-$ only in the initial part of a track starting with an axiom.

Recall that our axioms are of the form $\forall_{\vec{x}}A$ with A quantifier-free.

The *quasi-subformulas* of a formula A are defined by the clauses

(a) A, B are quasi-subformulas of $A \to B$;
(b) $A(i)$ is a quasi-subformula of $\forall_x A(x)$, for all numerals i;
(c) If A is a quasi-subformula of B, and C is an atomic formula, then $C \to A$ and $\forall_x A$ are quasi-subformulas of B;
(d) "... is quasi-subformula of ..." is a reflexive and transitive relation.

For example, $Q \to \forall_x(P \to A)$ with P, Q atomic is a quasi-subformula of $A \to B$.

We now transfer the subformula property for normal derivations in $\mathbf{Z}^\infty$ to a quasi-subformula property for quasi-normal derivations.

THEOREM (Quasi-subformula property). *Let d be a quasi-normal derivation in $\mathbf{Z}^\infty + \mathrm{Prog}(P)$ for $\Gamma \vdash A$. Then each formula in d is a quasi-subformula of a formula in $\Gamma \cup \{A\}$.*

PROOF. Again prove this for tracks of order n, by induction on n. □

COROLLARY. *Let d be a quasi-normal derivation in $\mathbf{Z}^\infty + \mathrm{Prog}(P)$ of a formula $\forall_{\vec{x}}A$ with A quantifier-free from quantifier-free assumptions. Then any track in d of positive order ends with a quantifier-free formula.*

PROOF. If not, then the major premise of the $\to^-$ whose minor premise is the offending end formula of the track, would contain a quantifier to the left of $\to$. This contradicts the theorem. □

6.4.3. Elimination of the ω-rule. Our next aim is to eliminate the ω-rule. For this we need the notion of an *instance* of a formula, defined by the following clauses.

(a) If B' is an instance of B and A is quantifier-free, then $A \to B'$ is an instance of $A \to B$;
(b) $A(i)$ is an instance of $\forall_x A(x)$, for all numerals i;

(c) The relation "... is an instance of ..." is reflexive and transitive.

LEMMA. *Let d be a quasi-normal derivation in $\mathbf{Z}^\infty + \mathrm{Prog}(P)$ of a formula A without $\forall$ to the left of $\to$, and from quantifier-free assumptions. Then for any quantifier-free instance A' of A we can find a quasi-normal derivation d' of A' from the same assumptions such that*

(a) *d' does not use the ω-rule,*
(b) *d' contains $\forall^-$ only in the initial elimination part of a track starting with an axiom, and*
(c) *$|d'| \leq |d|$.*

PROOF. By induction on the depth of d. We distinguish cases according to the last rule in d.

Case $\to^-$.

$$\frac{A \to B \qquad A}{B}\,{\to^-}$$

By the quasi-subformula property A must be quantifier-free. Let B' be a quantifier-free instance of B. Then by definition $A \to B'$ is a quantifier-free instance of $A \to B$. The claim now follows from the induction hypothesis.

Case $\to^+$.

$$\frac{B}{A \to B}\,{\to^+}$$

Any instance of $A \to B$ has the form $A \to B'$ with B' an instance of B. Hence the claim follows from the induction hypothesis.

Case $\forall^-$.

$$\frac{\forall_x A(x) \qquad i}{A(i)}\,{\forall^-}$$

Then any quantifier-free instance of $A(i)$ is also a quantifier-free instance of $\forall_x A(x)$, and hence the claim follows from the induction hypothesis.

Case ω.

$$\frac{\ldots \quad A(i) \quad \ldots \quad (\text{all } i < \omega)}{\forall_x A(x)}\,{\omega}$$

Any quantifier-free instance of $\forall_x A(x)$ has the form $A(i)'$ with $A(i)'$ a quantifier-free instance of $A(i)$. Hence the claim again follows from the induction hypothesis. □

A derivation d in $\mathbf{Z}^\infty + \mathrm{Prog}(P)$ is called a *$P\vec{\alpha}, \neg P\vec{\beta}$-refutation* if

(i) $\vec{\alpha}$ and $\vec{\beta}$ are disjoint, and

(ii) d derives a formula $\vec{A} \to B := A_1 \to \cdots \to A_k \to B$ with $\vec{A}$ and the free assumptions in d among $P^\ulcorner \alpha_1 \urcorner, \ldots, P^\ulcorner \alpha_m \urcorner, \neg P^\ulcorner \beta_1 \urcorner, \ldots, \neg P^\ulcorner \beta_n \urcorner$ or true quantifier-free formulas without P, and B a false quantifier-free formula without P or else among $P^\ulcorner \beta_1 \urcorner, \ldots, P^\ulcorner \beta_n \urcorner$

Intuitively, (ii) says that $\bigwedge_i P^\ulcorner \alpha_i \urcorner$ implies $\bigvee_j P^\ulcorner \beta_j \urcorner$.

LEMMA. *Let d be a quasi-normal $P\vec{\alpha}, \neg P\vec{\beta}$-refutation. Then*

$$\min(\vec{\beta}) \leq |d| + \mathrm{lh}(\vec{\alpha}'),$$

where $\vec{\alpha}'$ is the sublist of $\vec{\alpha}$ consisting of all $\alpha_i < \min(\vec{\beta})$, and $\mathrm{lh}(\vec{\alpha}')$ denotes the length of the list $\vec{\alpha}'$.

PROOF. By induction on $|d|$. By the previous lemma we may assume that d does not contain the ω-rule, and contains $\forall^-$ only in a context where leading universal quantifiers of an axiom are removed. We distinguish cases according to the last rule in d.

Case $\to^+$. By our definition of refutations the claim follows immediately from the induction hypothesis.

Case $\to^-$. Then $d = f^{C \to \vec{A} \to B} e^C$. If C is a true quantifier-free formula without P or of the form $P^\ulcorner \gamma \urcorner$ with $\gamma < \min(\vec{\beta})$, the claim follows from the induction hypothesis for f:

$$\min(\vec{\beta}) \leq |f| + \mathrm{lh}(\vec{\alpha}') + 1 \leq |d| + \mathrm{lh}(\vec{\alpha}').$$

If C is a false quantifier-free formula without P or of the form $P^\ulcorner \gamma \urcorner$ with $\min(\vec{\beta}) \leq \gamma$, the claim follows from the induction hypothesis for e:

$$\min(\vec{\beta}) \leq |e| + \mathrm{lh}(\vec{\alpha}') + 1 \leq |d| + \mathrm{lh}(\vec{\alpha}').$$

It remains to consider the case when C is a quantifier-free implication involving P. Then $\mathrm{lev}(C) \geq 1$, hence $\mathrm{lev}(C \to \vec{A} \to B) \geq 2$ and therefore (since $\mathrm{cr}(d) \leq 1$) f must be a simple application starting with an axiom. Now our only axioms involving P are $\mathrm{Eq}_P \colon \forall_{x,y}(x = y \to Px \to Py)$ and $\mathrm{Stab}_P \colon \forall_x(\neg\neg Px \to Px)$, and of these only Stab_P has the right form. Hence $f = \mathrm{Stab}_P{}^\ulcorner \gamma \urcorner$ and therefore $e \colon \neg\neg P^\ulcorner \gamma \urcorner$. Now from $\mathrm{lev}(\neg\neg P^\ulcorner \gamma \urcorner) = 2$, the assumption $\mathrm{cr}(e) \leq 1$ and again the form of our axioms involving P, it follows that e must end with $\to^+$, i.e., $e = \lambda_{u^{\neg P^\ulcorner \gamma \urcorner}} e_0^\bot$. So we have

$$\dfrac{\dfrac{f}{\neg\neg P^\ulcorner \gamma \urcorner \to P^\ulcorner \gamma \urcorner} \qquad \dfrac{\dfrac{[u \colon \neg P^\ulcorner \gamma \urcorner]}{\dfrac{e_0}{\bot}}}{\neg\neg P^\ulcorner \gamma \urcorner}}{P^\ulcorner \gamma \urcorner}$$

The claim now follows from the induction hypothesis for e_0.

Case $\forall^-$. By assumption we then are in the initial part of a track starting with an axiom. Since d is a $P\vec{\alpha}, \neg P\vec{\beta}$-refutation, that axiom must contain P. It cannot be the equality axiom $\mathrm{Eq}_P \colon \forall_{x,y}(x = y \to Px \to Py)$, since $\ulcorner\gamma\urcorner = \ulcorner\delta\urcorner \to P\ulcorner\gamma\urcorner \to P\ulcorner\delta\urcorner$ can never be (whether $\gamma = \delta$ or $\gamma \neq \delta$) the end formula of a $P\vec{\alpha}, \neg P\vec{\beta}$-refutation. For the same reason it can not be the stability axiom $\mathrm{Stab}_P \colon \forall_x(\neg\neg Px \to Px)$. Hence the case $\forall^-$ cannot occur.

Case $\mathrm{Prog}(P)$. Then $d = \langle d_\delta^{P\ulcorner\delta\urcorner}\rangle_{\delta<\gamma}^{P\ulcorner\gamma\urcorner}$. By assumption on d, γ is in $\vec{\beta}$. We may assume $\gamma = \beta_i := \min(\vec{\beta})$, for otherwise the premise deduction $d_{\beta_i} \colon P\ulcorner\beta_i\urcorner$ would be a quasi-normal $P\vec{\alpha}, \neg P\vec{\beta}$-refutation, to which we could apply the induction hypothesis.

If there are no $\alpha_j < \gamma$, then the argument is simple: every d_δ is a $P\vec{\alpha}, \neg P\vec{\beta}, \neg P\delta$-refutation, so by induction hypothesis, since also no $\alpha_j < \delta$,

$$\min(\vec{\beta}, \delta) = \delta \leq |d_\delta|,$$

hence $\gamma = \min(\vec{\beta}) \leq |d|$.

To deal with the situation that some α_j are less than γ, we observe that there can be at most finitely many α_j immediately preceding γ; so let ε be the least ordinal such that

$$\forall_\delta(\varepsilon \leq \delta < \gamma \to \delta \in \vec{\alpha}).$$

Then $\varepsilon, \varepsilon+1, \ldots, \varepsilon+k-1 \in \vec{\alpha}$, and $\varepsilon + k = \gamma$. ε is either a successor or a limit. *Case* $\varepsilon = \varepsilon' + 1$. Since $d_{\varepsilon'}$ is a $P\vec{\alpha}, \neg P\vec{\beta}, \neg P(\varepsilon-1)$-refutation, it follows by the induction hypothesis that

$$\varepsilon - 1 \leq |d_{\varepsilon-1}| + \mathrm{lh}(\vec{\alpha}') - k,$$

where $\vec{\alpha}'$ is the sequence of $\alpha_j < \gamma$. Hence $\varepsilon \leq |d| + \mathrm{lh}(\vec{\alpha}') - k$, and therefore

$$\gamma \leq |d| + \mathrm{lh}(\vec{\alpha}').$$

Case ε is a limit. Then there is a sequence $\langle\delta_{f(n)}\rangle_n$ with limit ε, and with all $\alpha_j < \varepsilon$ below $\delta_{f(0)}$, and therefore by induction hypothesis

$$\delta_{f(n)} \leq |d_{f(n)}| + \mathrm{lh}(\vec{\alpha}') - k.$$

Hence $\varepsilon \leq |d_{f(n)}| + \mathrm{lh}(\vec{\alpha}') - k$, so $\gamma \leq |d| + \mathrm{lh}(\vec{\alpha}')$. □

Theorem (Underivability of transfinite induction in $\mathbf{Z}$). *Transfinite induction up to ε_0 is underivable in $\mathbf{Z}$, i.e.*

$$\mathbf{Z} \nvdash \forall_x(\forall_{y\prec x} Py \to Px) \to \forall_x Px$$

with a relation symbol P, and for $k \geq 3$ transfinite induction up to ω_{k+1} is underivable in $\mathbf{Z}_k$, i.e.,

$$\mathbf{Z}_k \nvdash \forall_x(\forall_{y\prec x} Py \to Px) \to \forall_{x\prec\omega_{k+1}} Px.$$

PROOF. We restrict ourselves to the second part. So assume that transfinite induction up to ω_{k+1} is derivable in $\mathbf{Z}_k$. Then by the embedding of $\mathbf{Z}_k$ into $\mathbf{Z}^\infty$ and the normal derivability of the progressiveness of P in $\mathbf{Z}^\infty + \mathrm{Prog}(P)$ with finite depth (proved in 6.4.1) we can conclude that $\forall_{x\prec\omega_{k+1}} Px$ is derivable in $\mathbf{Z}^\infty + \mathrm{Prog}(P)$ with depth $< \omega^{m+1}$ and cutrank $\leq k$. (Note that here we need $k \geq 3$, since the formula expressing progressiveness of P has level 3). Now $k-1$ applications of the rank reduction lemma yield a derivation of the same formula $\forall_{x\prec\omega_{k+1}} Px$ in $\mathbf{Z}^\infty + \mathrm{Prog}(P)$ with depth $\gamma < 2_{k-1}^{\omega^{m+1}} < \omega_{k+1}$ and cutrank ≤ 1.

Hence there is also a quasi-normal derivation of $P^\ulcorner\gamma + 3^\urcorner$ in $\mathbf{Z}^\infty + \mathrm{Prog}(P)$ with depth $\gamma + 2$ and cutrank ≤ 1, of the form

$$\dfrac{\dfrac{\begin{matrix}d\\ \forall_{x\prec\omega_{k+1}} Px\end{matrix}}{\ulcorner\gamma+3\urcorner \prec \omega_{k+1} \to P^\ulcorner\gamma+3\urcorner} \qquad \begin{matrix}d'\\ \ulcorner\gamma+3\urcorner \prec \omega_{k+1}\end{matrix}}{P^\ulcorner\gamma+3\urcorner}$$

where d' is a deduction of finite depth (it may even be an axiom, depending on the precise choice of axioms for $\mathbf{Z}$); this contradicts the last lemma. □

6.4.4. Normalization for arithmetic is impossible. The normalization theorem for first-order logic applied to one of our arithmetical systems $\mathbf{Z}$ is not particularly useful since we may have used in our derivation induction axioms of arbitrary complexity. Hence it is tempting to first eliminate the induction schema in favour of an induction rule allowing us to conclude $\forall_x A(x)$ from a derivation of $A(0)$ and a derivation of $A(\mathrm{S}x)$ with an additional assumption $A(x)$ to be cancelled at this point (note that this rule is equivalent to the induction schema), and then to try to normalize the resulting derivation in the new system $\mathbf{Z}$ with the induction rule. We will apply the theorems on underivability of transfinite induction up to ε_0 in $\mathbf{Z}$ and on provable initial cases of transfinite induction in $\mathbf{Z}$ to show that even a very weak form of the normalization theorem cannot hold in $\mathbf{Z}$ with the induction rule.

THEOREM. *The following weak form of a normalization theorem for* $\mathbf{Z}$ *with the induction rule is false: "For any derivation* d^B *with free assumption variables among* $\vec{u}^{\vec{A}}$ *for formulas* $\vec{A}, B$ *of level* $\leq l$ *there is a derivation* $(d^*)^B$*, with free assumption variables contained in those of* d*, which contains only formulas of level* $\leq k$*, where* k *depends on* l *only."*

PROOF. Assume that such a normalization theorem holds. Consider the formula

$$\forall_x(\forall_{y\prec x} Py \to Px) \to \forall_{x\prec\omega_{n+1}} Px$$

expressing transfinite induction up to ω_{n+1}, which is of level 3. By the theorem on provable initial cases of transfinite induction in $\mathbf{Z}$ it is derivable in $\mathbf{Z}$. Now from our assumption it follows that there exists a derivation of this formula containing only formulas of level $\leq k$, for some k independent of n. Hence $\mathbf{Z}_k$ derives transfinite induction up to ω_{n+1} for any n. But this clearly contradicts the theorem on underivability of transfinite induction up to ε_0 in $\mathbf{Z}$. □

APPENDIX A

Normal Functions

Veblen (1908) investigated the notion of a continuous monotonic function on a segment of the ordinals, and introduced a certain hierarchy of normal functions. His goal was to generalize Cantor's (1897) theory of ε-numbers.

A.1. Closed Unbounded Classes

Let Ω be a regular cardinal $> \omega$ or $\Omega = \mathrm{On}$. An important example is $\Omega = \aleph_1$, that is the case where Ω is the set of all countable ordinals. Let $\alpha, \beta, \gamma, \delta, \varepsilon, \xi, \eta, \zeta$ denote elements of Ω. A function $\varphi\colon \Omega \to \Omega$ is *monotone* if $\alpha < \beta$ implies $\varphi\alpha < \varphi\beta$. φ is *continuous* if $\varphi\alpha = \sup_{\xi<\alpha} \varphi\xi$ for every limit α. φ is *normal* if φ is monotone and continuous.

LEMMA. *For every monotone function φ we have $\alpha \leq \varphi\alpha$.*

PROOF. Induction on α. *Case* 0. $0 \leq \varphi 0$. *Case* $\alpha + 1$. $\alpha \leq \varphi\alpha < \varphi(\alpha + 1)$. *Case* α limit. $\alpha = \sup_{\xi<\alpha} \xi \leq \sup_{\xi<\alpha} \varphi\xi \leq \varphi\alpha$. □

A class $\mathcal{B} \subseteq \Omega$ is *bounded* if $\sup(\mathcal{B}) \in \Omega$. A class $\mathcal{A} \subseteq \Omega$ is *closed* if for every bounded subclass $\mathcal{B} \subseteq \mathcal{A}$ we have $\sup(\mathcal{B}) \in \mathcal{A}$. Closed unbounded classes $\mathcal{A} \subseteq \Omega$ are called *normal* or *closed unbounded* in Ω (club for short).

If for instance $\Omega = \Omega_1$, then every $\mathcal{B} \subseteq \Omega$ is a set, and $\mathcal{B}$ is bounded iff $\mathcal{B}$ is countable. If $\Omega = \mathrm{On}$, then $\mathcal{B}$ is bounded iff $\mathcal{B}$ is a set.

By the corollary to Mostowski's isomorphy theorem for every $\mathcal{A} \subseteq \mathrm{On}$ we have a uniquely determined isomorphism of an ordinal class onto $\mathcal{A}$, that is an $f\colon \mathrm{On} \to \mathcal{A}$ (or $f\colon \alpha \to \mathcal{A}$). This isomorphism is called the *ordering function* of $\mathcal{A}$. Notice that f is the *monotone enumeration* of $\mathcal{A}$.

LEMMA. *The range of a normal function is a normal class. Conversely, the ordering function of a normal class is a normal function.*

PROOF. Let φ be a normal function. $\varphi[\Omega]$ is unbounded, since for every α we have $\alpha \leq \varphi\alpha$. We now show that $\varphi[\Omega]$ is closed. Let $\mathcal{B} = \{\, \varphi\xi \mid \xi \in \mathcal{A} \,\}$ be bounded, i.e., $\sup(\mathcal{B}) \in \Omega$. Because of $\xi \leq \varphi\xi$ then also $\mathcal{A}$ is bounded. We must show $\sup(\mathcal{B}) = \varphi\alpha$ for some α. If $\mathcal{A}$ has a maximal element we are done. Otherwise $\alpha := \sup(\mathcal{A})$ is a limit. Then $\varphi\alpha = \sup_{\xi<\alpha} \varphi\xi =$

$\sup_{\xi\in\mathcal{A}}\varphi\xi = \sup(\mathcal{B})$. Conversely, let $\mathcal{A}$ be closed and unbounded. We define a function $\varphi\colon \Omega \to \mathcal{A}$ by transfinite recursion, as follows.

$$\varphi\alpha := \min\{\,\gamma \in \mathcal{A} \mid \forall_\xi.\xi < \alpha \to \varphi\xi < \gamma\,\}.$$

φ is well defined, since $\mathcal{A}$ is unbounded. Clearly φ is the ordering function of $\mathcal{A}$ and hence monotone. It remains to be shown that φ is continuous. Let α be a limit. Since $\varphi[\alpha]$ is bounded (this follows from $\varphi\xi < \varphi\alpha$ for $\xi < \alpha$) and $\mathcal{A}$ is closed, we have $\sup_{\xi\in\alpha}\varphi\xi \in \mathcal{A}$, hence by definition $\varphi\alpha = \sup_{\xi\in\alpha}\varphi\xi$. □

LEMMA. *The fixed points of a normal function form a normal class.*

PROOF. (Cf. Cantor (1897, p. 242)). Let φ be a normal function. For every ordinal α we can construct a fixed point $\beta \geq \alpha$ of φ by

$$\beta := \sup\{\,\varphi^n\alpha \mid n \in \mathbb{N}\,\}.$$

Hence the class of fixed points of φ is unbounded. It is closed as well, since for every class $\mathcal{B}$ of fixed points of φ we have $\varphi(\sup(\mathcal{B})) = \sup\{\,\varphi\alpha \mid \alpha \in \mathcal{B}\,\} = \sup\{\,\alpha \mid \alpha \in \mathcal{B}\,\} = \sup(\mathcal{B})$, i.e., $\sup(\mathcal{B})$ is a fixed point of φ. □

A.2. The Veblen Hierarchy of Normal Functions

The ordering function of the class of fixed points of a normal function φ has been called by Veblen the *first derivative* φ' of φ. For example, the first derivative of the function ω^ξ is the function ε_ξ.

LEMMA (Veblen). *Let $(\mathcal{A}_\gamma)_{\gamma<\beta}$ with β limit be a decreasing sequence of normal classes. Then the intersection $\bigcap_{\gamma<\beta}\mathcal{A}_\gamma$ is normal as well.*

PROOF. Unboundedness. Let α be given and $\delta_\gamma := \min\{\,\xi \in \mathcal{A}_\gamma \mid \xi > \alpha\,\}$. Then $(\delta_\gamma)_{\gamma<\beta}$ is weakly monotonic. Let $\delta := \sup_{\gamma<\beta}\delta_\gamma$. Then $\delta \in \mathcal{A}_\gamma$ for every $\gamma < \beta$, since the $\mathcal{A}_\gamma$ decrease. Hence $\alpha < \delta \in \bigcap_{\gamma<\beta}\mathcal{A}_\gamma$.

Closedness. Let $\mathcal{B} \subseteq \bigcap_{\gamma<\beta}\mathcal{A}_\gamma$, $\mathcal{B}$ bounded. Then $\mathcal{B} \subseteq \mathcal{A}_\gamma$ for every $\gamma < \beta$ and therefore $\sup(\mathcal{B}) \in \mathcal{A}_\gamma$. Hence $\sup(\mathcal{B}) \in \bigcap_{\gamma<\beta}\mathcal{A}_\gamma$. □

We now define the *Veblen hierarchy of normal functions*. It is based on an arbitrary given normal function $\varphi\colon \Omega \to \Omega$. We use transfinite recursion to define for every $\beta \in \Omega$ a normal function $\varphi_\beta\colon \Omega \to \Omega$:

$\varphi_0 := \varphi$,

$\varphi_{\beta+1} := (\varphi_\beta)'$,

for limits β let φ_β be the ordering function of $\bigcap_{\gamma<\beta}\varphi_\gamma[\Omega]$.

For example, for $\varphi\alpha := 1 + \alpha$ we obtain $\varphi_\beta\alpha = \omega^\beta + \alpha$. If we start with $\varphi\alpha := \omega^\alpha$, then $\varphi_1\alpha = \varepsilon_\alpha$ and φ_2 enumerates the critical ε-numbers, i.e., the ordinals α such that $\varepsilon_\alpha = \alpha$.

LEMMA. *Let $\beta > 0$. Then φ_β is the ordering function of the class of all common fixed points of all φ_γ for $\gamma < \beta$.*

PROOF. We must show $\varphi_\beta[\Omega] = \{\, \xi \mid \forall_\gamma.\gamma < \beta \to \varphi_\gamma\xi = \xi \,\}$.

$\subseteq$. This is proved by transfinite induction on β. In case $\beta + 1$ every $\varphi_{\beta+1}\alpha$ is a fixed point of φ_β and hence by induction hypothesis also a fixed point of all φ_γ for $\gamma < \beta$. If β is a limit, then the claim follows from $\varphi_\beta[\Omega] = \bigcap_{\gamma<\beta} \varphi_\gamma[\Omega]$.

$\supseteq$. Let ξ such that $\forall_\gamma.\gamma < \beta \to \varphi_\gamma\xi = \xi$ be given. If β is a successor, then $\xi \in \varphi_\beta[\Omega]$ by definition of φ_β. If β is a limit, then $\xi \in \bigcap_{\gamma<\beta} \varphi_\gamma[\Omega] = \varphi_\beta[\Omega]$. □

It follows that $\varphi_\gamma(\varphi_\beta\xi) = \varphi_\beta\xi$ for every $\gamma < \beta$.

A further normal function can be obtained as follows. From each of the normal classes $\varphi_\beta[\Omega]$ pick the least fixed point. The class formed in this way again is normal, hence can be enumerated by a normal function. This normal function assigns to every β the ordinal $\varphi_\beta 0$.

LEMMA. *If φ is a normal function with $0 < \varphi 0$, then $\lambda_\beta \varphi_\beta 0$ is a normal function as well.*

PROOF. We first show

$$\beta < \gamma \to \varphi_\beta 0 < \varphi_\gamma 0,$$

by induction on γ. Let $\beta < \gamma$. Observe that $0 < \varphi_\beta 0$ by induction hypothesis or in case $\beta = 0$ by assumption. Hence 0 is not a fixed point of φ_β and therefore $0 < \varphi_\gamma 0$. But this implies $\varphi_\beta 0 < \varphi_\beta(\varphi_\gamma 0) = \varphi_\gamma 0$.

We now show that $\lambda_\beta \varphi_\beta 0$ is continuous. Let $\delta := \sup_{\beta<\gamma} \varphi_\beta 0$ with γ limit. We must show $\delta = \varphi_\gamma 0$. Because of $\varphi_\beta 0 \in \varphi_\alpha[\Omega]$ for all $\alpha \leq \beta < \gamma$ and since $\varphi_\alpha[\Omega]$ is closed we have $\delta \in \varphi_\alpha[\Omega]$, hence $\delta \in \bigcap_{\alpha<\gamma} \varphi_\alpha[\Omega] = \varphi_\gamma[\Omega]$ and therefore $\delta \geq \varphi_\gamma 0$. On the other hand $\varphi_\beta 0 < \varphi_\beta(\varphi_\gamma 0) = \varphi_\gamma 0$, hence $\delta \leq \varphi_\gamma 0$. □

The fixed points of this function, i.e., the ordinals α such that $\varphi_\alpha 0 = \alpha$, are called *strongly critical* ordinals. Observe that they depend on the given normal function $\varphi = \varphi_0$. Their ordering function is usually denoted by Γ. Hence by definition $\Gamma_0 := \Gamma 0$ is the least ordinal β such that $\varphi_\beta 0 = \beta$.

A.3. φ Normal Form

We generalize Cantor's normal form, using the Veblen hierarchy instead of ω^ξ.

LEMMA.

$$\varphi_{\beta_0}\alpha_0 < \varphi_{\beta_1}\alpha_1 \leftrightarrow \begin{cases} \alpha_0 < \varphi_{\beta_1}\alpha_1, & \textit{if } \beta_0 < \beta_1, \\ \alpha_0 < \alpha_1, & \textit{if } \beta_0 = \beta_1, \\ \varphi_{\beta_0}\alpha_0 < \alpha_1, & \textit{if } \beta_0 > \beta_1, \end{cases} \tag{A.1}$$

$$\varphi_{\beta_0}\alpha_0 = \varphi_{\beta_1}\alpha_1 \leftrightarrow \begin{cases} \alpha_0 = \varphi_{\beta_1}\alpha_1, & \textit{if } \beta_0 < \beta_1, \\ \alpha_0 = \alpha_1, & \textit{if } \beta_0 = \beta_1, \\ \varphi_{\beta_0}\alpha_0 = \alpha_1, & \textit{if } \beta_0 > \beta_1. \end{cases} \tag{A.2}$$

PROOF. $\leftarrow$. (A.1). If $\beta_0 < \beta_1$ and also $\alpha_0 < \varphi_{\beta_1}\alpha_1$, then $\varphi_{\beta_0}\alpha_0 < \varphi_{\beta_0}\varphi_{\beta_1}\alpha_1 = \varphi_{\beta_1}\alpha_1$. If $\beta_0 = \beta_1$ and $\alpha_0 < \alpha_1$, then $\varphi_{\beta_0}\alpha_0 < \varphi_{\beta_1}\alpha_1$. If $\beta_0 > \beta_1$ and $\varphi_{\beta_0}\alpha_0 < \alpha_1$, then $\varphi_{\beta_0}\alpha_0 = \varphi_{\beta_1}\varphi_{\beta_0}\alpha_0 < \varphi_{\beta_1}\alpha_1$. For (A.2) one argues similarly.

$\rightarrow$. If the right hand side of (A.1) is false, we have

$$\begin{cases} \alpha_1 \leq \varphi_{\beta_0}\alpha_0, & \text{if } \beta_1 < \beta_0, \\ \alpha_1 \leq \alpha_0, & \text{if } \beta_1 = \beta_0, \\ \varphi_{\beta_1}\alpha_1 \leq \alpha_0, & \text{if } \beta_1 > \beta_0, \end{cases}$$

hence by $\Leftarrow$ (with 0 and 1 exchanged) $\varphi_{\beta_1}\alpha_1 < \varphi_{\beta_0}\alpha_0$ or $\varphi_{\beta_1}\alpha_1 = \varphi_{\beta_0}\alpha_0$, hence $\neg(\varphi_{\beta_0}\alpha_0 < \varphi_{\beta_1}\alpha_1)$. If the right hand side of (A.2) is false, we have

$$\begin{cases} \alpha_0 \neq \varphi_{\beta_1}\alpha_1, & \text{if } \beta_0 < \beta_1, \\ \alpha_0 \neq \alpha_1, & \text{if } \beta_0 = \beta_1, \\ \varphi_{\beta_0}\alpha_0 \neq \alpha_1, & \text{if } \beta_0 > \beta_1, \end{cases}$$

and hence by $\Leftarrow$ in (A.1) either $\varphi_{\beta_0}\alpha_0 < \varphi_{\beta_1}\alpha_1$ or $\varphi_{\beta_1}\alpha_1 < \varphi_{\beta_0}\alpha_0$, hence $\varphi_{\beta_0}\alpha_0 \neq \varphi_{\beta_1}\alpha_1$. □

COROLLARY. *If $\beta_0 \leq \beta_1$, then $\varphi_{\beta_0}\alpha \leq \varphi_{\beta_1}\alpha$.*

PROOF. Assume $\beta_0 < \beta_1$. By the lemma (for $\leq$) it suffices to show $\alpha \leq \varphi_{\beta_1}\alpha$. But this follows from the first lemma in A.1. □

COROLLARY. *If $\varphi_{\beta_0}\alpha_0 = \varphi_{\beta_1}\alpha_1$, then $\alpha_0 = \alpha_1$ and $\beta_0 = \beta_1$, provided $\alpha_0 < \varphi_{\beta_0}\alpha_0$ and $\alpha_1 < \varphi_{\beta_1}\alpha_1$.*

PROOF. *Case* $\beta_0 = \beta_1$. Then $\alpha_0 = \alpha_1$ follows from the lemma. *Case* $\beta_0 < \beta_1$. By the lemma we have $\alpha_0 = \varphi_{\beta_1}\alpha_1 = \varphi_{\beta_0}\alpha_0$, contradicting our assumption. *Case* $\beta_1 < \beta_0$. Similar. □

COROLLARY. *If φ is a normal function with $0 < \varphi 0$, then every fixed point α of $\varphi = \varphi_0$ can be written uniquely in the form $\alpha = \varphi_\beta \alpha'$ with $\alpha' < \alpha$.*

PROOF. We have $\alpha + 1 \leq \varphi_{\alpha+1}0$ by the last lemma in A.2 and hence $\alpha < \varphi_{\alpha+1}\alpha$. Now let β be minimal such that $\alpha < \varphi_\beta\alpha$. By assumption $0 < \beta$. Since α is a fixed point of all φ_γ with $\gamma < \beta$, we have $\alpha = \varphi_\beta\alpha'$ for some α'. Using $\alpha < \varphi_\beta\alpha$ it follows that $\alpha' < \alpha$.

Uniqueness. Let in addition $\alpha = \varphi_{\beta_1}\alpha_1$ with $\alpha_1 < \alpha$. Then $\alpha_1 < \varphi_{\beta_1}\alpha_1$, hence $\beta \leq \beta_1$ by the choice of β. Now if $\beta < \beta_1$, then we would obtain $\varphi_\beta\alpha = \varphi_\beta\varphi_{\beta_1}\alpha_1 = \varphi_{\beta_1}\alpha_1 = \alpha$, contradicting our choice of β. Hence $\beta = \beta_1$ and therefore $\alpha_1 = \alpha$. □

We now show that every ordinal can be written uniquely in a certain φ normal form. Here we assume that our initial normal function $\varphi_0 = \varphi$ is the exponential function with base ω.

THEOREM (φ normal form). *Let $\varphi_0\xi := \omega^\xi$. Then every ordinal α can be written uniquely in the form*

$$\alpha = \varphi_{\beta_1}\alpha_1 + \cdots + \varphi_{\beta_n}\alpha_n$$

with $\varphi_{\beta_1}\alpha_1 \geq \cdots \geq \varphi_{\beta_n}\alpha_n$ and $\alpha_i < \varphi_{\beta_i}\alpha_i$ for $i = 1, \ldots, k$. If $\alpha < \Gamma_0$, then in addition we have $\beta_i < \varphi_{\beta_i}\alpha_i$ for $i = 1, \ldots, n$.

PROOF. Existence. First write α in Cantor normal form $\alpha = \varphi_0\delta_1 + \cdots + \varphi_0\delta_n$ with $\delta_1 \geq \cdots \geq \delta_n$. Every summand with $\delta_i < \varphi_0\delta_i$ is left unchanged. Every other summand satisfies $\delta_i = \varphi_0\delta_i$ and hence by the last corollary can be replaced by $\varphi_\beta\alpha'$ where $\alpha' < \varphi_\beta\alpha'$.

Uniqueness. Let

$$\alpha = \varphi_{\beta_1}\alpha_1 + \cdots + \varphi_{\beta_n}\alpha_n = \varphi_{\beta'_1}\alpha'_1 + \cdots + \varphi_{\beta'_m}\alpha'_m$$

and assume that both representations are different. Since no such sum can extend the other, we must have $i \leq n, m$ such that $(\beta_i, \alpha_i) \neq (\beta'_i, \alpha'_i)$. By property (d) of ordinal addition we can assume $i = 1$. Now if say $\varphi_{\beta_1}\alpha_1 < \varphi_{\beta'_1}\alpha'_1$, then we would have (since $\varphi_{\beta'_1}\alpha'_1$ is an additive principal number and $\varphi_{\beta_1}\alpha_1 \geq \cdots \geq \varphi_{\beta_n}\alpha_n$)

$$\varphi_{\beta_1}\alpha_1 + \cdots + \varphi_{\beta_n}\alpha_n < \varphi_{\beta'_1}\alpha'_1 \leq \varphi_{\beta'_1}\alpha'_1 + \cdots + \varphi_{\beta'_m}\alpha'_m,$$

a contradiction.

We must show that in case $\alpha < \Gamma_0$ we have $\beta_i < \varphi_{\beta_i}\alpha_i$ for $i = 1, \ldots, n$. So assume $\varphi_{\beta_i}\alpha_i \leq \beta_i$ for some i. Then

$$\varphi_{\beta_i}0 \leq \varphi_{\beta_i}\alpha_i \leq \beta_i \leq \varphi_{\beta_i}0,$$

hence $\varphi_{\beta_i}0 = \beta_i$ and hence

$$\Gamma_0 \leq \beta_i = \varphi_{\beta_i}0 \leq \varphi_{\beta_i}\alpha_i \leq \alpha.$$ □

From the $\varphi_\beta(\alpha)$ one obtains a unique notation system for ordinals below $\Gamma_0 := \Gamma 0$. Observe however that $\Gamma_0 = \varphi_{\Gamma_0}0$. by definition of Γ_0.

A.4. Notes

The hierarchy of normal functions defined in Section A.2 has been extended by Veblen (1908) to functions with more than one argument. Schütte (1954) studied these functions carefully and could show that they can be used for a cosntructive representation of a segment of the ordinals far bigger than Γ_0. To this end he introduced so-called "Klammersymbole" to denote the multiary Veblen-functions.

Bachmann extended the Veblen hierarchy using the first uncountable ordinal Ω. His approach has later been extended by means of symbols for higher number classes, first by Pfeiffer for finite number clasees and then by Isles for transfinite number classes. However, the resulting theory was quite complicated and difficult to work with. An idea of Feferman then simplified the subject considerably. He introduced functions $\theta_\alpha\colon \mathrm{On} \to \mathrm{On}$ for $\alpha \in \mathrm{On}$, that form again a hierarchy of normal functions and extend the Veblen hierarchy. One usually writes $\theta\alpha\beta$ instad of $\theta_\alpha(\beta)$ and views θ as a binay function. The ordinals $\theta\alpha\beta$ can be defined byransfinite recursion on α, as follows. Assume thatd θ_ξ for every $\xi < \alpha$ is defined already. Let $C(\alpha, \beta)$ be the set of all ordinals that can be generated from ordinals $< \beta$ and say the constants $0, \aleph_1, \ldots, \aleph_\omega$ by means of the functions $+$ and $\theta{\restriction}\{\, \xi \mid \xi < \alpha \,\} \times \mathrm{On}$. An ordinal β is aclled α-critical if $\beta \notin C(\alpha, \beta)$. Than $\theta_\alpha\colon \mathrm{On} \to \mathrm{On}$ is defined as the ordering function of the class of all α-critical ordinals.

Buchholz (1986) observed that the second argument β in $\theta\alpha\beta$ is not used in any essential way, and that the functions $\alpha \mapsto \theta\alpha\aleph_v$ with $v = 0, 1, \ldots, \omega$ generate a notation system for ordinals of the same strength as the system with the binary θ-function. He then went on and defined directly functions ψ_v with $v \leq \omega$, that correspond to $\alpha \mapsto \theta\alpha\aleph_v$. More precisely he defined $\psi_v\alpha$ for $\alpha \in \mathrm{On}$ and $v \leq \omega$ by transfinite recursion on α (simultaneously for all v), as follows.

$$\psi_v\alpha := \min\{\, \gamma \mid \gamma \notin C_v(\alpha) \,\},$$

where $C_v(\alpha)$ is the set of all ordinals that can be generated from the ordinals $< \aleph_v$ by the functions $+$ and all $\psi_u{\restriction}\{\, \xi \mid \xi < \alpha \,\}$ with $u \leq \omega$.

Bibliography

E. Beth. Semantic construction of intuitionistic logic. *Medelingen de KNAW N.S.*, 19(11), 1956.

E. Börger, E. Grädel, and Y. Gurevich. *The Classical Decision Problem.* Perspectives in Mathematical Logic. Springer Verlag, Berlin, Heidelberg, New York, 1997.

W. Buchholz. A new system of proof-theoretic ordinal functions. *Annals of Pure and Applied Logic*, 32(3):195–207, 1986.

G. Cantor. Beträge zur Begründung der transfiniten Mengenlehre. *Mathematische Annalen*, 49, 1897.

C. Chang and H. Keisler. *Model Theory*, volume 73 of *Studies in Logic.* North–Holland, Amsterdam, 3rd edition, 1990.

N. G. de Bruijn. Lambda calculus notation with nameless dummies, a tool for automatic formula manipulation, with application to the Church-Rosser theorem. *Indagationes Math.*, 34:381–392, 1972.

H.-D. Ebbinghaus, J. Flum, and W. Thomas. *Einführung in die mathematische Logik.* Spektrum Akademischer Verlag, Heidelberg, Berlin, Oxford, 4. edition, 1996.

S. Feferman. Arithmetization of metamathematics in a general setting. *Fundamenta Mathematicae*, XLIX:35–92, 1960.

S. Feferman, J. W. Dawson, et al., editors. *Kurt Gödel Collected Works, Volume I-V.* Oxford University Press, 1986, 1990, 1995, 2002, 2002.

G. Gentzen. Untersuchungen über das logische Schließen I, II. *Mathematische Zeitschrift*, 39:176–210, 405–431, 1935.

G. Gentzen. Beweisbarkeit und Unbeweisbarkeit von Anfangsfällen der transfiniten Induktion in der reinen Zahlentheorie. *Mathematische Annalen*, 119:140–161, 1943.

J.-Y. Girard. *Proof Theory and Logical Complexity.* Bibliopolis, Napoli, 1987.

K. Gödel. Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I. *Monatshefte für Mathematik und Physik*, 38: 173–198, 1931.

D. Hilbert and P. Bernays. *Grundlagen der Mathematik II*, volume 50 of *Grundlehren der mathematischen Wissenschaften.* Springer Verlag,

Berlin, Heidelberg, New York, Berlin, second edition, 1970.

I. Johansson. Der Minimalkalkül, ein reduzierter intuitionistischer Formalismus. *Compositio Mathematica*, 4:119–136, 1937.

L. Kalmár. Ein einfaches Beispiel für ein unentscheidbares Problem (hungarian, with german summary). *Mat. Fiz. Lapok*, 50:1–23, 1943.

A. N. Kolmogorov. Zur Deutung der intuitionistischen Logik. *Math. Zeitschr.*, 35:58–65, 1932.

J. Łoś. Quelques remarques, théorèmes et problèmes sur les classes définissables d'algèbres. In *Mathematical Interpretation of Formal Systems*, pages 98–113. North–Holland, Amsterdam, 1955.

V. Orevkov. Lower bounds for increasing complexity of derivations after cut elimination. *Zapiski Nauchnykh Seminarov Leningradskogo*, 88:137–161, 1979.

R. M. Robinson. An essentially undecidable axiom system. In *Proceedings of the International Congress of Mathematicians (Cambridge 1950)*, volume I, pages 729–730, 1950.

K. Schütte. Kennzeichnung von Ordinalzahlen durch rekursiv definierte Funktionen. *Mathematische Annalen*, 127:16–32, 1954.

J. Shepherdson and H. Sturgis. Computability of recursive functions. *J. Ass. Computing Machinery*, 10:217–255, 1963.

J. R. Shoenfield. *Mathematical Logic*. Addison–Wesley Publ. Comp., Reading, Massachusetts, 1967.

C. Smoryński. *Logical Number Theory I*. Universitext. Springer Verlag, Berlin, Heidelberg, New York, 1991.

W. W. Tait. Infinitely long terms of transfinite type I. In J. Crossley and M. Dummett, editors, *Formal Systems and Recursive Functions*, pages 176–185. North–Holland, Amsterdam, 1965.

A. S. Troelstra and H. Schwichtenberg. *Basic Proof Theory*. Cambridge University Press, second edition, 2000.

O. Veblen. Continuous increasing functions of finite and transfinite ordinals. *Transactions AMS*, 9:280–292, 1908.

Index