The Syntax of First-Order Logic

Marc Hoyois

Table of Contents

Introduction

The reader is invited to read Chapter 1 of Joe Shoenfield's book *Mathematical Logic* for a remarkable introduction to the subject. Here I will only explain in what respects this text differs from the numerous other texts on mathematical logic.

The distinguishing particularity of this text is that it is exclusively concerned with what Shoenfield calls the *syntactical study* of first-order axiom systems, and herein only with those results that deal with concrete objects and can be proved in a constructive manner. We use the adjective *finitary* to describe such objects and such proofs. Without exceptions, all the definitions and theorems in this text fall into this category.

Following is a quick review each chapter.

I First-Order Theories

There is not much to say about this chapter whose content is completely standard (except, as explained above, that all the semantic notions are absent). The presentation often follows Shoenfield's closely, and with few exceptions I have used the same terminology so as not to confuse readers that are familiar with his book.

II The Question of Consistency

This chapter is roughly the equivalent of Shoenfield's Chapter 4 (again, minus the semantic part), to which I owe many of the proofs. One apparent difference is the early treatment of extensions by definitions. This is made possible by a new direct proof of the conservativity of extensions by definitions of function symbols, which is more efficient than Shoenfield's. In the section on interpretations some more material will be found, such as the concepts of isomorphism of interpretations and absoluteness, here presented in a general setting.

The proofs of the consistency theorem and Herbrand's theorem in the next section are those from Shoenfield, but I have made explicit some interesting corollaries that do not appear in Shoenfield's book, such as the Herbrand–Skolem theorem which implies that Shoenfield's version of Herbrand's theorem is true of arbitrary first-order theories (not only of those without nonlogical axioms) and the fact that any first-order theory has a conservative Skolem extension, i.e., an extension in which every instantiation has a witnessing term. For the latter result a new proof of the conservativity of the Henkin extension has been devised that generalizes easily to the Skolem extension. Finally, a constructive proof of Craig's interpolation lemma is given.

III The Incompleteness Theorem

This chapter gives two detailed proofs of the (first) incompleteness theorem. The first relies on the notion of recursive function and Church's theorem on undecidability, and the second uses Rosser's explicit construction of an undecidable formula. I have taken great care to make it apparent that *either* approach is completely constructive. In the first case, this requires a slight deviation from the usual proof, analogous to Rosser's improvement of Gödel's original argument to remove the hypothesis of ω-consistency, but happening on the metamathematical level.

Finally, in the last section, minimal arithmetic is introduced and it is proved that any first-order theory in which minimal arithmetic has an interpretation satisfies the hypotheses of both versions of the incompleteness theorem. It is also proved, using the finitary methods of Chapter II, that minimal arithmetic itself is consistent.

IV First-Order Number Theory

In the first section a syntactical analogue to recursiveness is introduced, based on the notion of RE-formula (which is very close to the more standard notion of Σ_1-formula). I then prove a generalization of the so-called Σ_1-completeness of minimal arithmetic, which is an effective tool to derive representability conditions and serves as a substitute for the Hilbert–Bernays method of formalizing primitive recursive definitions (primitive recursive functions are not discussed in this text). The reader is warned that several notions introduced at this point, such as that of recursive extension, may not be equivalent to those defined in other texts.

In the rest of this chapter, Peano arithmetic, PA, is discussed. Beside basic number-theoretic results, sequences and definitions by recursions are developped in PA, paralleling the number-theoretic developments of Chapter III.

V Arithmetical Theories

The main goal of this chapter is to arrive at a precise and appropriately general statement of the result known as "Gödel's second incompleteness theorem". I call it instead the "theorem on consistency proofs", following Shoenfield. To my knowledge only two published texts contain a proof of some form of this result: Hilbert and Bernay's *Grundlagen der Mathematik* of 1934 and Volume 1 of Tourlakis' *Lectures in Logic and Set Theory* of 2003. The version given here is more general than either of them.

The chapter starts with a discussion of the derivability conditions formulated by Löb to obtain a criterion on first-order theories subject to the second incompleteness theorem. With this general result in mind, the notions of arithmetical language, theory, and interpretation are introduced as formalizations within PA of the corresponding metamathematical notions. I define what it means for such an arithmetical object to describe a first-order language, theory, or interpretation. The main result is then that an arithmetical theory that describes a first-order theory T provides a provability predicate for T that satisfies the derivability conditions, and the theorem on consistency proofs is an immediate corollary. Another less well-known result that is discussed is the arithmetical completeness theorem, which is a formalization within PA of the arguments in the proof of the classical completeness theorem. It constructs an interpretation of any reasonable first-order theory T in Peano arithmetic, supplemented with a suitable axiom expressing the consistency of T. The chapter ends with an application of the theorem on consistency proofs to Zermelo–Fraenkel set theory.

VI First-Order Set Theory

This chapter develops the basics of Zermelo–Fraenkel set theory in a standard way. It is heavily inspired by Shoenfield's Chapter 9.

VII The Consistency Proofs

In this chapter it is proved that the axiom of choice and the generalized continuum hypothesis are consistent with ZF. This is done as usual in two steps: firstly an extension ZFL of ZF is constructed together with an interpretation of ZFL in ZF and secondly it is proved that the axiom of choice and the generalized continuum hypothesis are theorems of ZFL. Most of this chapter is again inspired by Shoenfield's Chapter 9. In particular, the predicate of constructibility is defined as Gödel originally did and not using the notion of definable subset: although the latter is closer to our intuition, perhaps even dangerously so, it also requires much more work. A notable difference from Shoenfield's treatment is the internal cardinality theorem which is here proved in ZFC and not just in ZFL: it says that one can conservatively add a "model" of ZFC within itself of arbitrary infinite cardinality.

Chapter One
First-Order Theories

§1 Formal systems

1.1 Sequences. The notion of *sequence* is omnipresent in the study of formal systems, so we establish some terminology about sequences. A sequence determines and is determined by the following data: a natural number n called the *length* of the sequence, and for each natural number i with $1 \leq i \leq n$, an object called the ith *member* of the sequence. In particular, there is exactly one sequence of length 0, called the *empty sequence*. A sequence shall not be a collection. Sequences $\mathbf{u}_1, \ldots, \mathbf{u}_n$ may be *concatenated* to yield a new sequence which we denote by $\mathbf{u}_1\mathbf{u}_2 \ldots \mathbf{u}_n$. An *occurrence* in a sequence $\mathbf{u}$, written $\langle \mathbf{u}_1, \mathbf{u}_2 \rangle$, consists of two sequences $\mathbf{u}_1$ and $\mathbf{u}_2$ such that $\mathbf{u}$ is $\mathbf{u}_1\mathbf{v}\mathbf{u}_2$ for some sequence $\mathbf{v}$; the sequence $\mathbf{v}$ is then uniquely determined, and we say that $\langle \mathbf{u}_1, \mathbf{u}_2 \rangle$ is an occurrence *of* $\mathbf{v}$ in $\mathbf{u}$. A *subsequence* of $\mathbf{u}$ is a sequence of which there is an occurrence in $\mathbf{u}$. Let $\langle \mathbf{u}_1, \mathbf{u}_2 \rangle$ and $\langle \mathbf{u}_1', \mathbf{u}_2' \rangle$ be two occurrences in $\mathbf{u}$. We say that $\langle \mathbf{u}_1, \mathbf{u}_2 \rangle$ *happens within* $\langle \mathbf{u}_1', \mathbf{u}_2' \rangle$ if $\mathbf{u}_1'$ is a subsequence of $\mathbf{u}_1$ and $\mathbf{u}_2'$ is a subsequence of $\mathbf{u}_2$. If an occurrence of $\mathbf{v}_1$ happens within an occurrence of $\mathbf{v}_2$, then $\mathbf{v}_1$ is a subsequence of $\mathbf{v}_2$. The sequence obtained from $\mathbf{u}$ *by replacing the occurrence* $\langle \mathbf{u}_1, \mathbf{u}_2 \rangle$ *by* $\mathbf{v}$ is defined to be the sequence $\mathbf{u}_1\mathbf{v}\mathbf{u}_2$.

An *appearance* of an object will be synonymous with an occurrence of the sequence of length 1 whose member is that object.

1.2 Languages. A sequence of length 1 whose only member is not a sequence will be called a *symbol*. An *alphabet* is a collection of symbols. The *expressions* of an alphabet A are defined inductively by the following clauses: the empty sequence is an expression of A; if $\mathbf{u}$ is an expression of A and $\mathbf{s}$ is a symbol of A, then $\mathbf{su}$ is an expression of A. From now on we use boldface letters exclusively to denote expressions of some alphabet.

A *language* L consists of an alphabet A together with a collection of expressions of A, called the *formulae* of L. Symbols and expressions of A are also called symbols and expressions of L. We let $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ vary through formulae. (By this we mean: from now on, the letters $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$, possibly decorated with subscripts or superscripts, will be used exclusively to denote formulae of some language, namely the language being discussed.) A *subformula* of a formula $\mathbf{A}$ of L is a subsequence of $\mathbf{A}$ which is also a formula of L.

1.3 Formal systems. Let L be a language. A *rule of inference* for L consists of a finite collection of formulae of L called the *premises* of the rule, and a single formula of L called the *conclusion* of the rule. A rule of inference with no premises is also called an *axiom*, and we identify such a rule with its conclusion. Thus any formula of L is a rule of inference for L.

A *formal system* F consists of a language $L(F)$ together with a collection of rules of inference for $L(F)$, called the rules of inference *of* F. Symbols, expressions, and formulae of $L(F)$ are also called symbols, expressions, and formulae of F. Let $\mathbf{A}$ be a formula of $L(F)$ and Γ a collection of formulae of $L(F)$. A *derivation of* $\mathbf{A}$ *from* Γ *in* F is a sequence of formulae of $L(F)$ ending with $\mathbf{A}$, each of whose member is either in Γ or the conclusion of a rule of inference of F whose premises appear previously in the sequence. We say that $\mathbf{A}$ is *derivable* or *inferrable* from Γ in F, and we write $\Gamma \vdash_F \mathbf{A}$, when there exists a derivation of $\mathbf{A}$ from Γ in F. A rule of inference for $L(F)$ is said to be derivable in F if its conclusion is derivable in F from its premises.

Proposition. If $\Gamma \vdash_F \mathbf{A}$ and if $\Delta \vdash_F \mathbf{B}$ for any formula $\mathbf{B}$ of Γ, then $\Delta \vdash_F \mathbf{A}$.

Proof. A derivation of $\mathbf{A}$ from Δ in F is obtained by replacing in a derivation of $\mathbf{A}$ from Γ in F any appearance of a formula $\mathbf{B}$ of Γ by a derivation of $\mathbf{B}$ from Δ in F. □

If Γ consists of $\mathbf{B}_1, \ldots, \mathbf{B}_n$, we also write $\mathbf{B}_1, \ldots, \mathbf{B}_n \vdash_F \mathbf{A}$ instead of $\Gamma \vdash_F \mathbf{A}$. We write $\mathbf{A} \equiv_F \mathbf{B}$ when $\mathbf{A} \vdash_F \mathbf{B}$ and $\mathbf{B} \vdash_F \mathbf{A}$. When Γ is empty, we simply remove the reference to Γ in the above definitions and we write $\vdash_F \mathbf{A}$ instead of $\Gamma \vdash_F \mathbf{A}$. A *theorem* of F is a formula $\mathbf{A}$ of F such that $\vdash_F \mathbf{A}$, i.e., a formula of which there is a derivation in F. Note that our convention of identifying formulae with rules of inference is compatible with this terminology: a formula is derivable in F if and only if it is derivable in F as a rule of inference.

1.4 Extensions. Let L be a language. A language L' is an *extension* of L if the alphabet of L' includes the alphabet of L and if every formula of L is a formula of L'.

Let F be a formal system. A formal system F' is an *extension* of F if $L(F')$ is an extension of $L(F)$ and if every theorem of F is a theorem of F'. It is a *conservative extension* if moreover any formula of $L(F)$ which is a theorem of F' is a theorem of F. A *simple extension* of a formal system F is an extension of F whose language is $L(F)$. Two formal systems are *equivalent* if they are extensions of each other; this is easily seen to be the case if and only if one is a simple conservative extension of the other. It follows immediatly from these definitions that if F'' is a (conservative; simple) extension of F' and F' is a (conservative; simple) extension of F, then F'' is a (conservative; simple) extension of F.

If R is a collection of rules of inference for $L(F)$, we denote by $F[R]$ the formal system whose language is $L(F)$ and whose rules of inference are those of F and those in R. Observe that $F[R]$ is a simple extension of F. With our identification of formulae with rules of inference, we see that $\Gamma \vdash_F \mathbf{A}$ if and only if $\vdash_{F[\Gamma]} \mathbf{A}$. In this way the notion of formula derivable from Γ is reduced to the notion of theorem. Observe that if F' is an extension of F and if R is a collection of rules of inference for $L(F)$, then $F'[R]$ need not be an extension of $F[R]$. This is the case if however F' is of the form $F[R']$.

1.5 Induction on theorems. To prove that all the theorems of a formal system F have a given property, it suffices to prove that whenever that property holds for the premises of a rule of inference of F (an assumption called the *induction hypothesis*), it holds for the conclusion of the rule as well; this ensures that all formulae appearing in some derivation in F have the given property. Such a proof will be called a proof by *induction on theorems in F*.

1.6 Convention. The remaining sections of this chapter are mostly devoted to proving assertions of the form $\Gamma \vdash_F \mathbf{A}$. To gain space, in the proof of such an assertion, we shall use the sign $\vdash_F$ as an abbreviation for $\Gamma \vdash_F$ (unless of course the context clearly indicates otherwise). See §3.2 for examples of applications of this convention.

§2 First-order languages and theories

2.1 Logical symbols. We choose once and for all an infinite collection of symbols,

$$x, y, z, w, x', y', z', w', x'', \dots$$

which we call the *variables*. The order in which they are listed above is called the *alphabetical order*. We also choose four distinct symbols written $\vee$, $\neg$, $\exists$, and $=$ that are not variables. The variables and the symbols $\vee$, $\neg$, $\exists$, and $=$ are called the *logical symbols*. The symbol $=$ is called the *equality symbol*. (We also use the sign $=$ in the usual way, as in $a = b$ to signify that a and b are the same object. The meaning of the sign $=$ will however always be clear from the context, just as it is clear that the sign y in *symbol* does not stand for a variable.) We let $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$, and $\mathbf{w}$ vary through variables.

Instead of choosing infinitely many symbols as variables, one can also use two symbols, say x and $'$, and define the variables by induction to be the expressions x, $'x$, $''x$, etc. This is, however, merely a cosmetic variation of the above definition.

2.2 Signatures. Suppose given, for each natural number n, a collection of symbols called *n-ary function symbols* and a collection of symbols called *n-ary predicate symbols*. Assume that the following conditions are satisfied: a function symbol is not a predicate symbol; the arity of a function or predicate symbol is uniquely determined; the symbol $=$ is a binary predicate symbol, and other logical symbols are neither function symbols nor predicate symbols. This data is then said to define a *signature* S. A 0-ary predicate symbol is called a *truth value*, and a 0-ary function symbol is called a *constant*. We let $\mathbf{p}$, $\mathbf{q}$, and $\mathbf{r}$ vary through predicate symbols, $\mathbf{f}$, $\mathbf{g}$, and $\mathbf{h}$ through function symbols, and $\mathbf{e}$ through constants.

A signature S has an underlying alphabet $A(S)$ consisting of the logical symbols, the function symbols of S, and the predicate symbols of S. To every symbol of $A(S)$ we associate an *index* as follows:

(i) variables are symbols of index 0;

(ii) $\vee$ and $\exists$ have index 2, and $\neg$ has index 1;

(iii) an n-ary function symbol has index n;

(iv) an n-ary predicate symbol has index n.

2.3 First-order languages. Let S be a signature. We shall associate to it a language $L(S)$ with alphabet $A(S)$. We first define the *terms* of S inductively as follows:

(i) a variable is a term;

(ii) if $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are terms and $\mathbf{f}$ is an n-ary function symbol, then $\mathbf{f}\mathbf{a}_1 \ldots \mathbf{a}_n$ is a term.

The *atomic formulae* of S are defined as follows:

(iii) if $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are terms and $\mathbf{p}$ is an n-ary predicate symbol, then $\mathbf{p}\mathbf{a}_1 \ldots \mathbf{a}_n$ is an atomic formula.

The formulae of $L(S)$ are then defined inductively by the following clauses:

(iv) an atomic formula is a formula;

(v) if $\mathbf{A}$ and $\mathbf{B}$ are formulae, $\vee\mathbf{AB}$ is a formula;

(vi) if $\mathbf{A}$ is a formula, $\neg\mathbf{A}$ is a formula;

(vii) if $\mathbf{A}$ is a formula and $\mathbf{x}$ is a variable, $\exists\mathbf{x}\mathbf{A}$ is a formula.

A language L is a *first-order language* if it is of the form $L(S)$ for some signature S. We observe that L determines S, for a nonlogical symbol $\mathbf{s}$ of L is an n-ary function symbol of S (resp. an n-ary predicate symbol of S) if and only if the expression $=y\mathbf{s}x \ldots x$ (resp. $\mathbf{s}x \ldots x$) with n occurrences of x is a formula of L. Thus we may speak of the n-ary function symbols of L, the n-ary predicate symbols of L, the index of a symbol of L, the terms of L, and the atomic formulae of L. We can prove in the same way that given two signatures S and S', $L(S')$ is an extension of $L(S)$ if and only if the n-ary function symbols of S are n-ary function symbols of S' and the n-ary predicate symbols of S are n-ary predicate symbols of S'.

It should be noted that the terms (resp. the formulae) of a first-order language are the *theorems* of the formal system whose formulae are all expressions and whose rules of inference are defined by (i)–(ii) (resp. by (iv)–(vii)). An important property of those two formal systems is that they have an obvious *decision method*, i.e., an algorithm that will decide whether a formula is a theorem or not. This is not the case in general for an arbitrary formal system, not even for the very particular formal systems that we intend to study.

2.4 Designators. Let L be a first-order language. A *designator* of L is either a term of L or a formula of L. As we pointed out in §2.3, the designators are the theorems of some formal system. This formal system has two very pleasant properties: any designator has an "essentially unique" derivation in this formal system, and a designator occurs in a designator $\mathbf{u}$ if and only if it appears in any derivation of $\mathbf{u}$. This is the content of the two theorems of this paragraph. These are simple properties of the so-called Polish notation, and we omit the proofs. By definition, any designator has the form $\mathbf{s}\mathbf{u}_1 \ldots \mathbf{u}_n$ where $\mathbf{s}$ is a symbol of index n and $\mathbf{u}_1, \ldots, \mathbf{u}_n$ are designators.

FORMATION THEOREM. Let $\mathbf{v}$ be a designator of L. If $\mathbf{v}$ can be written as $\mathbf{s}\mathbf{u}_1 \ldots \mathbf{u}_n$ and as $\mathbf{s}'\mathbf{u}_1' \ldots \mathbf{u}_m'$, where $\mathbf{s}$ is a symbol of index n, $\mathbf{s}'$ is a symbol of index m, and $\mathbf{u}_1, \ldots, \mathbf{u}_n$, $\mathbf{u}_1', \ldots, \mathbf{u}_m'$ are designators, then $n = m$, $\mathbf{s}$ is $\mathbf{s}'$, $\mathbf{u}_1$ is $\mathbf{u}_1', \ldots, \mathbf{u}_n$ is $\mathbf{u}_n'$.

OCCURRENCE THEOREM. Let $\mathbf{s}\mathbf{u}_1 \ldots \mathbf{u}_n$ and $\mathbf{v}$ be distinct designators where $\mathbf{s}$ is a symbol of index n and $\mathbf{u}_1, \ldots, \mathbf{u}_n$ are designators. Any occurrence of $\mathbf{v}$ in $\mathbf{u}$ happens within an occurrence of the form $\langle \mathbf{s}\mathbf{u}_1 \ldots \mathbf{u}_{i-1}, \mathbf{u}_{i+1} \ldots \mathbf{u}_n \rangle$ for some i.

The formation theorem implies that a term is either a variable or can be written in only one way as $\mathbf{f}\mathbf{a}_1 \ldots \mathbf{a}_n$ where $\mathbf{f}$ is n-ary and $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are terms. Similarly, a formula is either atomic or can be written in only one way as $\vee\mathbf{AB}$, $\neg\mathbf{A}$, or $\exists\mathbf{x}\mathbf{A}$. The occurrence theorem, applied to formulae, says that an occurrence of a formula in $\vee\mathbf{AB}$ (resp. $\neg\mathbf{A}$; $\exists\mathbf{x}\mathbf{A}$) that is not the whole formula must happen within either $\langle \vee, \mathbf{B} \rangle$ or $\langle \vee\mathbf{A}, \rangle$ (resp. within $\langle \neg, \rangle$; within $\langle \exists, \mathbf{A} \rangle$ or $\langle \exists\mathbf{x}, \rangle$). These consequences of the formation and occurrence theorems will most often be used tacitly.

2.5 Abbreviations. We introduce some *abbreviations* for first-order languages:

(i) x_1, x_2, x_3, etc. abbreviate x', x'', x''', etc., and similarly for y, z, and w;

(ii) x^n abbreviates $x_1 \dots x_n$, and similarly for y, z, and w (of course, x^0 will be the empty sequence);

(iii) $(\mathbf{A} \vee \mathbf{B})$ abbreviates $\vee\mathbf{AB}$;

(iv) $(\mathbf{A} \wedge \mathbf{B})$ abbreviates $\neg(\neg\mathbf{A} \vee \neg\mathbf{B})$;

(v) $(\mathbf{A} \to \mathbf{B})$ abbreviates $(\neg\mathbf{A} \vee \mathbf{B})$;

(vi) $(\mathbf{A} \leftrightarrow \mathbf{B})$ abbreviates $((\mathbf{A} \to \mathbf{B}) \wedge (\mathbf{B} \to \mathbf{A}))$;

(vii) $\forall\mathbf{x}$ abbreviates $\neg\exists\mathbf{x}\neg$.

This is to be understood as follows: whenever one of the symbols above appears, it must be expanded using its definition in order to recover the actual formula. For example, $=xy$ is a formula of any first-order language, but $(=xy \wedge =xy)$ is not: it only abbreviates an actual formula, namely $\neg\vee\neg=xy\neg=xy$.

At any given time we allow ourselves to introduce new abbreviations, either for any first-order language or for a specific one. This is in fact absolutely necessary, for otherwise we would be overwhelmed by the length and complexity of expressions.

(viii) If $\mathbf{p}$ is binary, then $(\mathbf{apb})$ abbreviates $\mathbf{pab}$;

(ix) $(\mathbf{a} \neq \mathbf{b})$ abbreviates $\neg(\mathbf{a} = \mathbf{b})$.

We also drop the parentheses in the above abbreviations whenever there can be no confusion about the intended meaning. To be able to drop even more parentheses, we use the following convention: when given a choice, a formula shall be of the form $\mathbf{A} \to \mathbf{B}$ or $\mathbf{A} \leftrightarrow \mathbf{B}$ rather than of the form $\mathbf{A} \vee \mathbf{B}$ or $\mathbf{A} \wedge \mathbf{B}$, and it shall be of any of those four forms rather than of the form $\exists\mathbf{xA}$ or $\forall\mathbf{xA}$. Other ambiguous cases are settled using association from the right. Thus: $\mathbf{A} \vee \mathbf{B} \to \mathbf{C} \wedge \mathbf{D}$ is to be read as $((\mathbf{A} \vee \mathbf{B}) \to (\mathbf{C} \wedge \mathbf{D}))$, $\mathbf{A} \to \mathbf{B} \leftrightarrow \mathbf{C}$ as $(\mathbf{A} \to (\mathbf{B} \leftrightarrow \mathbf{C}))$, $\exists\mathbf{xA} \to \mathbf{B}$ as $(\exists\mathbf{xA} \to \mathbf{B})$ and not $\exists\mathbf{x}(\mathbf{A} \to \mathbf{B})$, etc.

Another use of parentheses is the following. Suppose for instance that to each formula $\mathbf{A}$ we have associated two formulae which we decided to denote by $\mathbf{A}^*$ and $\mathbf{A}_*$, then we may write $(\neg\mathbf{A})^*$, $(\mathbf{A}_1)_*$, $(\mathbf{A}^*)_*$, etc.

2.6 Some terminology. Let L be a first-order language. It is often useful to have names for some formulae of L. The formula $\mathbf{A}_1 \vee \cdots \vee \mathbf{A}_n$ is called the *disjunction* of $\mathbf{A}_1, \dots, \mathbf{A}_n$; $\mathbf{A}_1 \wedge \cdots \wedge \mathbf{A}_n$ the *conjunction* of $\mathbf{A}_1, \dots, \mathbf{A}_n$; $\neg\mathbf{A}$ the *negation* of $\mathbf{A}$; $\mathbf{A}_1 \to \cdots \to \mathbf{A}_n \to \mathbf{B}$ the *implication* of $\mathbf{B}$ by $\mathbf{A}_1, \dots, \mathbf{A}_n$; $\mathbf{A} \leftrightarrow \mathbf{B}$ the *equivalence* of $\mathbf{A}$ and $\mathbf{B}$; $\exists\mathbf{xA}$ the *instantiation* of $\mathbf{A}$ by $\mathbf{x}$; $\forall\mathbf{xA}$ the *generalization* of $\mathbf{A}$ by $\mathbf{x}$.

An occurrence $\langle\mathbf{u}_1, \mathbf{u}_2\rangle$ of a variable $\mathbf{x}$ in a designator $\mathbf{u}$ is *not meaningful* if $\mathbf{u}_1$ is of the form $\mathbf{v}\exists$; *bound* if it happens within an occurrence of an instantiation; *free* if it is not bound. The variable $\mathbf{x}$ itself is said to be *bound* (resp. *free*) in $\mathbf{u}$ if *some* occurrence of $\mathbf{x}$ is bound (resp. free) in $\mathbf{u}$. We say that a designator is *closed* if no variable is free in it; *open* if $\exists$ does not occur in it. A formula which is either atomic or an instantiation is called *elementary*. The *closure* of $\mathbf{A}$ is the formula $\forall\mathbf{x}_1 \dots \forall\mathbf{x}_n\mathbf{A}$ where $\mathbf{x}_1, \dots, \mathbf{x}_n$ are the variables free in $\mathbf{A}$ in reverse alphabetical order.

2.7 Substitution. Let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be variables and $\mathbf{a}_1, \dots, \mathbf{a}_n$ terms of a first-order language L such that whenever $\mathbf{a}_i$ and $\mathbf{a}_j$ are distinct, $\mathbf{x}_i$ and $\mathbf{x}_j$ are distinct. We let $\mathbf{b}[\mathbf{x}_1, \dots, \mathbf{x}_n|\mathbf{a}_1, \dots, \mathbf{a}_n]$ abbreviate the term obtained from $\mathbf{b}$ by replacing each occurrence of the variable $\mathbf{x}_i$ by the term $\mathbf{a}_i$, for all i simultaneously. Similarly, we let $\mathbf{A}[\mathbf{x}_1, \dots, \mathbf{x}_n|\mathbf{a}_1, \dots, \mathbf{a}_n]$ abbreviate the formula obtained from $\mathbf{A}$ by replacing each *free* occurrence of the variable $\mathbf{x}_i$ by the term $\mathbf{a}_i$, for all i simultaneously.† We say that $\mathbf{a}$ is *substitutible* for $\mathbf{x}$ in $\mathbf{A}$ if for any variable $\mathbf{y}$ that occurs in $\mathbf{a}$, no occurrence of $\mathbf{x}$ in $\mathbf{A}$ happening within an occurrence of $\exists\mathbf{yB}$ is free in $\mathbf{A}$. *We restrict the use of the abbreviation* $\mathbf{A}[\mathbf{x}_1, \dots, \mathbf{x}_n|\mathbf{a}_1, \dots, \mathbf{a}_n]$ *to those* $\mathbf{A}$, $\mathbf{x}_1, \dots, \mathbf{x}_n$, $\mathbf{a}_1, \dots, \mathbf{a}_n$ *such that* $\mathbf{a}_i$ *is substitutible for* $\mathbf{x}_i$ *in* $\mathbf{A}$, *for all* i. This ensures that any occurrence of a variable in $\mathbf{a}_i$ does not become bound in $\mathbf{A}[\mathbf{x}_1, \dots, \mathbf{x}_n|\mathbf{a}_1, \dots, \mathbf{a}_n]$. With this restriction, if $\mathbf{y}_1, \dots, \mathbf{y}_n$ are not free in $\mathbf{A}$, then $\mathbf{A}[\mathbf{x}_1, \dots, \mathbf{x}_n|\mathbf{y}_1, \dots, \mathbf{y}_n][\mathbf{y}_1, \dots, \mathbf{y}_n|\mathbf{a}_1, \dots, \mathbf{a}_n]$ is the same as $\mathbf{A}[\mathbf{x}_1, \dots, \mathbf{x}_n|\mathbf{a}_1, \dots, \mathbf{a}_n]$. Observe that a closed term is substitutible for any variable in any formula.

Let $\mathbf{A}$ be a formula. A *variant* of $\mathbf{A}$ is a formula obtained from $\mathbf{A}$ by repeated replacements of occurrences of subformulae of the form $\exists\mathbf{xB}$ by $\exists\mathbf{yB}[\mathbf{x}|\mathbf{y}]$ for some $\mathbf{y}$ not free in $\mathbf{B}$. An *instance* of $\mathbf{A}$ is a formula of the form $\mathbf{A}[\mathbf{x}_1, \dots, \mathbf{x}_n|\mathbf{a}_1, \dots, \mathbf{a}_n]$. A *version* of $\mathbf{A}$ is an instance of a variant of $\mathbf{A}$. Note that an instance of

†We should here verify, using induction on the lengths of $\mathbf{b}$ and $\mathbf{A}$, that $\mathbf{b}[\mathbf{x}_1, \dots, \mathbf{x}_n|\mathbf{a}_1, \dots, \mathbf{a}_n]$ is a term and that $\mathbf{A}[\mathbf{x}_1, \dots, \mathbf{x}_n|\mathbf{a}_1, \dots, \mathbf{a}_n]$ is a formula. Usually, verifications of this kind will be entirely left out.

an instance of **A** is an instance of **A** and that any formula obtained from **A** by taking successively variants and instances is a version of **A**.

2.8 The standard meaning. We make a few remarks on the intended meaning of the symbols and formulae of a first-order language. These remarks are not required for the formal exposition of first-order theories but are important nonetheless. The terms of a first-order language are meant to represent the individuals whose behaviour we intend to formalize—it must be assumed that there is at least one such individual—, and the formulae represent propositions about those individuals. The predicate and function symbols are of course meant to represent predicates and functions, so that $\mathbf{pa}_1 \dots \mathbf{a}_n$ means "$\mathbf{a}_1, \dots, \mathbf{a}_n$ together have the predicate **p**", and $\mathbf{fa}_1 \dots \mathbf{a}_n$ represents the individual that is the image of the individuals $\mathbf{a}_1, \dots, \mathbf{a}_n$ by the function **f**. The predicate symbol = is a symbol for identity of individuals. The meanings of closed formulae $\vee\mathbf{AB}$, $\neg\mathbf{A}$, and $\exists\mathbf{xA}$ are respectively "**A** is true or **B** is true", "**A** is false", and "for some **x**, **A** is true". If a formula is not closed, then its meaning is that of its closure. This is the *standard meaning* of a first-order language, and it is used, in the informal exposition, to translate back and forth between English and the first-order language. However, other meanings are possible: a first-order language is wholly independant from the meaning we have in mind for it.

We then see that the standard meaning of the abbreviations $\mathbf{A} \wedge \mathbf{B}$, $\mathbf{A} \to \mathbf{B}$, $\mathbf{A} \leftrightarrow \mathbf{B}$, and $\forall\mathbf{xA}$ is as expected. Our choice of the "primitive symbols" $\vee$, $\neg$, and $\exists$ rather than, say, $\wedge$, $\neg$, and $\forall$ is completely arbitrary. However, it is worthwhile to note that it would have been possible to use only one symbol instead of $\vee$ and $\neg$: for example a symbol whose meaning, when applied to **A** and **B**, is "both **A** and **B** are false".

2.9 First-order theories. Let L be a first-order language. The following rules of inference are called the *logical rules* for L:

(i) infer $\neg\mathbf{A} \vee \mathbf{A}$ (*propositional axioms*);
(ii) infer **A** from $\mathbf{A} \vee \mathbf{A}$ (*contraction rules*);
(iii) infer $\mathbf{B} \vee \mathbf{A}$ from **A** (*expansion rules*);
(iv) infer $(\mathbf{A} \vee \mathbf{B}) \vee \mathbf{C}$ from $\mathbf{A} \vee \mathbf{B} \vee \mathbf{C}$ (*associativity rules*);
(v) infer $\mathbf{B} \vee \mathbf{C}$ from $\mathbf{A} \vee \mathbf{B}$ and $\neg\mathbf{A} \vee \mathbf{C}$ (*cut rules*);
(vi) infer $\mathbf{A}[\mathbf{x}|\mathbf{a}] \to \exists\mathbf{xA}$ (*substitution axioms*);
(vii) if **x** is not free in **B**, infer $\exists\mathbf{xA} \to \mathbf{B}$ from $\mathbf{A} \to \mathbf{B}$ (*∃-introduction rules*);
(viii) infer $\mathbf{x} = \mathbf{x}$ (*identity axioms*);
(ix) infer $\mathbf{x}_1 = \mathbf{y}_1 \to \cdots \to \mathbf{x}_n = \mathbf{y}_n \to \mathbf{fx}_1 \dots \mathbf{x}_n = \mathbf{fy}_1 \dots \mathbf{y}_n$ (*functional equality axioms*);
(x) infer $\mathbf{x}_1 = \mathbf{y}_1 \to \cdots \to \mathbf{x}_n = \mathbf{y}_n \to \mathbf{px}_1 \dots \mathbf{x}_n \to \mathbf{py}_1 \dots \mathbf{y}_n$ (*predicative equality axioms*).

Other rules of inference for L are called *nonlogical rules*. The rules (i)–(v) are called the *propositional rules*, the rules (vi)–(vii) the *quantification rules*, and the rules (viii)–(x) the *equality rules*. In what follows we shall often refer to one of (i)–(x) as a *rule* in the singular, even though each of them is an infinite collection of rules of inferences.

A formal system T is called a *first-order theory* if its language $L(T)$ is a first-order language, if its rules of inference include the logical rules for $L(T)$, and if its nonlogical rules are axioms. This last restriction deserves an explanation. It is easily seen that most of the results of this chapter are true even if we allow any nonlogical rules. But this is not the case of many fundamental results of Chapter II. For instance, we shall prove in ch. II §1.1 that if **A** is a theorem of $T[\neg\mathbf{A}]$, where T is a first-order theory and **A** is a closed formula of T, then **A** is a theorem of T. This simple result can be false if T is allowed to have arbitrary nonlogical rules. Indeed, we shall discuss in Chapter III a first-order theory N in which there is a closed formula **A** such that neither **A** nor $\neg\mathbf{A}$ is a theorem of N. If F is the formal system obtained from N by adding the rule of inference "infer **A** from $\neg\mathbf{A}$", then the sequence $\neg\mathbf{A}$, **A** is a derivation of **A** in $F[\neg\mathbf{A}]$, but **A** is not a theorem of F.

We must also comment briefly on the equality rules. Unsurprisingly, all the general results on first-order theories which do not deal explicitely with the equality symbol = are in fact true if we do not require the presence of the equality symbol nor of the equality rules. The systematic inclusion of the equality rules has the consequence that, if we want to obtain a first-order theory from a given first-order theory by extending its signature, i.e., by adding nonlogical symbols, then we have to add all the equality rules featuring the new symbols as well. This may seem like a trivial variation, but in fact it introduces considerable difficulties, as we shall see in ch. II §4. There are a number of reasons for us not to consider the more general situation of first-order theories "without equality". The first is that *all* the proofs of these more general results are contained in the proofs given here. The second is that it is tedious to deal with various sets of hypotheses all the time. Finally, first-order theories without equality rarely occur in practice.

Let T and T' be first-order theories such that $L(T')$ is an extension of $L(T)$. For T' to be an extension of T, it is obviously necessary that every nonlogical axiom of T be a theorem of T'. This is also sufficient, for if Γ is the collection of nonlogical axioms of T, a derivation of $\mathbf{A}$ in T is a derivation of $\mathbf{A}$ from Γ in T', so our claim follows from the proposition of §1.3.

§3 Tautologies

3.1 Truth valuations. Let L be a first-order language. We say that a *truth valuation* V on L has been given if to each elementary formula of L is associated one of the two symbols T and F. We denote by $V(\mathbf{A})$ the symbol that V assigns to an elementary formula $\mathbf{A}$. We want to extend the domain of truth valuations to all formulae. To do this we define the mapping $f_\vee$ by

$$f_\vee(\mathsf{T},\mathsf{T}) = \mathsf{T},\ f_\vee(\mathsf{T},\mathsf{F}) = \mathsf{T},\ f_\vee(\mathsf{F},\mathsf{T}) = \mathsf{T}, \text{ and } f_\vee(\mathsf{F},\mathsf{F}) = \mathsf{F},$$

and the mapping $f_\neg$ by

$$f_\neg(\mathsf{T}) = \mathsf{F} \text{ and } f_\neg(\mathsf{F}) = \mathsf{T}.$$

Given a truth valuation V on L and a formula $\mathbf{A}$ of L, we define $V(\mathbf{A})$ by induction on the length of $\mathbf{A}$ as follows:

(i) if $\mathbf{A}$ is elementary, then $V(\mathbf{A})$ is already defined;
(ii) if $\mathbf{A}$ is $\mathbf{B} \vee \mathbf{C}$, then $V(\mathbf{A})$ is $f_\vee(V(\mathbf{B}), V(\mathbf{C}))$;
(iii) if $\mathbf{A}$ is $\neg\mathbf{B}$, then $V(\mathbf{A})$ is $f_\neg(V(\mathbf{B}))$.

Let Γ be a *finite* collection of formulae of L. A formula $\mathbf{A}$ is a *tautological consequence* of Γ if any truth valuation which assigns T to the formulae of Γ assigns T to $\mathbf{A}$. If the latter holds when Γ is empty, we say that $\mathbf{A}$ is a *tautology*. Two formulae are *tautologically equivalent* if they are tautological consequences of one another. It is an easy exercise to show that $\mathbf{A}$ is a tautological consequence of $\mathbf{B}_1, \dots, \mathbf{B}_n$ if and only if $\mathbf{B}_1 \to \cdots \to \mathbf{B}_n \to \mathbf{A}$ is a tautology.

It is not at once clear that the notions of tautology and tautological consequence are finitary, since there is an infinite number of elementary formulae and hence an infinite number of possible truth valuations to consider. This is not actually so, for in a given formula $\mathbf{A}$ there are only finitely many occurrences of elementary subformulae, and hence all the possibilities of assignment may be checked in a finite number of steps in order to decide whether $\mathbf{A}$ is a tautology or not. This also shows that the notion of tautology does not depend on L, but only on the expression $\mathbf{A}$ itself. These remarks also apply to the notion of tautological consequence that can be seen as a particular case of the notion of tautology.

Obviously, given the definition of a truth valuation, the meaning of a tautology is true. In fact, a tautology is a formula that can be seen to be true using only the meanings of $\neg$ and $\vee$. However, the notion of tautology does not take into account the meanings of $\exists$ and $=$. In this section we shall prove that the logical axioms and rules of a first-order theory are sufficiently strong to derive any tautology.

3.2 The tautology theorem. In this paragraph, we fix a first-order language L and we let F be the formal system with language L whose rules of inference are the propositional rules for L. It follows from the definition of a first-order theory and some remarks in §1.4 that for any first-order theory T with language L, if $\Gamma \vdash_F \mathbf{A}$, then $\Gamma \vdash_T \mathbf{A}$. We write $\vdash$ and $\equiv$ instead of $\vdash_F$ and $\equiv_F$.

COMMUTATIVITY RULE. $\mathbf{A} \vee \mathbf{B} \vdash \mathbf{B} \vee \mathbf{A}$.

Proof. Apply the cut rule to $\mathbf{A} \vee \mathbf{B}$ and $\neg\mathbf{A} \vee \mathbf{A}$ which is a propositional axiom. ☐

DETACHMENT RULE. $\mathbf{A} \to \mathbf{B}, \mathbf{A} \vdash \mathbf{B}$.

Proof. We have $\vdash \mathbf{A} \vee \mathbf{B}$ by the expansion rule and the commutativity rule. Applying the cut rule to $\mathbf{A} \vee \mathbf{B}$ and $\mathbf{A} \to \mathbf{B}$, we get $\vdash \mathbf{B} \vee \mathbf{B}$, whence $\vdash \mathbf{B}$ by the contraction rule. ☐

The deduction theorem that we shall prove in §4.3 is a reciprocal to the detachment rule.

Lemma 1. Let $\mathbf{A}_1, \dots, \mathbf{A}_n$ be formulae, each $\mathbf{A}_i$ being either elementary or the negation of an elementary formula. The formula $\mathbf{A}_1 \vee \cdots \vee \mathbf{A}_n$ is a tautology if and only if some $\mathbf{A}_i$ is the negation of some $\mathbf{A}_j$.

Proof. The sufficiency is obvious. Suppose now that there are no i and j such that $\mathbf{A}_i$ is the negation of $\mathbf{A}_j$. Define a truth valuation V by letting $V(\mathbf{A})$ be T if and only if $\neg\mathbf{A}$ is some $\mathbf{A}_i$, for elementary $\mathbf{A}$. Then $V(\mathbf{A}_i)$ is F for all i, and from this we easily conclude that $V(\mathbf{A}_1 \vee \cdots \vee \mathbf{A}_n)$ is F. □

Lemma 2. Let $i_1, \dots, i_m$ be natural numbers among $1, \dots, n$. Then $\mathbf{A}_{i_1} \vee \cdots \vee \mathbf{A}_{i_m} \vdash \mathbf{A}_1 \vee \cdots \vee \mathbf{A}_n$.

Proof. Let $\mathbf{A}$ be $\mathbf{A}_1 \vee \cdots \vee \mathbf{A}_n$. We proceed by induction on m. Suppose that $m = 1$. By the expansion rule, $\vdash(\mathbf{A}_{i_1+1} \vee \cdots \vee \mathbf{A}_n) \vee \mathbf{A}_{i_1}$, whence $\vdash \mathbf{A}_{i_1} \vee \cdots \vee \mathbf{A}_n$ by the commutativity rule. Using the expansion rule $i_1 - 1$ more times, we obtain $\vdash\mathbf{A}$.

Suppose that m is 2. If i_1 is i_2, then the contraction rule yields $\vdash\mathbf{A}_{i_1}$, whence $\vdash\mathbf{A}$ by the first case. Suppose that $i_1 < i_2$. We prove the result in this case by induction on n, which is necessarily greater than or equal to 2. If n is exactly 2, then there is nothing to prove. Suppose that $n \geq 3$. We distinguish the following cases:

(i) $i_1 \geq 2$;
(ii) $i_1 = 1$ and $i_2 \geq 3$;
(iii) $i_1 = 1$ and $i_2 = 2$.

In case (i) we have $\vdash\mathbf{A}_2 \vee \cdots \vee \mathbf{A}_n$ by the induction hypothesis, whence $\vdash\mathbf{A}$ by the expansion rule. In case (ii), we have $\vdash\mathbf{A}_1 \vee \mathbf{A}_3 \vee \cdots \vee \mathbf{A}_n$ by the induction hypothesis. By the commutativity rule, the expansion rule, and the associativity rule, we get $\vdash(\mathbf{A}_2 \vee \mathbf{A}_3 \vee \cdots \vee \mathbf{A}_n) \vee \mathbf{A}_1$, whence $\vdash\mathbf{A}$ by the commutativity rule. In case (iii), the hypothesis is $\vdash\mathbf{A}_1 \vee \mathbf{A}_2$. By the expansion rule and the associativity rule, $\vdash((\mathbf{A}_3 \vee \cdots \vee \mathbf{A}_n) \vee \mathbf{A}_1) \vee \mathbf{A}_2$. Using the commutativity rule, the associativity rule, and again the commutativity rule, we obtain $\vdash\mathbf{A}$ as desired. If $i_2 < i_1$, we have $\vdash\mathbf{A}_{i_2} \vee \mathbf{A}_{i_1}$ by the commutativity rule, whence $\vdash\mathbf{A}$ by the case $i_1 < i_2$.

Suppose finally that $m \geq 3$. By the associativity rule, $\vdash(\mathbf{A}_{i_1} \vee \mathbf{A}_{i_2}) \vee \mathbf{A}_{i_3} \vee \cdots \vee \mathbf{A}_{i_m}$. Hence $\vdash(\mathbf{A}_{i_1} \vee \mathbf{A}_{i_2}) \vee \mathbf{A}$ by the induction hypothesis. By the commutativity rule and the associativity rule, $\vdash(\mathbf{A} \vee \mathbf{A}_{i_1}) \vee \mathbf{A}_{i_2}$. Hence $\vdash(\mathbf{A} \vee \mathbf{A}_{i_1}) \vee \mathbf{A}$ by the induction hypothesis. Again by the commutativity rule and the associativity rule, $\vdash(\mathbf{A} \vee \mathbf{A}) \vee \mathbf{A}_{i_1}$. Hence $\vdash(\mathbf{A} \vee \mathbf{A}) \vee \mathbf{A} \vee \mathbf{A}$ by the induction hypothesis. Applying the contraction rule twice, we obtain $\vdash\mathbf{A}$ as desired. □

Lemma 3. $\mathbf{A} \vee \mathbf{B} \vdash \neg\neg\mathbf{A} \vee \mathbf{B}$.

Proof. The formula $\neg\neg\mathbf{A} \vee \neg\mathbf{A}$ is a propositional axiom. Hence $\vdash\neg\mathbf{A} \vee \neg\neg\mathbf{A}$ by the commutativity rule. From $\mathbf{A} \vee \mathbf{B}$ and $\neg\mathbf{A} \vee \neg\neg\mathbf{A}$ we infer $\vdash\mathbf{B} \vee \neg\neg\mathbf{A}$ by the cut rule. Hence $\vdash\neg\neg\mathbf{A} \vee \mathbf{B}$ by the commutativity rule. □

Lemma 4. $\neg\mathbf{A} \vee \mathbf{C}, \neg\mathbf{B} \vee \mathbf{C} \vdash \neg(\mathbf{A} \vee \mathbf{B}) \vee \mathbf{C}$.

Proof. The formula $\neg(\mathbf{A} \vee \mathbf{B}) \vee \mathbf{A} \vee \mathbf{B}$ is a propositional axiom. Hence $\vdash\mathbf{A} \vee \mathbf{B} \vee \neg(\mathbf{A} \vee \mathbf{B})$ by Lemma 2. From $\mathbf{A} \vee \mathbf{B} \vee \neg(\mathbf{A} \vee \mathbf{B})$ and $\neg\mathbf{A} \vee \mathbf{C}$ we infer $\vdash(\mathbf{B} \vee \neg(\mathbf{A} \vee \mathbf{B})) \vee \mathbf{C}$ by the cut rule. Hence $\vdash\mathbf{C} \vee \mathbf{B} \vee \neg(\mathbf{A} \vee \mathbf{B})$ by the commutativity rule, whence $\vdash\mathbf{B} \vee \mathbf{C} \vee \neg(\mathbf{A} \vee \mathbf{B})$ by Lemma 2. From $\mathbf{B} \vee \mathbf{C} \vee \neg(\mathbf{A} \vee \mathbf{B})$ and $\neg\mathbf{B} \vee \mathbf{C}$ we infer $\vdash(\mathbf{C} \vee \neg(\mathbf{A} \vee \mathbf{B})) \vee \mathbf{C}$ by the cut rule. Hence $\vdash\mathbf{C} \vee \mathbf{C} \vee \neg(\mathbf{A} \vee \mathbf{B})$ by the commutativity rule. From this and Lemma 2, we obtain $\vdash\neg(\mathbf{A} \vee \mathbf{B}) \vee \mathbf{C}$ as desired. □

Tautology Theorem. Every tautology is a theorem. If $\mathbf{A}$ is a tautological consequence of $\mathbf{B}_1, \dots, \mathbf{B}_n$, then $\mathbf{B}_1, \dots, \mathbf{B}_n \vdash \mathbf{A}$.

Proof. It suffices to prove the first statement, for $\mathbf{A}$ is a tautological consequence of $\mathbf{B}_1, \dots, \mathbf{B}_n$ if and only if $\mathbf{B}_1 \to \cdots \to \mathbf{B}_n \to \mathbf{A}$ is a tautology, and $\mathbf{B}_1 \to \cdots \to \mathbf{B}_n \to \mathbf{A}, \mathbf{B}_1, \dots, \mathbf{B}_n \vdash \mathbf{A}$ by n applications of the detachment rule. Now clearly if $\mathbf{A}$ is a tautology, so is $\mathbf{A} \vee \mathbf{A}$, and if $\vdash\mathbf{A} \vee \mathbf{A}$, then $\vdash\mathbf{A}$ by the contraction rule. Thus, it suffices to prove that if $\mathbf{A} \vee \mathbf{A}$ is a tautology, then it is a theorem. We prove more generally that if $n \geq 2$ and $\mathbf{A}_1 \vee \cdots \vee \mathbf{A}_n$ is a tautology, then it is a theorem. We proceed by induction on the sum of the lengths of the $\mathbf{A}_i$.

Suppose that each $\mathbf{A}_i$ is either elementary or the negation of an elementary formula. By Lemma 1, some $\mathbf{A}_i$ is the negation of some $\mathbf{A}_j$. Then $\mathbf{A}_i \vee \mathbf{A}_j$ is a propositional axiom, and by Lemma 2, $\vdash\mathbf{A}_1 \vee \cdots \vee \mathbf{A}_n$.

Suppose that some $\mathbf{A}_i$ is neither elementary nor the negation of an elementary formula. By Lemma 2, we have $\vdash \mathbf{A}_1 \vee \cdots \vee \mathbf{A}_n$ if and only if $\vdash \mathbf{A}_i \vee \cdots \vee \mathbf{A}_n \vee \mathbf{A}_1 \vee \cdots \vee \mathbf{A}_{i-1}$. Since those two formulae are tautologically equivalent, we may suppose that i is 1. Then $\mathbf{A}_1$ is either

(i) a disjunction;
(ii) the negation of a negation; or
(iii) the negation of a disjunction.

We prove the result in each case. Suppose that $\mathbf{A}_1$ is $\mathbf{B} \vee \mathbf{C}$. Then it is easy to prove that $\mathbf{B} \vee \mathbf{C} \vee \mathbf{A}_2 \vee \cdots \vee \mathbf{A}_n$ is a tautology. By induction hypothesis, it is a theorem. Hence $\vdash (\mathbf{B} \vee \mathbf{C}) \vee \mathbf{A}_2 \vee \cdots \vee \mathbf{A}_n$ by the associativity rule, that is, $\vdash \mathbf{A}_1 \vee \cdots \vee \mathbf{A}_n$. Suppose that $\mathbf{A}_1$ is $\neg\neg\mathbf{B}$. Then $\mathbf{B} \vee \mathbf{A}_2 \vee \cdots \vee \mathbf{A}_n$ is clearly a tautology; by induction hypothesis it is a theorem. By Lemma 3, $\vdash \mathbf{A}_1 \vee \cdots \vee \mathbf{A}_n$. Finally, suppose that $\mathbf{A}_1$ is $\neg(\mathbf{B} \vee \mathbf{C})$. Then both $\neg\mathbf{B} \vee \mathbf{A}_2 \vee \cdots \vee \mathbf{A}_n$ and $\neg\mathbf{C} \vee \mathbf{A}_2 \vee \cdots \vee \mathbf{A}_n$ are clearly tautologies, hence theorems by the induction hypothesis. By Lemma 4, we obtain $\vdash \mathbf{A}_1 \vee \cdots \vee \mathbf{A}_n$ as desired. □

COROLLARY. If $\mathbf{A}$ and $\mathbf{B}$ are tautologically equivalent, then $\vdash \mathbf{A} \leftrightarrow \mathbf{B}$.

Proof. For any truth valuation V, $V(\mathbf{A})$ is the same symbol as $V(\mathbf{B})$. It is then easily verified using the definition of the abbreviation $\mathbf{A} \leftrightarrow \mathbf{B}$ that $\mathbf{A} \leftrightarrow \mathbf{B}$ is a tautology. □

From the tautology theorem, we see that

(i) if $\vdash \mathbf{A} \leftrightarrow \mathbf{B}$, then $\vdash \mathbf{A}$ if and only if $\vdash \mathbf{B}$.

This result, as well as the detachment rule, should be kept in mind whenever we assert a statement of the form $\vdash \mathbf{A} \to \mathbf{B}$ or $\vdash \mathbf{A} \leftrightarrow \mathbf{B}$. Sometimes we refer to such a statement but we actually use "if $\vdash \mathbf{A}$ then $\vdash \mathbf{B}$" or "$\vdash \mathbf{A}$ if and only if $\vdash \mathbf{B}$", respectively. The tautology theorem allows us to prove the following results, among others:

(ii) if $\vdash \mathbf{A}$ or $\vdash \mathbf{B}$, then $\vdash \mathbf{A} \vee \mathbf{B}$;
(iii) $\vdash \mathbf{A} \wedge \mathbf{B}$ if and only if $\vdash \mathbf{A}$ and $\vdash \mathbf{B}$;
(iv) $\vdash (\mathbf{A} \to \mathbf{B}) \to (\mathbf{B} \to \mathbf{C}) \to \mathbf{A} \to \mathbf{C}$;
(v) $\vdash (\mathbf{A} \leftrightarrow \mathbf{B}) \to (\mathbf{B} \leftrightarrow \mathbf{C}) \to \mathbf{A} \leftrightarrow \mathbf{C}$;
(vi) $\vdash \mathbf{A} \leftrightarrow \neg\neg\mathbf{A}$;
(vii) $\vdash (\mathbf{A} \to \mathbf{B}) \leftrightarrow (\neg\mathbf{B} \to \neg\mathbf{A})$;
(viii) $\vdash (\neg\mathbf{A} \to \mathbf{A}) \to \mathbf{A}$;
(ix) $\vdash (\mathbf{A} \to \mathbf{B}) \to (\mathbf{A} \to \neg\mathbf{B}) \to \neg\mathbf{A}$;
(x) $\vdash \mathbf{A} \to (\mathbf{B} \to \neg\mathbf{A}) \to \neg\mathbf{B}$.

By the detachment rule, (i), (ii), and (iii), all of the above theorems have consequences on derivability. For example, from (iv) we obtain that $(\mathbf{B} \to \mathbf{C}) \to \mathbf{A} \to \mathbf{C}$ is derivable from $\mathbf{A} \to \mathbf{B}$, and hence that $\mathbf{A} \to \mathbf{C}$ is derivable from $\mathbf{A} \to \mathbf{B}$ and $\mathbf{B} \to \mathbf{C}$. Many more results can be deduced from the tautology theorem, such as properties of associativity and distributivity of $\vee$ and $\wedge$, and it is hardly possible to make an exhaustive list of even the most used ones. All of these results will be referred to generically as the tautology theorem.

3.3 Tautological induction. The tautology theorem allows for a useful characterization of the theorems of a first-order theory. More generally, let L and F be as in §3.2, and let T be of the form $F[R]$. We define a formal system T^* as follows. The language of T^* is L. The rules of inference of T^* are the rules in R together with all the rules with premises $\mathbf{B}_1, \ldots, \mathbf{B}_n$ and conclusion $\mathbf{A}$ whenever $\mathbf{A}$ is a tautological consequence of $\mathbf{B}_1, \ldots, \mathbf{B}_n$.

PROPOSITION 1. T and T^* are equivalent.

Proof. Using the tautology theorem and induction on theorems in T^*, it is clear that a theorem of T^* is a theorem of T. The converse is equally clear using induction on theorems in T and noting that the conclusions of the propositional rules are tautological consequences of their premises. □

To prove that all the theorems of T have a given property, we may thus use induction on theorems in T^*. Such a proof will be called a proof by *tautological induction on theorems in T*. The following proposition is often used in such proofs.

PROPOSITION 2. Let L and L' be first-order languages. Suppose that to each formula $\mathbf{A}$ of L' is associated a formula $\mathbf{A}^*$ of L in such a way that if $\mathbf{A}$ is $\mathbf{B} \vee \mathbf{C}$, then $\mathbf{A}^*$ is $\mathbf{B}^* \vee \mathbf{C}^*$ and if $\mathbf{A}$ is $\neg\mathbf{B}$, then $\mathbf{A}^*$ is $\neg\mathbf{B}^*$. If $\mathbf{A}$ is a tautological consequence of $\mathbf{B}_1, \ldots, \mathbf{B}_n$, then $\mathbf{A}^*$ is a tautological consequence of $\mathbf{B}_1^*, \ldots, \mathbf{B}_n^*$.

Proof. We may suppose $n = 0$. Let V be any truth valuation on L. Define a truth valuation V' on L' by setting $V'(\mathbf{A})$ to be $V(\mathbf{A}^*)$ for $\mathbf{A}$ elementary. Let $\mathbf{A}$ be any formula of L'. We prove by induction on the length of $\mathbf{A}$ that $V'(\mathbf{A})$ is $V(\mathbf{A}^*)$. If $\mathbf{A}$ is elementary, this is given. If $\mathbf{A}$ is $\mathbf{B} \vee \mathbf{C}$, then $\mathbf{A}^*$ is $\mathbf{B}^* \vee \mathbf{C}^*$ and by induction hypothesis $V'(\mathbf{B})$ is $V(\mathbf{B}^*)$ and $V'(\mathbf{C})$ is $V(\mathbf{C}^*)$; thus $V'(\mathbf{A})$ is $V(\mathbf{A}^*)$. If $\mathbf{A}$ is $\neg\mathbf{B}$, then $\mathbf{A}^*$ is $\neg\mathbf{B}^*$ and by induction hypothesis $V'(\mathbf{B})$ is $V(\mathbf{B}^*)$; thus $V'(\mathbf{A})$ is $V(\mathbf{A}^*)$. Now, if $\mathbf{A}$ is a tautology, $V'(\mathbf{A})$ is $\mathbf{T}$; hence $V(\mathbf{A}^*)$ is $\mathbf{T}$. □

§4 Theorems and rules in first-order theories

In this section, a first-order theory T is fixed. We write $\vdash$ and $\equiv$ instead of $\vdash_T$ and $\equiv_T$.

4.1 Quantification. We now prove some rules involving the symbol $\exists$.

$\forall$-INTRODUCTION RULE. If $\mathbf{x}$ is not free in $\mathbf{A}$, then $\mathbf{A} \to \mathbf{B} \vdash \mathbf{A} \to \forall\mathbf{x}\mathbf{B}$.

Proof. By the tautology theorem, $\vdash \neg\mathbf{B} \to \neg\mathbf{A}$. Since $\mathbf{x}$ is not free in $\neg\mathbf{A}$, the $\exists$-introduction rule yields $\vdash \exists\mathbf{x}\neg\mathbf{B} \to \neg\mathbf{A}$. Then by the tautology theorem, $\vdash \mathbf{A} \to \forall\mathbf{x}\mathbf{B}$. □

GENERALIZATION RULE. $\mathbf{A} \vdash \forall\mathbf{x}\mathbf{A}$.

Proof. By the tautology theorem, $\vdash \neg\forall\mathbf{x}\mathbf{A} \to \mathbf{A}$. By the $\forall$-introduction rule, $\vdash \neg\forall\mathbf{x}\mathbf{A} \to \forall\mathbf{x}\mathbf{A}$, whence $\vdash \forall\mathbf{x}\mathbf{A}$ by the tautology theorem. □

SUBSTITUTION RULE. If $\mathbf{A}'$ is an instance of $\mathbf{A}$, then $\mathbf{A} \vdash \mathbf{A}'$.

Proof. We first deal with the special case where $\mathbf{A}'$ is $\mathbf{A}[\mathbf{x}|\mathbf{a}]$. By the generalization rule, we have $\vdash \forall\mathbf{x}\mathbf{A}$, and by the substitution axioms $\vdash \neg\mathbf{A}[\mathbf{x}|\mathbf{a}] \to \exists\mathbf{x}\neg\mathbf{A}$. From these by the tautology theorem $\vdash \mathbf{A}[\mathbf{x}|\mathbf{a}]$. We now prove the general case. Suppose that $\mathbf{A}'$ is $\mathbf{A}[\mathbf{x}_1, \ldots, \mathbf{x}_n|\mathbf{a}_1, \ldots, \mathbf{a}_n]$. Let $\mathbf{y}_1, \ldots, \mathbf{y}_n$ be distinct variables not occurring in either $\mathbf{A}$ or $\mathbf{A}'$. By n applications of the special case, we find successively $\vdash \mathbf{A}[\mathbf{x}_1|\mathbf{y}_1], \ldots, \vdash \mathbf{A}[\mathbf{x}_1, \ldots, \mathbf{x}_n|\mathbf{y}_1, \ldots, \mathbf{y}_n]$. Applying again n times the special case, we find

$$\vdash \mathbf{A}[\mathbf{x}_1, \ldots, \mathbf{x}_n|\mathbf{y}_1, \ldots, \mathbf{y}_n][\mathbf{y}_1|\mathbf{a}_1], \ldots, \vdash \mathbf{A}[\mathbf{x}_1, \ldots, \mathbf{x}_n|\mathbf{y}_1, \ldots, \mathbf{y}_n][\mathbf{y}_1, \ldots, \mathbf{y}_n|\mathbf{a}_1, \ldots, \mathbf{a}_n],$$

i.e., $\vdash \mathbf{A}[\mathbf{x}_1, \ldots, \mathbf{x}_n|\mathbf{a}_1, \ldots, \mathbf{a}_n]$. □

SUBSTITUTION THEOREM. $\vdash \mathbf{A}[\mathbf{x}_1, \ldots, \mathbf{x}_n|\mathbf{a}_1, \ldots, \mathbf{a}_n] \to \exists\mathbf{x}_1 \ldots \exists\mathbf{x}_n\mathbf{A}$ and $\vdash \forall\mathbf{x}_1 \ldots \forall\mathbf{x}_n\mathbf{A} \to \mathbf{A}[\mathbf{x}_1, \ldots, \mathbf{x}_n|\mathbf{a}_1, \ldots, \mathbf{a}_n]$.

Proof. For each i, $\exists\mathbf{x}_{i+1} \ldots \exists\mathbf{x}_n\mathbf{A} \to \exists\mathbf{x}_i\exists\mathbf{x}_{i+1} \ldots \exists\mathbf{x}_n\mathbf{A}$ is a substitution axiom. A tautological consequence of all these is $\mathbf{A} \to \exists\mathbf{x}_1 \ldots \exists\mathbf{x}_n\mathbf{A}$, from which we obtain the first result by the substitution rule. The formula $\neg\mathbf{B} \to \exists\mathbf{x}\neg\mathbf{B}$ is a substitution axiom, from which we infer $\vdash \forall\mathbf{x}\mathbf{B} \to \mathbf{B}$ by the tautology theorem. Hence for each i, we have $\vdash \forall\mathbf{x}_i\forall\mathbf{x}_{i+1} \ldots \forall\mathbf{x}_n\mathbf{A} \to \forall\mathbf{x}_{i+1} \ldots \forall\mathbf{x}_n\mathbf{A}$. A tautological consequence of all these is $\forall\mathbf{x}_1 \ldots \forall\mathbf{x}_n\mathbf{A} \to \mathbf{A}$, from which we get the second result by the substitution rule. □

DISTRIBUTION RULE. $\mathbf{A} \to \mathbf{B} \vdash \exists\mathbf{x}\mathbf{A} \to \exists\mathbf{x}\mathbf{B}$ and $\mathbf{A} \to \mathbf{B} \vdash \forall\mathbf{x}\mathbf{A} \to \forall\mathbf{x}\mathbf{B}$.

Proof. The formula $\mathbf{B} \to \exists\mathbf{x}\mathbf{B}$ is a substitution axiom. Hence $\vdash \mathbf{A} \to \exists\mathbf{x}\mathbf{B}$ by the tautology theorem, and $\vdash \exists\mathbf{x}\mathbf{A} \to \exists\mathbf{x}\mathbf{B}$ by the $\exists$-introduction rule. Similarly, we have $\vdash \forall\mathbf{x}\mathbf{A} \to \mathbf{A}$ by the substitution theorem, from which $\vdash \forall\mathbf{x}\mathbf{A} \to \mathbf{B}$ by the tautology theorem, and $\vdash \forall\mathbf{x}\mathbf{A} \to \forall\mathbf{x}\mathbf{B}$ by the $\forall$-introduction rule. □

The following corollary will also be referred to as the distribution rule.

COROLLARY. $\mathbf{A} \leftrightarrow \mathbf{B} \vdash \exists\mathbf{x}\mathbf{A} \leftrightarrow \exists\mathbf{x}\mathbf{B}$ and $\mathbf{A} \leftrightarrow \mathbf{B} \vdash \forall\mathbf{x}\mathbf{A} \leftrightarrow \forall\mathbf{x}\mathbf{B}$.

Proof. By the tautology theorem, $\vdash \mathbf{A} \to \mathbf{B}$ and $\vdash \mathbf{B} \to \mathbf{A}$, whence $\vdash \exists \mathbf{x}\mathbf{A} \to \exists \mathbf{x}\mathbf{B}$ and $\vdash \exists \mathbf{x}\mathbf{B} \to \exists \mathbf{x}\mathbf{A}$ by the distribution rule. From these by the tautology theorem, $\vdash \exists \mathbf{x}\mathbf{A} \leftrightarrow \exists \mathbf{x}\mathbf{B}$. The proof of $\vdash \forall \mathbf{x}\mathbf{A} \leftrightarrow \forall \mathbf{x}\mathbf{B}$ is identical. □

Closure Theorem. $\mathbf{A} \equiv \forall \mathbf{x}_1 \dots \forall \mathbf{x}_n \mathbf{A}$.

Proof. We have $\mathbf{A} \vdash \forall \mathbf{x}_1 \dots \forall \mathbf{x}_n \mathbf{A}$ by n applications of the generalization rule. By the second part of the substitution theorem, we have $\vdash \forall \mathbf{x}_1 \dots \forall \mathbf{x}_n \mathbf{A} \to \mathbf{A}$, whence $\forall \mathbf{x}_1 \dots \forall \mathbf{x}_n \mathbf{A} \vdash \mathbf{A}$ by the detachment rule. □

We use the results of this paragraph to derive some useful theorems.

(i) If $\mathbf{x}$ is not free in $\mathbf{A}$, $\vdash \exists \mathbf{x}\mathbf{A} \leftrightarrow \mathbf{A}$;
(ii) if $\mathbf{x}$ is not free in $\mathbf{A}$, $\vdash \forall \mathbf{x}\mathbf{A} \leftrightarrow \mathbf{A}$;
(iii) $\vdash \exists \mathbf{x} \exists \mathbf{y}\mathbf{A} \leftrightarrow \exists \mathbf{y} \exists \mathbf{x}\mathbf{A}$;
(iv) $\vdash \forall \mathbf{x} \forall \mathbf{y}\mathbf{A} \leftrightarrow \forall \mathbf{y} \forall \mathbf{x}\mathbf{A}$;
(v) $\vdash \exists \mathbf{x} \forall \mathbf{y}\mathbf{A} \to \forall \mathbf{y} \exists \mathbf{x}\mathbf{A}$;
(vi) $\vdash \exists \mathbf{x}(\mathbf{A} \vee \mathbf{B}) \leftrightarrow \exists \mathbf{x}\mathbf{A} \vee \exists \mathbf{x}\mathbf{B}$;
(vii) $\vdash \forall \mathbf{x}(\mathbf{A} \wedge \mathbf{B}) \leftrightarrow \forall \mathbf{x}\mathbf{A} \wedge \forall \mathbf{x}\mathbf{B}$;
(viii) $\vdash \exists \mathbf{x}(\mathbf{A} \wedge \mathbf{B}) \to \exists \mathbf{x}\mathbf{A} \wedge \exists \mathbf{x}\mathbf{B}$;
(ix) $\vdash \forall \mathbf{x}\mathbf{A} \vee \forall \mathbf{x}\mathbf{B} \to \forall \mathbf{x}(\mathbf{A} \vee \mathbf{B})$.

The formula $\mathbf{A} \to \exists \mathbf{x}\mathbf{A}$ is a substitution axiom and if $\mathbf{x}$ is not free in $\mathbf{A}$, we obtain $\vdash \exists \mathbf{x}\mathbf{A} \to \mathbf{A}$ by the $\exists$-introduction rule from the tautology $\mathbf{A} \to \mathbf{A}$; so (i) by the tautology theorem. By the substitution theorem $\vdash \mathbf{A} \to \exists \mathbf{y} \exists \mathbf{x}\mathbf{A}$, whence $\vdash \exists \mathbf{x} \exists \mathbf{y}\mathbf{A} \to \exists \mathbf{y} \exists \mathbf{x}\mathbf{A}$ by the $\exists$-introduction rule. Similarly, $\vdash \exists \mathbf{y} \exists \mathbf{x}\mathbf{A} \to \exists \mathbf{x} \exists \mathbf{y}\mathbf{A}$, whence (iii) by the tautology theorem. By the substitution axioms and the distribution rule, $\vdash \forall \mathbf{y}\mathbf{A} \to \forall \mathbf{y} \exists \mathbf{x}\mathbf{A}$, whence (v) by the $\exists$-introduction rule. By the substitution axioms, $\vdash \mathbf{A} \to \exists \mathbf{x}\mathbf{A}$ and $\vdash \mathbf{B} \to \exists \mathbf{x}\mathbf{B}$, whence $\vdash \exists \mathbf{x}(\mathbf{A} \vee \mathbf{B}) \to \exists \mathbf{x}\mathbf{A} \vee \exists \mathbf{x}\mathbf{B}$ by the tautology theorem and the $\exists$-introduction rule. The same method proves (viii). The formula $\mathbf{A} \to \mathbf{A} \vee \mathbf{B}$ is a tautology; hence $\vdash \exists \mathbf{x}\mathbf{A} \to \exists \mathbf{x}(\mathbf{A} \vee \mathbf{B})$ by the distribution rule, so $\vdash \exists \mathbf{x}\mathbf{A} \vee \exists \mathbf{x}\mathbf{B} \to \exists \mathbf{x}(\mathbf{A} \vee \mathbf{B})$ by the tautology theorem. Together with the previous result, this proves (vi). Items (ii), (iv), (vii), and (ix) are proved in the same way as (i), (iii), (vi), and (viii), respectively, using the $\forall$-introduction rule instead of the $\exists$-introduction rule, and the other parts of the substitution theorem and the distribution rule.

4.2 Adjunction of nonlogical symbols. Let T be a first-order theory. We may form a first-order theory from T by adding new function and predicate symbols to the signature of $L(T)$ while leaving the nonlogical rules of T unchanged (but of course new formulae and logical rules featuring the new symbols are to be added). We may also form a new first-order theory by adding new nonlogical axioms. We then say that such a theory is *obtained from T by the adjunction of* those symbols and of those nonlogical axioms. The next proposition says that the mere adjunction of nonlogical symbols will not allow us to derive any formula of the original language that was not already derivable without using the new symbols.

Proposition. Let T be a first-order theory and let T' be obtained from T by the adjunction of new nonlogical symbols. Then T' is a conservative extension of T.

Proof. We shall prove a slightly more general result: if $\mathbf{A}$ is a theorem of T', if $\mathbf{x}$ is a variable not occurring in any member of some derivation of $\mathbf{A}$ in T', and if $\mathbf{A}^*$ is obtained from $\mathbf{A}$ by replacing every occurrence of a term of the form $\mathbf{f}\mathbf{a}_1 \dots \mathbf{a}_n$, where $\mathbf{f}$ is not a symbol of T, by $\mathbf{x}$ and every occurrence of an atomic formula of the form $\mathbf{p}\mathbf{a}_1 \dots \mathbf{a}_n$, where $\mathbf{p}$ is not a symbol of T, by $\mathbf{x} = \mathbf{x}$, then $\mathbf{A}^*$ is a theorem of T (note that the order in which those replacements are carried out has no influence on the resulting formula $\mathbf{A}^*$). First observe that this implies the proposition, since $\mathbf{A}^*$ is $\mathbf{A}$ whenever $\mathbf{A}$ is a formula of T. Now to prove the more general statement, we modify the given derivation of $\mathbf{A}$ as follows: replace any appearance of a functional equality axiom $\mathbf{x}_1 = \mathbf{y}_1 \to \cdots \to \mathbf{x}_n = \mathbf{y}_n \to \mathbf{f}\mathbf{x}_1 \dots \mathbf{x}_n = \mathbf{f}\mathbf{y}_1 \dots \mathbf{y}_n$ by a derivation in T of $\mathbf{x}_1 = \mathbf{y}_1 \to \cdots \to \mathbf{x}_n = \mathbf{y}_n \to \mathbf{x} = \mathbf{x}$; replace any appearance of a predicative equality axiom $\mathbf{x}_1 = \mathbf{y}_1 \to \cdots \to \mathbf{x}_n = \mathbf{y}_n \to \mathbf{p}\mathbf{x}_1 \dots \mathbf{x}_n \to \mathbf{p}\mathbf{y}_1 \dots \mathbf{y}_n$ by a derivation in T of $\mathbf{x}_1 = \mathbf{y}_1 \to \cdots \to \mathbf{x}_n = \mathbf{y}_n \to \mathbf{x} = \mathbf{x} \to \mathbf{x} = \mathbf{x}$; replace any occurrence of a term of the form $\mathbf{f}\mathbf{a}_1 \dots \mathbf{a}_n$, where $\mathbf{f}$ is not a symbol of T, by $\mathbf{x}$; finally, replace any occurrence of an atomic formula of the form $\mathbf{p}\mathbf{a}_1 \dots \mathbf{a}_n$, where $\mathbf{p}$ is not a symbol of T, by $\mathbf{x} = \mathbf{x}$. A quick inspection reveals that, with the exception of the equality axioms, an application of a logical rule of inference becomes an application of the same type of rule, and nonlogical rules are unaffected. The only nontrivial case is that of the substitution axioms. But if $\mathbf{u}^*$ denotes the designator of T obtained from a designator $\mathbf{u}$ of T' in the same way as $\mathbf{A}^*$ was obtained from $\mathbf{A}$, a straightforward induction on the length of $\mathbf{u}$ shows that, if $\mathbf{x}$ does not occur in $\mathbf{u}$ and is distinct from $\mathbf{y}$, $(\mathbf{u}[\mathbf{y}|\mathbf{a}])^*$ is $\mathbf{u}^*[\mathbf{y}|\mathbf{a}^*]$. It follows that a substitution axiom $\mathbf{A}[\mathbf{y}|\mathbf{a}] \to \exists\mathbf{y}\mathbf{A}$ becomes the substitution axiom $\mathbf{A}^*[\mathbf{y}|\mathbf{a}^*] \to \exists\mathbf{y}\mathbf{A}^*$. Thus, we have indeed obtained a derivation of $\mathbf{A}^*$ in T. □

When the only nonlogical symbols added are constants, we have the following stronger result.

Theorem on Constants. Let T be a first-order theory. Let T' be obtained from T by the adjunction of n new constants $\mathbf{e}_1, \dots, \mathbf{e}_n$, and let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be distinct variables. Then $\vdash_T \mathbf{A}$ if and only if $\vdash_{T'} \mathbf{A}[\mathbf{x}_1, \dots, \mathbf{x}_n|\mathbf{e}_1, \dots, \mathbf{e}_n]$.

Proof. Suppose that $\vdash_T \mathbf{A}$. Then $\vdash_{T'} \mathbf{A}$, whence $\vdash_{T'} \mathbf{A}[\mathbf{x}_1, \dots, \mathbf{x}_n|\mathbf{e}_1, \dots, \mathbf{e}_n]$ by the substitution rule. Conversely, suppose that $\vdash_{T'} \mathbf{A}[\mathbf{x}_1, \dots, \mathbf{x}_n|\mathbf{e}_1, \dots, \mathbf{e}_n]$. This means that there is a derivation of $\mathbf{A}[\mathbf{x}_1, \dots, \mathbf{x}_n|\mathbf{e}_1, \dots, \mathbf{e}_n]$ in T'. Let $\mathbf{y}_1, \dots, \mathbf{y}_n$ be distinct variables not appearing in that derivation. We replace every occurrence of $\mathbf{e}_i$ in members of that derivation by $\mathbf{y}_i$, for all i, and what we obtain is a derivation of $\mathbf{A}[\mathbf{x}_1, \dots, \mathbf{x}_n|\mathbf{y}_1, \dots, \mathbf{y}_n]$ in T. Indeed, an application of a logical rule of inference becomes an application of the same type of rule, and nonlogical rules are unaffected. Thus $\vdash_T \mathbf{A}[\mathbf{x}_1, \dots, \mathbf{x}_n|\mathbf{y}_1, \dots, \mathbf{y}_n]$, whence $\vdash_T \mathbf{A}$ by the substitution rule. □

4.3 The deduction theorem.

Lemma. Let T be a first-order theory and let $\mathbf{A}$ be a closed formula of T. Then $\vdash_T \mathbf{A} \to \mathbf{B}$ if and only if $\mathbf{A} \vdash_T \mathbf{B}$.

Proof. The necessity follows at once from the detachment rule. To prove the converse, we use tautological induction on theorems in $T[\mathbf{A}]$. If $\mathbf{B}$ is an axiom of $T[\mathbf{A}]$ other than $\mathbf{A}$, then $\mathbf{B}$ is an axiom of T and hence $\vdash_T \mathbf{A} \to \mathbf{B}$ by the tautology theorem. If $\mathbf{B}$ is $\mathbf{A}$, then $\mathbf{A} \to \mathbf{B}$ is a tautology, and hence $\vdash_T \mathbf{A} \to \mathbf{B}$. Suppose that $\mathbf{B}$ is a tautological consequence of $\mathbf{C}_1, \dots, \mathbf{C}_n$. By induction hypothesis, $\vdash_T \mathbf{A} \to \mathbf{C}_i$ for all i. Then $\vdash_T \mathbf{A} \to \mathbf{B}$ because it is a tautological consequence of all the $\mathbf{A} \to \mathbf{C}_i$. Suppose that $\mathbf{B}$ is inferred from $\mathbf{C} \to \mathbf{D}$ by the $\exists$-introduction rule with the variable $\mathbf{x}$. By induction hypothesis, $\vdash_T \mathbf{A} \to \mathbf{C} \to \mathbf{D}$. Then by the tautology theorem, $\vdash_T \mathbf{C} \to \mathbf{A} \to \mathbf{D}$. Since $\mathbf{x}$ is not free in $\mathbf{D}$ and $\mathbf{A}$ is closed, $\mathbf{x}$ is not free in $\mathbf{A}$; hence it is not free in $\mathbf{A} \to \mathbf{D}$. By the $\exists$-introduction rule, $\vdash_T \exists\mathbf{x}\mathbf{C} \to \mathbf{A} \to \mathbf{D}$. From the latter and the tautology theorem, $\vdash_T \mathbf{A} \to \exists\mathbf{x}\mathbf{C} \to \mathbf{D}$, that is, $\vdash_T \mathbf{A} \to \mathbf{B}$. □

Combining the Lemma and the theorem on constants, we obtain the following result:

Deduction Theorem. Let T be a first-order theory and let $\mathbf{A}$ be a formula of T whose free variables are among $\mathbf{x}_1, \dots, \mathbf{x}_n$. Let T' be obtained from T by the adjunction of n new constants $\mathbf{e}_1, \dots, \mathbf{e}_n$. Then $\mathbf{A} \to \mathbf{B}$ is a theorem of T if and only if $\mathbf{A}[\mathbf{x}_1, \dots, \mathbf{x}_n|\mathbf{e}_1, \dots, \mathbf{e}_n] \vdash_{T'} \mathbf{B}[\mathbf{x}_1, \dots, \mathbf{x}_n|\mathbf{e}_1, \dots, \mathbf{e}_n]$.

Reduction Theorem. Let T be a first-order theory, Γ a collection of formulae of T, and $\mathbf{A}$ a formula of T. Then $\Gamma \vdash_T \mathbf{A}$ if and only if there are formulae $\mathbf{B}_1, \dots, \mathbf{B}_n$ among the closures of the formulae of Γ such that $\vdash_T \mathbf{B}_1 \to \cdots \to \mathbf{B}_n \to \mathbf{A}$.

Proof. Suppose that $\Gamma \vdash_T \mathbf{A}$. This means that there exists a derivation of $\mathbf{A}$ in $T[\Gamma]$. Let $\mathbf{B}_1, \dots, \mathbf{B}_n$ be the closures of the formulae of Γ that appear in the derivation. Then by the closure theorem $\mathbf{A}$ is a theorem of $T[\mathbf{B}_1, \dots, \mathbf{B}_n]$. So $\vdash_T \mathbf{B}_1 \to \cdots \to \mathbf{B}_n \to \mathbf{A}$ by n applications of the deduction theorem. The converse follows from the closure theorem and n applications of the detachment rule. □

4.4 The equivalence theorem.

Equivalence Theorem. Let $\mathbf{A}'$ be a formula obtained from $\mathbf{A}$ by replacing some occurrences of $\mathbf{B}_1, \dots, \mathbf{B}_n$ by $\mathbf{B}'_1, \dots, \mathbf{B}'_n$. Then $\mathbf{B}_1 \leftrightarrow \mathbf{B}'_1, \dots, \mathbf{B}_n \leftrightarrow \mathbf{B}'_n \vdash \mathbf{A} \leftrightarrow \mathbf{A}'$.

Proof. If $\mathbf{B}_i$ is all of $\mathbf{A}$ for some i and if $\mathbf{A}$ is replaced by $\mathbf{B}'_i$, then $\mathbf{A}'$ is $\mathbf{B}'_i$ and $\vdash \mathbf{A} \leftrightarrow \mathbf{A}'$ by hypothesis. We now exclude this case, and we prove the result by induction on the length of $\mathbf{A}$. If $\mathbf{A}$ is atomic, then $\mathbf{A}$ has no subformula distinct from $\mathbf{A}$; hence $\mathbf{A}'$ is $\mathbf{A}$ and $\vdash \mathbf{A} \leftrightarrow \mathbf{A}'$ by the tautology theorem. Suppose that $\mathbf{A}$ is $\mathbf{C} \vee \mathbf{D}$. By the occurrence theorem, any subformula of $\mathbf{A}$ different from $\mathbf{A}$ occurs either in $\mathbf{C}$ or in $\mathbf{D}$. Denote by $\mathbf{C}'$ and $\mathbf{D}'$ the formulae such that $\mathbf{A}'$ is $\mathbf{C}' \vee \mathbf{D}'$. By induction hypothesis, $\vdash \mathbf{C} \leftrightarrow \mathbf{C}'$ and $\vdash \mathbf{D} \leftrightarrow \mathbf{D}'$, whence $\vdash \mathbf{A} \leftrightarrow \mathbf{A}'$ by the tautology theorem. Suppose that $\mathbf{A}$ is $\neg\mathbf{C}$. By the occurrence theorem, any subformula of $\mathbf{A}$ different from $\mathbf{A}$ occurs in $\mathbf{C}$. Let $\mathbf{C}'$ be the formula such that $\mathbf{A}'$ is $\neg\mathbf{C}'$. Then $\vdash \mathbf{C} \leftrightarrow \mathbf{C}'$ by induction hypothesis, whence $\vdash \mathbf{A} \leftrightarrow \mathbf{A}'$ by the tautology theorem. Suppose finally that $\mathbf{A}$ is $\exists\mathbf{x}\mathbf{C}$. By the occurrence theorem and the induction hypothesis, $\mathbf{A}'$ is $\exists\mathbf{x}\mathbf{C}'$ where $\vdash \mathbf{C} \leftrightarrow \mathbf{C}'$. By the distribution rule, $\vdash \mathbf{A} \leftrightarrow \mathbf{A}'$. ☐

VARIANT THEOREM. If $\mathbf{A}'$ is a variant of $\mathbf{A}$, then $\vdash \mathbf{A} \leftrightarrow \mathbf{A}'$.

Proof. Suppose first that $\mathbf{A}$ is $\exists\mathbf{x}\mathbf{B}$ and that $\mathbf{A}'$ is $\exists\mathbf{y}\mathbf{B}[\mathbf{x}|\mathbf{y}]$ where $\mathbf{y}$ is not free in $\mathbf{B}$. By the substitution axioms $\vdash \mathbf{B}[\mathbf{x}|\mathbf{y}] \to \exists\mathbf{x}\mathbf{B}$ and $\vdash \mathbf{B} \to \exists\mathbf{y}\mathbf{B}[\mathbf{x}|\mathbf{y}]$. By the $\exists$-introduction rule, $\vdash \exists\mathbf{y}\mathbf{B}[\mathbf{x}|\mathbf{y}] \to \exists\mathbf{x}\mathbf{B}$ and $\vdash \exists\mathbf{x}\mathbf{B} \to \exists\mathbf{y}\mathbf{B}[\mathbf{x}|\mathbf{y}]$. From these we get $\vdash \exists\mathbf{x}\mathbf{B} \leftrightarrow \exists\mathbf{y}\mathbf{B}[\mathbf{x}|\mathbf{y}]$ by the tautology theorem. In the general case, $\mathbf{A}'$ is obtained from $\mathbf{A}$ through $\mathbf{A}_1, \ldots, \mathbf{A}_{n-1}$ in the following way: setting $\mathbf{A}_0$ to be $\mathbf{A}$ and $\mathbf{A}_n$ to be $\mathbf{A}'$, $\mathbf{A}_i$ is obtained from $\mathbf{A}_{i-1}$ by replacement of an occurrence of a subformula of the form $\exists\mathbf{x}\mathbf{B}$ by $\exists\mathbf{y}\mathbf{B}[\mathbf{x}|\mathbf{y}]$ with $\mathbf{y}$ not free in $\mathbf{B}$. At each step, we have $\vdash \mathbf{A}_{i-1} \leftrightarrow \mathbf{A}_i$ by the special case and the equivalence theorem. Hence $\vdash \mathbf{A} \leftrightarrow \mathbf{A}'$ by the tautology theorem. ☐

From the variant theorem, the tautology theorem, and the substitution rule, we obtain:

VERSION THEOREM. If $\mathbf{A}'$ is a version of $\mathbf{A}$, then $\mathbf{A} \vdash \mathbf{A}'$.

4.5 The equality theorem.

SYMMETRY THEOREM. $\vdash \mathbf{a} = \mathbf{b} \leftrightarrow \mathbf{b} = \mathbf{a}$.

Proof. The formula $x = y \to x = x \to x = x \to y = x$ is an equality axiom. By the identity axiom $x = x$ and the tautology theorem, we obtain $\vdash x = y \to y = x$. By the substitution rule, we have $\vdash \mathbf{a} = \mathbf{b} \to \mathbf{b} = \mathbf{a}$ and $\vdash \mathbf{b} = \mathbf{a} \to \mathbf{a} = \mathbf{b}$. From these we get $\vdash \mathbf{a} = \mathbf{b} \leftrightarrow \mathbf{b} = \mathbf{a}$ by the tautology theorem. ☐

EQUALITY THEOREM FOR TERMS. Let $\mathbf{a}'$ be a term obtained from $\mathbf{a}$ by replacing some occurrences of $\mathbf{b}_1, \ldots, \mathbf{b}_n$ by $\mathbf{b}'_1, \ldots, \mathbf{b}'_n$. Then $\mathbf{b}_1 = \mathbf{b}'_1, \ldots, \mathbf{b}_n = \mathbf{b}'_n \vdash \mathbf{a} = \mathbf{a}'$.

Proof. If $\mathbf{b}_i$ is all of $\mathbf{a}$ for some i and if $\mathbf{a}$ is replaced by $\mathbf{b}'_i$, then $\mathbf{a}'$ is $\mathbf{b}'_i$ and $\vdash \mathbf{a} = \mathbf{a}'$ by hypothesis. We now exclude this case, and we prove the result by induction on the length of $\mathbf{a}$. If $\mathbf{a}$ is a variable, then there is no occurrence in $\mathbf{a}$ of a term distinct from $\mathbf{a}$; hence $\mathbf{a}'$ is $\mathbf{a}$ and $\vdash \mathbf{a} = \mathbf{a}'$ by the identity axioms and the substitution rule. Suppose that $\mathbf{a}$ is $\mathbf{f}\mathbf{c}_1 \ldots \mathbf{c}_k$. By the occurrence theorem, any occurrence of a term in $\mathbf{a}$ different from $\mathbf{a}$ is in one of the $\mathbf{c}_i$. Denote by $\mathbf{c}'_1, \ldots, \mathbf{c}'_k$ the terms such that $\mathbf{a}'$ is $\mathbf{f}\mathbf{c}'_1 \ldots \mathbf{c}'_k$. By induction hypothesis, $\vdash \mathbf{c}_i = \mathbf{c}'_i$ for all i. By the equality axioms and the substitution rule, $\vdash \mathbf{c}_1 = \mathbf{c}'_1 \to \cdots \to \mathbf{c}_k = \mathbf{c}'_k \to \mathbf{a} = \mathbf{a}'$, hence $\vdash \mathbf{a} = \mathbf{a}'$ by k applications of the detachment rule. ☐

EQUALITY THEOREM FOR FORMULAE. Let $\mathbf{A}'$ be a formula obtained from $\mathbf{A}$ by replacing some meaningful occurrences of $\mathbf{b}_1, \ldots, \mathbf{b}_n$ by $\mathbf{b}'_1, \ldots, \mathbf{b}'_n$. Then $\mathbf{b}_1 = \mathbf{b}'_1, \ldots, \mathbf{b}_n = \mathbf{b}'_n \vdash \mathbf{A} \leftrightarrow \mathbf{A}'$.

Proof. We prove the result by induction on the length of $\mathbf{A}$. Suppose that $\mathbf{A}$ is $\mathbf{p}\mathbf{c}_1 \ldots \mathbf{c}_k$. By the occurrence theorem, any occurrence of a term in $\mathbf{A}$ is in one of the $\mathbf{c}_i$. Denote by $\mathbf{c}'_1, \ldots, \mathbf{c}'_k$ the terms such that $\mathbf{A}'$ is $\mathbf{p}\mathbf{c}'_1 \ldots \mathbf{c}'_k$. By the equality theorem for terms, $\vdash \mathbf{c}_i = \mathbf{c}'_i$ for all i; also $\vdash \mathbf{c}'_i = \mathbf{c}_i$ by the symmetry theorem. By the equality axioms and the substitution rule, $\vdash \mathbf{c}_1 = \mathbf{c}'_1 \to \cdots \to \mathbf{c}_k = \mathbf{c}'_k \to \mathbf{A} \to \mathbf{A}'$ and $\vdash \mathbf{c}'_1 = \mathbf{c}_1 \to \cdots \to \mathbf{c}'_k = \mathbf{c}_k \to \mathbf{A}' \to \mathbf{A}$, hence $\vdash \mathbf{A} \to \mathbf{A}'$ and $\vdash \mathbf{A}' \to \mathbf{A}$ by k applications of the detachment rule. From these we get $\vdash \mathbf{A} \leftrightarrow \mathbf{A}'$ by the tautology theorem. Suppose that $\mathbf{A}$ is $\mathbf{C} \vee \mathbf{D}$. By the occurrence theorem, any occurrence of a term in $\mathbf{A}$ is either in $\mathbf{C}$ or in $\mathbf{D}$. Denote by $\mathbf{C}'$ and $\mathbf{D}'$ the formulae such that $\mathbf{A}'$ is $\mathbf{C}' \vee \mathbf{D}'$. By induction hypothesis, $\vdash \mathbf{C} \leftrightarrow \mathbf{C}'$ and $\vdash \mathbf{D} \leftrightarrow \mathbf{D}'$, whence $\vdash \mathbf{A} \leftrightarrow \mathbf{A}'$ by the tautology theorem. Suppose that $\mathbf{A}$ is $\neg\mathbf{C}$. By the occurrence theorem, any occurrence of a term in $\mathbf{A}$ is in $\mathbf{C}$. Let $\mathbf{C}'$ be the formula such that $\mathbf{A}'$ is $\neg\mathbf{C}'$. Then $\vdash \mathbf{C} \leftrightarrow \mathbf{C}'$ by induction hypothesis, whence $\vdash \mathbf{A} \leftrightarrow \mathbf{A}'$ by the tautology theorem. Suppose finally that $\mathbf{A}$ is $\exists\mathbf{x}\mathbf{C}$. By the hypothesis, the occurrence $\langle \exists, \mathbf{C} \rangle$ of $\mathbf{x}$ is not replaced. Hence by the occurrence theorem and the induction hypothesis, $\mathbf{A}'$ is $\exists\mathbf{x}\mathbf{C}'$ where $\vdash \mathbf{C} \leftrightarrow \mathbf{C}'$. By the distribution rule, $\vdash \mathbf{A} \leftrightarrow \mathbf{A}'$. ☐

The equality theorems can be combined with the deduction theorem to get the following useful corollary, which will also be referred to as the equality theorem. Its proof is straightforward.

COROLLARY.
(i) $\vdash \mathbf{a}_1 = \mathbf{a}'_1 \to \cdots \to \mathbf{a}_n = \mathbf{a}'_n \to \mathbf{b}[\mathbf{x}_1, \ldots, \mathbf{x}_n|\mathbf{a}_1, \ldots, \mathbf{a}_n] = \mathbf{b}[\mathbf{x}_1, \ldots, \mathbf{x}_n|\mathbf{a}'_1, \ldots, \mathbf{a}'_n]$;
(ii) $\vdash \mathbf{a}_1 = \mathbf{a}'_1 \to \cdots \to \mathbf{a}_n = \mathbf{a}'_n \to \mathbf{A}[\mathbf{x}_1, \ldots, \mathbf{x}_n|\mathbf{a}_1, \ldots, \mathbf{a}_n] \leftrightarrow \mathbf{A}[\mathbf{x}_1, \ldots, \mathbf{x}_n|\mathbf{a}'_1, \ldots, \mathbf{a}'_n]$.

REPLACEMENT THEOREM. If $\mathbf{x}$ does not occur in $\mathbf{a}$, then $\vdash \mathbf{A}[\mathbf{x}|\mathbf{a}] \leftrightarrow \exists \mathbf{x}(\mathbf{x} = \mathbf{a} \wedge \mathbf{A})$ and $\vdash \mathbf{A}[\mathbf{x}|\mathbf{a}] \leftrightarrow \forall \mathbf{x}(\mathbf{x} = \mathbf{a} \to \mathbf{A})$.

Proof. By the hypothesis, the formula $(\mathbf{a} = \mathbf{a} \wedge \mathbf{A}[\mathbf{x}|\mathbf{a}]) \to \exists \mathbf{x}(\mathbf{x} = \mathbf{a} \wedge \mathbf{A})$ is a substitution axiom. By the identity axioms and the substitution rule, we have $\vdash \mathbf{a} = \mathbf{a}$. A tautological consequence of these two formulae is

$$\mathbf{A}[\mathbf{x}|\mathbf{a}] \to \exists \mathbf{x}(\mathbf{x} = \mathbf{a} \wedge \mathbf{A}). \tag{1}$$

By the Corollary, we have $\vdash \mathbf{x} = \mathbf{a} \to \mathbf{A} \leftrightarrow \mathbf{A}[\mathbf{x}|\mathbf{a}]$, whence $\vdash \mathbf{x} = \mathbf{a} \wedge \mathbf{A} \to \mathbf{A}[\mathbf{x}|\mathbf{a}]$ by the tautology theorem. By the $\exists$-introduction rule,

$$\vdash \exists \mathbf{x}(\mathbf{x} = \mathbf{a} \wedge \mathbf{A}) \to \mathbf{A}[\mathbf{x}|\mathbf{a}]. \tag{2}$$

From (1) and (2), we get the desired result by the tautology theorem.

By the hypothesis and the substitution theorem, $\vdash \forall \mathbf{x}(\mathbf{x} = \mathbf{a} \to \mathbf{A}) \to \mathbf{a} = \mathbf{a} \to \mathbf{A}[\mathbf{x}|\mathbf{a}]$. A tautological consequence of this and $\mathbf{a} = \mathbf{a}$ is $\forall \mathbf{x}(\mathbf{x} = \mathbf{a} \to \mathbf{A}) \to \mathbf{A}[\mathbf{x}|\mathbf{a}]$. By the equality theorem and the tautology theorem, $\vdash \mathbf{A}[\mathbf{x}|\mathbf{a}] \to \mathbf{x} = \mathbf{a} \to \mathbf{A}$, whence $\vdash \mathbf{A}[\mathbf{x}|\mathbf{a}] \to \forall \mathbf{x}(\mathbf{x} = \mathbf{a} \to \mathbf{A})$ by the $\forall$-introduction rule. As above, we obtain the second result by the tautology theorem. ☐

4.6 Prenex form. A formula $\mathbf{A}$ is in *prenex form* if it is of the form $\mathbf{u}_1 \ldots \mathbf{u}_n \mathbf{B}$ where: $\mathbf{B}$ is open; each $\mathbf{u}_i$ is either $\exists \mathbf{x}_i$ or $\forall \mathbf{x}_i$; $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are distinct. It is clear that $\mathbf{u}_1, \ldots, \mathbf{u}_n$, and $\mathbf{B}$ are then uniquely determined; $\mathbf{u}_1 \ldots \mathbf{u}_n$ is called the *prefix* of $\mathbf{A}$ and $\mathbf{B}$ its *matrix*. If moreover each $\mathbf{u}_i$ is $\exists \mathbf{x}_i$ (resp. $\forall \mathbf{x}_i$), we say that $\mathbf{A}$ is *existential* (resp. *universal*). We shall show that for any formula $\mathbf{A}$, there is a formula $\mathbf{A}'$ in prenex form such that $\vdash \mathbf{A} \leftrightarrow \mathbf{A}'$.

Let $\mathbf{A}$ be a formula. The *prenex operations* that can be applied to $\mathbf{A}$ are the following:

(i) replace $\mathbf{A}$ by a variant;
(ii) if $\mathbf{x}$ is not free in $\mathbf{C}$, replace an occurrence of $\exists \mathbf{x}\mathbf{B} \vee \mathbf{C}$ by $\exists \mathbf{x}(\mathbf{B} \vee \mathbf{C})$;
(iii) if $\mathbf{x}$ is not free in $\mathbf{C}$, replace an occurrence of $\forall \mathbf{x}\mathbf{B} \vee \mathbf{C}$ by $\forall \mathbf{x}(\mathbf{B} \vee \mathbf{C})$;
(iv) if $\mathbf{x}$ is not free in $\mathbf{B}$, replace an occurrence of $\mathbf{B} \vee \exists \mathbf{x}\mathbf{C}$ by $\exists \mathbf{x}(\mathbf{B} \vee \mathbf{C})$;
(v) if $\mathbf{x}$ is not free in $\mathbf{B}$, replace an occurrence of $\mathbf{B} \vee \forall \mathbf{x}\mathbf{C}$ by $\forall \mathbf{x}(\mathbf{B} \vee \mathbf{C})$;
(vi) replace an occurrence of $\neg \exists \mathbf{x}\mathbf{B}$ by $\forall \mathbf{x} \neg \mathbf{B}$;
(vii) replace an occurrence of $\neg \forall \mathbf{x}\mathbf{B}$ by $\exists \mathbf{x} \neg \mathbf{B}$.

A formula in prenex form obtained from $\mathbf{A}$ by prenex operations is called a prenex form *of* $\mathbf{A}$.

THEOREM ON PRENEX OPERATIONS. If $\mathbf{A}'$ is obtained from $\mathbf{A}$ by prenex operations, then $\vdash \mathbf{A} \leftrightarrow \mathbf{A}'$.

Proof. By the tautology theorem, it suffices to consider the case where $\mathbf{A}'$ is obtained from $\mathbf{A}$ by just one of the seven prenex operations. If $\mathbf{A}'$ is a variant of $\mathbf{A}$, then the conclusion is the variant theorem. To prove that $\vdash \mathbf{A} \leftrightarrow \mathbf{A}'$ when $\mathbf{A}'$ is obtained from $\mathbf{A}$ by one of (ii)–(vii), we may suppose, by the equivalence theorem, that all of $\mathbf{A}$ is replaced. By the tautology theorem, it will suffice to prove

$$\vdash \exists \mathbf{x}(\mathbf{B} \vee \mathbf{C}) \to \exists \mathbf{x}\mathbf{B} \vee \mathbf{C}, \tag{3}$$
$$\vdash \exists \mathbf{x}\mathbf{B} \to \exists \mathbf{x}(\mathbf{B} \vee \mathbf{C}), \tag{4}$$
$$\vdash \mathbf{C} \to \exists \mathbf{x}(\mathbf{B} \vee \mathbf{C}), \tag{5}$$
$$\vdash \forall \mathbf{x}(\mathbf{B} \vee \mathbf{C}) \to \forall \mathbf{x}\mathbf{B} \vee \mathbf{C}, \tag{6}$$
$$\vdash \forall \mathbf{x}\mathbf{B} \to \forall \mathbf{x}(\mathbf{B} \vee \mathbf{C}), \tag{7}$$
$$\vdash \mathbf{C} \to \forall \mathbf{x}(\mathbf{B} \vee \mathbf{C}), \tag{8}$$
$$\vdash \mathbf{B} \vee \exists \mathbf{x}\mathbf{C} \leftrightarrow \exists \mathbf{x}(\mathbf{B} \vee \mathbf{C}), \tag{9}$$
$$\vdash \mathbf{B} \vee \forall \mathbf{x}\mathbf{C} \leftrightarrow \forall \mathbf{x}(\mathbf{B} \vee \mathbf{C}), \tag{10}$$
$$\vdash \neg \exists \mathbf{x}\mathbf{B} \leftrightarrow \neg \exists \mathbf{x} \neg \neg \mathbf{B}, \text{ and} \tag{11}$$
$$\vdash \neg \neg \exists \mathbf{x} \neg \mathbf{B} \leftrightarrow \exists \mathbf{x} \neg \mathbf{B}, \tag{12}$$

under the hypothesis that $\mathbf{x}$ is not free in $\mathbf{C}$ in (3)–(8) and not free in $\mathbf{B}$ in (9)–(10). Now $\mathbf{B} \to \exists\mathbf{x}\mathbf{B}$ is a substitution axiom. A tautological consequence of it is $\mathbf{B} \vee \mathbf{C} \to \exists\mathbf{x}\mathbf{B} \vee \mathbf{C}$. Hence we find (3) by the $\exists$-introduction rule. From the tautology $\mathbf{B} \to \mathbf{B} \vee \mathbf{C}$, we get (4) and (7) by the distriution rule. From the substitution axiom $\mathbf{B} \vee \mathbf{C} \to \exists\mathbf{x}(\mathbf{B} \vee \mathbf{C})$, we get (5) by the tautology theorem. By the substitution theorem, $\vdash \forall\mathbf{x}(\mathbf{B} \vee \mathbf{C}) \to \mathbf{B} \vee \mathbf{C}$. A tautological consequence of this is $\forall\mathbf{x}(\mathbf{B} \vee \mathbf{C}) \wedge \neg\mathbf{C} \to \mathbf{B}$, from which $\vdash \forall\mathbf{x}(\mathbf{B} \vee \mathbf{C}) \wedge \neg\mathbf{C} \to \forall\mathbf{x}\mathbf{B}$ by the $\forall$-introduction rule, whence (6) by the tautotlogy theorem. From the tautology $\mathbf{C} \to \mathbf{B} \vee \mathbf{C}$ we get (8) by the $\forall$-introduction rule. Interchanging $\mathbf{B}$ and $\mathbf{C}$ in (3)–(8) yields (9) and (10) by the equivalence theorem and the tautology theorem. We obtain (11) by the equivalence theorem and the tautology theorem. Finally, (12) is a tautology. □

Proposition. Every formula has a prenex form.

Proof. Let $\mathbf{A}$ be a formula. We prove our claim by induction on the length of $\mathbf{A}$. If $\mathbf{A}$ is atomic, then it is a prenex form of itself. If $\mathbf{A}$ is $\mathbf{B} \vee \mathbf{C}$, then by induction hypothesis $\mathbf{B}$ and $\mathbf{C}$ have prenex forms $\mathbf{B}'$ and $\mathbf{C}'$, and we may suppose by (i) that the bound variables in $\mathbf{B}'$ are distinct from the variables in $\mathbf{C}'$ and that the bound variables in $\mathbf{C}'$ are distinct from the variables in $\mathbf{B}'$. We then obtain a prenex form of $\mathbf{A}$ by applying successively (ii)–(v) to $\mathbf{B}' \vee \mathbf{C}'$. If $\mathbf{A}$ is $\neg\mathbf{B}$, then we obtain a prenex form of $\mathbf{A}$ from a prenex form of $\mathbf{B}$ and successive applications of (vi)–(vii). If $\mathbf{A}$ is $\exists\mathbf{x}\mathbf{B}$, then by the induction hypothesis and (i), $\mathbf{B}$ has a prenex form $\mathbf{B}'$ in which $\mathbf{x}$ is not bound. Then $\exists\mathbf{x}\mathbf{B}'$ is a prenex form of $\mathbf{A}$. □

The following operations are easily seen to be combinations of prenex operations.

(viii) if $\mathbf{x}$ is not free in $\mathbf{C}$, replace an occurrence of $\exists\mathbf{x}\mathbf{B} \wedge \mathbf{C}$ by $\exists\mathbf{x}(\mathbf{B} \wedge \mathbf{C})$;

(ix) if $\mathbf{x}$ is not free in $\mathbf{C}$, replace an occurrence of $\forall\mathbf{x}\mathbf{B} \wedge \mathbf{C}$ by $\forall\mathbf{x}(\mathbf{B} \wedge \mathbf{C})$;

(x) if $\mathbf{x}$ is not free in $\mathbf{B}$, replace an occurrence of $\mathbf{B} \wedge \exists\mathbf{x}\mathbf{C}$ by $\exists\mathbf{x}(\mathbf{B} \wedge \mathbf{C})$;

(xi) if $\mathbf{x}$ is not free in $\mathbf{B}$, replace an occurrence of $\mathbf{B} \wedge \forall\mathbf{x}\mathbf{C}$ by $\forall\mathbf{x}(\mathbf{B} \wedge \mathbf{C})$;

(xii) if $\mathbf{x}$ is not free in $\mathbf{C}$, replace an occurrence of $\exists\mathbf{x}\mathbf{B} \to \mathbf{C}$ by $\forall\mathbf{x}(\mathbf{B} \to \mathbf{C})$;

(xiii) if $\mathbf{x}$ is not free in $\mathbf{C}$, replace an occurrence of $\forall\mathbf{x}\mathbf{B} \to \mathbf{C}$ by $\exists\mathbf{x}(\mathbf{B} \to \mathbf{C})$;

(xiv) if $\mathbf{x}$ is not free in $\mathbf{B}$, replace an occurrence of $\mathbf{B} \to \exists\mathbf{x}\mathbf{C}$ by $\exists\mathbf{x}(\mathbf{B} \to \mathbf{C})$;

(xv) if $\mathbf{x}$ is not free in $\mathbf{B}$, replace an occurrence of $\mathbf{B} \to \forall\mathbf{x}\mathbf{C}$ by $\forall\mathbf{x}(\mathbf{B} \to \mathbf{C})$.

4.7 On English translation. We have already defined in §2.8 the *standard meaning* of a first-order language, which provides an efficient way of translating formulae of a first-order language into English. For example, $\forall x(x \neq x)$ may be translated as "for all individuals x, x does not equal x", where the kind of individuals may be further specified in a given theory. Of course, in order to translate formulae of an arbitrary first-order theory, it is also necessary to explain the meaning of the nonlogical symbols.

In Chapter VI we shall introduce a first-order theory called ZF, and we shall often translate derivations in English. This does not mean that we abandon the formalism we have described with great care in this chapter. Our objects of study will remain formal expressions, and our theorems (whose statements will always be accurate) will say something about formal systems and not, as is usual in mathematics, about abstract individuals. These informal derivations in English should be considered as guidelines from which one can recover derivations using the results of the previous sections. We intend each step of these guidelines to be easily translated back into the formal system by simple applications of these results. The problem is that no reference can be given where the derivations of the theorems of ZF can be found, so one must trust that these informal derivations actually work. The best way to be convinced of it is of course to try and translate some of them into formal derivations.

There is a device that is often used in informal derivations which is worth explaining. When proving $\vdash_T \mathbf{A} \to \mathbf{B}$, it is customary to start the derivation by saying "suppose that $\mathbf{A}$ holds". One way to translate such an argument formally is by using the deduction theorem.

Chapter Two

The Question of Consistency

§1 Consistency and completeness

1.1 Consistency. We say that a formal system F is *inconsistent* if every formula of F is a theorem of F; it is *consistent* otherwise. We shall only be interested in applying these definitions when $L(F)$ is a first-order language and when the propositional rules are derivable in F. We can then formulate a simple criterion for consistency with the tautology theorem. Since there is no truth valuation V for which both $V(\mathbf{A})$ and $V(\neg\mathbf{A})$ are T, it follows that every formula is a tautological consequence of $\mathbf{A}$ and $\neg\mathbf{A}$. Consequently, F is inconsistent if (and only if) there is a formula $\mathbf{A}$ of F such that $\vdash_F\mathbf{A}$ and $\vdash_F\neg\mathbf{A}$.

Proposition 1. Let T be a first-order theory, $\mathbf{A}$ a formula of T, and $\mathbf{A}'$ its closure. Then $\vdash_T\mathbf{A}$ if and only if $T[\neg\mathbf{A}']$ is inconsistent.

Proof. If $\vdash_T\mathbf{A}$, then $\vdash_{T[\neg\mathbf{A}']}\mathbf{A}$ and hence $\vdash_{T[\neg\mathbf{A}']}\mathbf{A}'$ by the closure theorem. Obviously $\vdash_{T[\neg\mathbf{A}']}\neg\mathbf{A}'$, so by the above remark $T[\neg\mathbf{A}']$ is inconsistent. Conversely, suppose that $T[\neg\mathbf{A}']$ is inconsistent. Then $\vdash_{T[\neg\mathbf{A}']}\mathbf{A}'$, so by the deduction theorem, $\vdash_T\neg\mathbf{A}' \to \mathbf{A}'$. Thus we obtain $\vdash_T\mathbf{A}'$ by the tautology theorem, whence $\vdash_T\mathbf{A}$ by the closure theorem. □

Proposition 2. Let F be a formal system whose language is a first-order language and whose rules of inference are the propositional rules for $L(F)$, and let Γ be a collection of formulae of $L(F)$. Then $F[\Gamma]$ is inconsistent if and only if some disjunction of negations of formulae in Γ is a tautology.

Proof. Recall that, by the tautology theorem, a formula of $L(F)$ is a theorem of $F[\Gamma]$ if and only if it is a tautological consequence of formulae in Γ. It follows that if $F[\Gamma]$ is inconsistent, then $x = x$ and $x \neq x$ are tautological consequences of formulae $\mathbf{A}_1, \dots, \mathbf{A}_n$ in Γ. Then no truth valuation assigns T to all of $\mathbf{A}_1, \dots, \mathbf{A}_n$, i.e., $\neg\mathbf{A}_1 \vee \cdots \vee \neg\mathbf{A}_n$ is a tautology. Conversely, if some disjunction of negations of formulae in Γ is a tautology, then any formula of $L(F)$ is a tautological consequence of formulae in Γ, so $F[\Gamma]$ is inconsistent by the tautology theorem. □

1.2 The consistency of first-order logic. Given the meaning we have in mind for first-order theories, we want them to be consistent. But the problem of determining the consistency of a first-order theory is not an easy one in general. In some simple cases it can be done by elementary (and finitary) arguments. This is fortunately the case for the first-order theories exempt of nonlogical axioms.

Proposition. A first-order theory with no nonlogical axioms is consistent.

Proof. Let T be a first-order theory with no nonlogical axioms, and let L be the first-order language consisting of $L(T)$ and a new constant $\mathbf{e}$. To every formula $\mathbf{A}$ of T, we let $\mathbf{A}^*$ be obtained from $\mathbf{A}$ by deleting all occurrences of $\exists\mathbf{x}$ and replacing all remaining terms by $\mathbf{e}$. Clearly $\mathbf{A}^*$ is a formula of L. We prove that if $\vdash_T\mathbf{A}$, then $\mathbf{A}^*$ is a tautological consequence of $\mathbf{e} = \mathbf{e}$. We proceed by tautological induction on theorems in T. If $\mathbf{A}$ is $\mathbf{B}[\mathbf{x}|\mathbf{a}] \to \exists\mathbf{x}\mathbf{B}$, then $\mathbf{A}^*$ is $\mathbf{B}^* \to \mathbf{B}^*$ which is a tautology. If $\mathbf{A}$ is $\mathbf{x} = \mathbf{x}$, then $\mathbf{A}^*$ is $\mathbf{e} = \mathbf{e}$ and is a tautological consequence of itself. If $\mathbf{A}$ is an equality axiom, then $\mathbf{A}^*$ is either $\mathbf{e} = \mathbf{e} \to \cdots \to \mathbf{e} = \mathbf{e}$, which is a tautological consequence of $\mathbf{e} = \mathbf{e}$, or $\mathbf{e} = \mathbf{e} \to \cdots \to \mathbf{e} = \mathbf{e} \to \mathbf{pe}\dots\mathbf{e} \to \mathbf{pe}\dots\mathbf{e}$, which is a tautology. If $\mathbf{A}$ is a tautological consequence of $\mathbf{B}_1, \dots, \mathbf{B}_n$, then $\mathbf{A}^*$ is a tautological consequence of $\mathbf{B}_1^*, \dots, \mathbf{B}_n^*$ by proposition 2 of ch. I §3.3, and so is a tautological consequence of $\mathbf{e} = \mathbf{e}$ by induction hypothesis. If $\mathbf{A}$ is obtained from $\mathbf{B} \to \mathbf{C}$ by the $\exists$-introduction rule with the variable $\mathbf{x}$, then $\mathbf{A}^*$ is $\mathbf{B}^* \to \mathbf{C}^*$, which is a tautological consequence of $\mathbf{e} = \mathbf{e}$ by induction hypothesis. Let now V be a truth valuation on L which assigns T to $\mathbf{e} = \mathbf{e}$. Then $V(\mathbf{e} \neq \mathbf{e})$ is F, and so $x \neq x$ is not a theorem of T. □

1.3 Completeness. A notion that is closely related to consistency is that of completeness. A first-order theory T is *complete* if for any closed formula $\mathbf{A}$ of T, $\vdash_T\mathbf{A}$ or $\vdash_T\neg\mathbf{A}$. Completeness, like consistency, is obviously a desirable property of a first-order theory. We will not pursue this notion any further in this chapter, but we shall obtain important results on the completeness of particular first-order theories in Chapter III.

§2 Extensions by definitions

2.1 Definitions of predicate symbols. Let T be a first-order theory, $\mathbf{D}$ a formula of T, and $\mathbf{x}_1, \ldots, \mathbf{x}_n$ distinct variables including the variables free in $\mathbf{D}$. Let T' be the first-order theory obtained from T by the adjunction of a new n-ary predicate symbol $\mathbf{p}$ and the new nonlogical axiom $\mathbf{p}\mathbf{x}_1 \ldots \mathbf{x}_n \leftrightarrow \mathbf{D}$. For any formula $\mathbf{A}$ of T', we choose a variant $\mathbf{D}'$ of $\mathbf{D}$ in which no variable of $\mathbf{A}$ is bound and we let $\mathbf{A}^*$ be the formula of T obtained from $\mathbf{A}$ by replacing each occurrence of $\mathbf{p}\mathbf{a}_1 \ldots \mathbf{a}_n$ in $\mathbf{A}$ by

$$\mathbf{D}'[\mathbf{x}_1, \ldots, \mathbf{x}_n|\mathbf{a}_1, \ldots, \mathbf{a}_n].$$

We now assume that for each formula $\mathbf{A}$ a formula $\mathbf{A}^*$ has been chosen once and for all. Since only occurrences of atomic formulae are replaced to form $\mathbf{A}^*$, it follows that if $\mathbf{A}$ is $\mathbf{B} \vee \mathbf{C}$, then $\mathbf{A}^*$ is a variant of $\mathbf{B}^* \vee \mathbf{C}^*$, if $\mathbf{A}$ is $\neg\mathbf{B}$, then $\mathbf{A}^*$ is a variant of $\neg\mathbf{B}^*$, and if $\mathbf{A}$ is $\exists\mathbf{x}\mathbf{B}$, then $\mathbf{A}^*$ is $\exists\mathbf{x}\mathbf{B}'$ where $\mathbf{B}'$ is a variant of $\mathbf{B}^*$.

Theorem on Predicative Definitions. With the notations of this paragraph, $\vdash_{T'}\mathbf{A} \leftrightarrow \mathbf{A}^*$ and T' is a conservative extension of T.

Proof. To prove the first assertion, it suffices, by the equivalence theorem, to show that $\vdash_{T'}\mathbf{p}\mathbf{a}_1 \ldots \mathbf{a}_n \leftrightarrow \mathbf{D}'[\mathbf{x}_1, \ldots, \mathbf{x}_n|\mathbf{a}_1, \ldots, \mathbf{a}_n]$. This follows from the axiom $\mathbf{p}\mathbf{x}_1 \ldots \mathbf{x}_n \leftrightarrow \mathbf{D}$ by the version theorem.

To prove that T' is a conservative extension of T, it suffices to prove $\vdash_T\mathbf{A}^*$ for any theorem $\mathbf{A}$ of T'; for if $\mathbf{A}$ is a formula of T, then $\mathbf{A}^*$ is $\mathbf{A}$. We use tautological induction on theorems in T'. If $\mathbf{A}$ is a substitution axiom $\mathbf{B}[\mathbf{x}|\mathbf{a}] \to \exists\mathbf{x}\mathbf{B}$, then $\mathbf{A}^*$ is easily seen to be a variant of the substitution axiom $\mathbf{B}^*[\mathbf{x}|\mathbf{a}] \to \exists\mathbf{x}\mathbf{B}^*$, and hence $\vdash_T\mathbf{A}^*$ by the variant theorem. If $\mathbf{A}$ is an identity axiom, then $\mathbf{p}$ does not occur in $\mathbf{A}$; hence $\mathbf{A}^*$ is $\mathbf{A}$ and $\vdash_T\mathbf{A}^*$. Similarly, if $\mathbf{A}$ is an equality axiom in which $\mathbf{p}$ does not occur, then $\mathbf{A}^*$ is $\mathbf{A}$ and $\vdash_T\mathbf{A}^*$. Suppose that $\mathbf{A}$ is an equality axiom of the form $\mathbf{y}_1 = \mathbf{y}'_1 \to \cdots \to \mathbf{y}_n = \mathbf{y}'_n \to \mathbf{p}\mathbf{y}_1 \ldots \mathbf{y}_n \to \mathbf{p}\mathbf{y}'_1 \ldots \mathbf{y}'_n$. Then $\mathbf{A}^*$ is

$$\mathbf{y}_1 = \mathbf{y}'_1 \to \cdots \to \mathbf{y}_n = \mathbf{y}'_n \to \mathbf{D}'[\mathbf{x}_1, \ldots, \mathbf{x}_n|\mathbf{y}_1 \ldots \mathbf{y}_n] \to \mathbf{D}''[\mathbf{x}_1, \ldots, \mathbf{x}_n|\mathbf{y}'_1 \ldots \mathbf{y}'_n]$$

for some variants $\mathbf{D}'$ and $\mathbf{D}''$ of $\mathbf{D}$, so it is a theorem of T by the equality theorem and the variant theorem. If $\mathbf{A}$ is a nonlogical axiom of T, then $\mathbf{p}$ does not occur in $\mathbf{A}$; hence $\mathbf{A}^*$ is $\mathbf{A}$ and $\vdash_T\mathbf{A}^*$. Finally, if $\mathbf{A}$ is $\mathbf{p}\mathbf{x}_1 \ldots \mathbf{x}_n \leftrightarrow \mathbf{D}$, then $\mathbf{A}^*$ is $\mathbf{D}' \leftrightarrow \mathbf{D}$, which is a theorem of T by the variant theorem.

Suppose that $\mathbf{A}$ is a tautological consequence of $\mathbf{B}_1, \ldots, \mathbf{B}_k$. Then by Proposition 2 of ch. I §3.3, $\mathbf{A}^*$ is a tautological consequence of variants of $\mathbf{B}^*_1, \ldots, \mathbf{B}^*_k$, and hence $\vdash_T\mathbf{A}^*$ by the variant theorem, the tautology theorem, and the induction hypothesis. Suppose that $\mathbf{A}$ is inferred from $\mathbf{B} \to \mathbf{C}$ by the $\exists$-introduction rule with the variable $\mathbf{x}$. Then $\mathbf{A}^*$ is $\exists\mathbf{x}\mathbf{B}' \to \mathbf{C}'$ where $\mathbf{B}'$ and $\mathbf{C}'$ are variants of $\mathbf{B}^*$ and $\mathbf{C}^*$. The induction hypothesis is $\vdash_T\mathbf{B}^* \to \mathbf{C}^*$. Since $\mathbf{x}$ is not free in $\mathbf{C}^*$, $\mathbf{A}$ is a theorem of T by the $\exists$-introduction rule and the variant theorem. □

2.2 Definitions of function symbols. Let T be a first-order theory, $\mathbf{D}$ a formula of T, and $\mathbf{x}_1, \ldots, \mathbf{x}_n$, $\mathbf{y}$, $\mathbf{y}'$ distinct variables such that $\mathbf{x}_1, \ldots, \mathbf{x}_n$, $\mathbf{y}$ include the variables free in $\mathbf{D}$. Let T' be the first-order theory obtained from T by the adjunction of a new n-ary function symbol $\mathbf{f}$ and the new nonlogical axiom $\mathbf{y} = \mathbf{f}\mathbf{x}_1 \ldots \mathbf{x}_n \leftrightarrow \mathbf{D}$. For any atomic formula $\mathbf{A}$ of T', we define a formula $\mathbf{A}^*$ of T by induction on the number of occurrences of $\mathbf{f}$ in $\mathbf{A}$. If there are no such occurrences, then $\mathbf{A}^*$ is $\mathbf{A}$. If $\mathbf{f}$ occurs in $\mathbf{A}$, consider the last occurrence of a term $\mathbf{f}\mathbf{a}_1 \ldots \mathbf{a}_n$ in $\mathbf{A}$, so that $\mathbf{f}$ does not occur in $\mathbf{a}_1, \ldots, \mathbf{a}_n$, and let $\mathbf{B}$ be obtained from $\mathbf{A}$ by replacing that occurrence by a variable $\mathbf{z}$ not occurring in $\mathbf{A}$. Then $\mathbf{B}$ is an atomic formula in which $\mathbf{f}$ occurs one less time than in $\mathbf{A}$; we choose a variant $\mathbf{D}'$ of $\mathbf{D}$ in which no variable of $\mathbf{A}$ is bound and we let $\mathbf{A}^*$ be the formula

$$\exists\mathbf{z}(\mathbf{D}'[\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{y}|\mathbf{a}_1, \ldots, \mathbf{a}_n, \mathbf{z}] \wedge \mathbf{B}^*).$$

We now assume that a formula $\mathbf{A}^*$ is fixed for any atomic formula $\mathbf{A}$ of T. If $\mathbf{A}$ is any formula of T', let $\mathbf{A}^*$ be the formula obtained from $\mathbf{A}$ by replacing each occurrence of an atomic formula $\mathbf{B}$ in $\mathbf{A}$ by $\mathbf{B}^*$. For the same reason as in §2.1, if $\mathbf{A}$ is $\mathbf{B} \vee \mathbf{C}$, then $\mathbf{A}^*$ is a variant of $\mathbf{B}^* \vee \mathbf{C}^*$, if $\mathbf{A}$ is $\neg\mathbf{B}$, then $\mathbf{A}^*$ is a variant of $\neg\mathbf{B}^*$, and if $\mathbf{A}$ is $\exists\mathbf{x}\mathbf{B}$, then $\mathbf{A}^*$ is $\exists\mathbf{x}\mathbf{B}'$ where $\mathbf{B}'$ is a variant of $\mathbf{B}^*$.

Lemma 1. Let $\mathbf{A}$ be a formula of T'. Assume that $\mathbf{A}^\circ$ is built as $\mathbf{A}^*$, except that, at each step, instead of replacing the last occurrence of a term $\mathbf{f}\mathbf{a}_1 \ldots \mathbf{a}_n$, we allow the replacement of any such occurrence as long as $\mathbf{f}$ does not occur in $\mathbf{a}_1, \ldots, \mathbf{a}_n$. Then $\vdash_T\mathbf{A}^* \leftrightarrow \mathbf{A}^\circ$.

Proof. By the equivalence theorem, it suffices to prove the result for $\mathbf{A}$ atomic. Replacing $\mathbf{A}^*$ and $\mathbf{A}^\circ$ by variants if necessary, we may assume that any given occurrence of $\mathbf{f}$ in $\mathbf{A}$ is replaced by the same variable in forming $\mathbf{A}^*$ and $\mathbf{A}^\circ$, and that this variable does not occur in all of $\mathbf{A}$. It follows that if $\mathbf{A}^*$ has the form $\exists\mathbf{z}_1(\mathbf{D}_1 \wedge \cdots \exists\mathbf{z}_m(\mathbf{D}_m \wedge \mathbf{B})\cdots)$ for some formula $\mathbf{B}$ of T, then $\mathbf{A}^\circ$ has the form $\exists\mathbf{z}'_1(\mathbf{D}'_1 \wedge \cdots \exists\mathbf{z}'_m(\mathbf{D}'_m \wedge \mathbf{B})\cdots)$ where for each i there is exactly one j such that $\mathbf{z}_i$ is $\mathbf{z}'_j$ and $\mathbf{D}_i$ is a variant of $\mathbf{D}'_j$. Thus $\mathbf{A}^\circ$ can be obtained from $\mathbf{A}^*$ using the following operations: prenex operations; replacing an occurrence of $\exists\mathbf{x}\exists\mathbf{x}'$ by $\exists\mathbf{x}'\exists\mathbf{x}$; replacing an occurrence of $\mathbf{C} \wedge \mathbf{C}'$ by $\mathbf{C}' \wedge \mathbf{C}$. By ch. I §4.1 (iii), the tautology theorem, the equivalence theorem, and the theorem on prenex operations, we have $\vdash_T \mathbf{A}^* \leftrightarrow \mathbf{A}^\circ$. □

Lemma 2. If $\vdash_T \exists\mathbf{x}\mathbf{A}$ and $\vdash_T \mathbf{A} \rightarrow \mathbf{A}[\mathbf{x}|\mathbf{x}'] \rightarrow \mathbf{x} = \mathbf{x}'$ for some $\mathbf{x}'$ not free in $\mathbf{A}$ and $\mathbf{B}$, then $\vdash_T \exists\mathbf{x}(\mathbf{A} \wedge \mathbf{B}) \leftrightarrow \forall\mathbf{x}(\mathbf{A} \rightarrow \mathbf{B})$ and $\vdash_T \neg\exists\mathbf{x}(\mathbf{A} \wedge \mathbf{B}) \leftrightarrow \exists\mathbf{x}(\mathbf{A} \wedge \neg\mathbf{B})$.

Proof. For the first assertion, we derive both implications. By the substitution theorem and the tautology theorem, $\vdash_T \forall\mathbf{x}(\mathbf{A} \rightarrow \mathbf{B}) \wedge \mathbf{A} \rightarrow \mathbf{A} \wedge \mathbf{B}$, whence $\vdash_T \forall\mathbf{x}(\mathbf{A} \rightarrow \mathbf{B}) \wedge \exists\mathbf{x}\mathbf{A} \rightarrow \exists\mathbf{x}(\mathbf{A} \wedge \mathbf{B})$ by the distribution rule and prenex operations. By the first hypothesis and the tautology theorem, we obtain $\vdash_T \forall\mathbf{x}(\mathbf{A} \rightarrow \mathbf{B}) \rightarrow \exists\mathbf{x}(\mathbf{A} \wedge \mathbf{B})$. Conversely, $\vdash_T \mathbf{A} \wedge \mathbf{B} \rightarrow \mathbf{A}[\mathbf{x}|\mathbf{x}'] \rightarrow \mathbf{B}[\mathbf{x}|\mathbf{x}']$ by the equality theorem, the second hypothesis, and the tautology theorem. So by the $\forall$-introduction rule and the $\exists$-introduction rule, we obtain $\vdash_T \exists\mathbf{x}(\mathbf{A} \wedge \mathbf{B}) \rightarrow \forall\mathbf{x}'(\mathbf{A}[\mathbf{x}|\mathbf{x}'] \rightarrow \mathbf{B}[\mathbf{x}|\mathbf{x}'])$, whence $\vdash_T \exists\mathbf{x}(\mathbf{A} \wedge \mathbf{B}) \rightarrow \forall\mathbf{x}(\mathbf{A} \rightarrow \mathbf{B})$ by the variant theorem. The second assertion follows from the first one by the tautology theorem and the equivalence theorem. □

Lemma 3. With the notations of this paragraph, suppose that

$$\vdash_T \exists\mathbf{y}\mathbf{D} \text{ and} \tag{1}$$

$$\vdash_T \mathbf{D} \rightarrow \mathbf{D}[\mathbf{y}|\mathbf{y}'] \rightarrow \mathbf{y} = \mathbf{y}'. \tag{2}$$

If $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are terms of T, $\mathbf{A}$ a formula of T', and $\mathbf{D}'$ a variant of $\mathbf{D}$ in which no variable of $\mathbf{A}, \mathbf{a}_1, \ldots, \mathbf{a}_n$ is bound, then $\vdash_T \mathbf{A}[\mathbf{z}|\mathbf{f}\mathbf{a}_1 \ldots \mathbf{a}_n]^* \leftrightarrow \exists\mathbf{z}(\mathbf{D}'[\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{y}|\mathbf{a}_1, \ldots, \mathbf{a}_n, \mathbf{z}] \wedge \mathbf{A}^*)$.

Proof. Throughout the proof, we let $\mathbf{D}_1$ abbreviate $\mathbf{D}'[\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{y}|\mathbf{a}_1, \ldots, \mathbf{a}_n, \mathbf{z}]$. We prove the result by induction on the length of $\mathbf{A}$. Suppose first that $\mathbf{A}$ is atomic, and proceed by induction on the number of free occurrences of $\mathbf{z}$ in $\mathbf{A}$. If there are none, then $\mathbf{z}$ is not free in $\mathbf{A}^*$. From the tautology $\mathbf{D}_1 \wedge \mathbf{A}^* \rightarrow \mathbf{A}^*$, we obtain $\vdash_T \exists\mathbf{z}(\mathbf{D}_1 \wedge \mathbf{A}^*) \rightarrow \mathbf{A}^*$ by the $\exists$-introduction rule. Conversely, from $\vdash_T \exists\mathbf{z}\mathbf{D}_1$, we obtain $\vdash_T \mathbf{A}^* \rightarrow \exists\mathbf{z}(\mathbf{D}_1 \wedge \mathbf{A}^*)$ by the tautology theorem and prenex operations. Hence the desired equivalence holds by the tautology theorem. Suppose that $\mathbf{z}$ has a free occurrence in $\mathbf{A}$, and let $\mathbf{B}$ be obtained from $\mathbf{A}$ by replacing all free occurrences of $\mathbf{z}$ save one by a variable $\mathbf{w}$ distinct from $\mathbf{z}$ and not occurring in $\mathbf{A}$, $\mathbf{a}_1$, …, $\mathbf{a}_n$, or $\mathbf{D}'$. Then by Lemma 1 $\vdash_T \mathbf{A}[\mathbf{z}|\mathbf{f}\mathbf{a}_1 \ldots \mathbf{a}_n]^* \leftrightarrow \exists\mathbf{z}(\mathbf{D}_1 \wedge \mathbf{B}[\mathbf{w}|\mathbf{f}\mathbf{a}_1 \ldots \mathbf{a}_n]^*)$, and by induction hypothesis $\vdash_T \mathbf{B}[\mathbf{w}|\mathbf{f}\mathbf{a}_1 \ldots \mathbf{a}_n]^* \leftrightarrow \exists\mathbf{w}(\mathbf{D}_1[\mathbf{z}|\mathbf{w}] \wedge \mathbf{B}^*)$. Hence by the equivalence theorem, $\vdash_T \mathbf{A}[\mathbf{z}|\mathbf{f}\mathbf{a}_1 \ldots \mathbf{a}_n]^* \leftrightarrow \exists\mathbf{z}(\mathbf{D}_1 \wedge \exists\mathbf{w}(\mathbf{D}_1[\mathbf{z}|\mathbf{w}] \wedge \mathbf{B}^*))$, and by prenex operations

$$\vdash_T \mathbf{A}[\mathbf{z}|\mathbf{f}\mathbf{a}_1 \ldots \mathbf{a}_n]^* \leftrightarrow \exists\mathbf{z}\exists\mathbf{w}(\mathbf{D}_1 \wedge \mathbf{D}_1[\mathbf{z}|\mathbf{w}] \wedge \mathbf{B}^*). \tag{3}$$

By a version of (2), the equality theorem, and the tautology theorem, $\vdash_T \mathbf{D}_1 \rightarrow \mathbf{D}_1[\mathbf{z}|\mathbf{w}] \rightarrow \mathbf{A}^* \leftrightarrow \mathbf{B}^*$, whence $\vdash_T \mathbf{D}_1[\mathbf{z}|\mathbf{w}] \wedge \mathbf{D}_1 \wedge \mathbf{A}^* \leftrightarrow \mathbf{D}_1 \wedge \mathbf{D}_1[\mathbf{z}|\mathbf{w}] \wedge \mathbf{B}^*$ by the tautology theorem. So by (3) and the equivalence theorem, we find $\vdash_T \mathbf{A}[\mathbf{z}|\mathbf{f}\mathbf{a}_1 \ldots \mathbf{a}_n]^* \leftrightarrow \exists\mathbf{z}\exists\mathbf{w}(\mathbf{D}_1[\mathbf{z}|\mathbf{w}] \wedge \mathbf{D}_1 \wedge \mathbf{A}^*)$. Since $\mathbf{w}$ is not free in $\mathbf{D}_1 \wedge \mathbf{A}^*$ and $\mathbf{z}$ is not free in $\mathbf{D}_1[\mathbf{z}|\mathbf{w}]$, we get

$$\vdash_T \mathbf{A}[\mathbf{z}|\mathbf{f}\mathbf{a}_1 \ldots \mathbf{a}_n]^* \leftrightarrow \exists\mathbf{w}\mathbf{D}_1[\mathbf{z}|\mathbf{w}] \wedge \exists\mathbf{z}(\mathbf{D}_1 \wedge \mathbf{A}^*) \tag{4}$$

by prenex operations. From the tautology $\mathbf{D}_1 \wedge \mathbf{A}^* \rightarrow \mathbf{D}_1$, we get $\vdash_T \exists\mathbf{z}(\mathbf{D}_1 \wedge \mathbf{A}^*) \rightarrow \exists\mathbf{w}\mathbf{D}_1[\mathbf{z}|\mathbf{w}]$ by the distribution rule and the variant theorem. From the latter and (4), we obtain $\vdash_T \mathbf{A}[\mathbf{z}|\mathbf{f}\mathbf{a}_1 \ldots \mathbf{a}_n]^* \leftrightarrow \exists\mathbf{z}(\mathbf{D}_1 \wedge \mathbf{A}^*)$ by the tautology theorem, which is the desired result.

Suppose now that $\mathbf{A}$ is $\mathbf{B} \vee \mathbf{C}$. Then $\mathbf{A}[\mathbf{z}|\mathbf{f}\mathbf{a}_1 \ldots \mathbf{a}_n]^*$ is a variant of $\mathbf{B}[\mathbf{z}|\mathbf{f}\mathbf{a}_1 \ldots \mathbf{a}_n]^* \vee \mathbf{C}[\mathbf{z}|\mathbf{f}\mathbf{a}_1 \ldots \mathbf{a}_n]^*$. By the induction hypotheses, the variant theorem, and the tautology theorem, $\vdash_T \mathbf{A}[\mathbf{z}|\mathbf{f}\mathbf{a}_1 \ldots \mathbf{a}_n]^* \leftrightarrow \exists\mathbf{z}(\mathbf{D}_1 \wedge \mathbf{B}^*) \vee \exists\mathbf{z}(\mathbf{D}_1 \wedge \mathbf{C}^*)$. We obtain the desired result by ch. I §4.1 (vi), the tautology theorem, and the equivalence theorem.

Suppose that $\mathbf{A}$ is $\neg\mathbf{B}$. Then $\mathbf{A}[\mathbf{z}|\mathbf{f}\mathbf{a}_1 \ldots \mathbf{a}_n]^*$ is a variant $\neg\mathbf{B}[\mathbf{z}|\mathbf{f}\mathbf{a}_1 \ldots \mathbf{a}_n]^*$. By the induction hypothesis, the variant theorem, and the tautology theorem, $\vdash_T \mathbf{A}[\mathbf{z}|\mathbf{f}\mathbf{a}_1 \ldots \mathbf{a}_n]^* \leftrightarrow \neg\exists\mathbf{z}(\mathbf{D}_1 \wedge \mathbf{B}^*)$. So the desired result follows from Lemma 2 and the tautology theorem.

Finally, suppose that $\mathbf{A}$ is $\exists\mathbf{xB}$. Since $\mathbf{fa}_1 \ldots \mathbf{a}_n$ is substitutible for $\mathbf{z}$ in $\mathbf{A}$, $\mathbf{x}$ does not occur in $\mathbf{a}_1, \ldots, \mathbf{a}_n$. In particular, either $\mathbf{x}$ is $\mathbf{z}$ or $\mathbf{x}$ is not free in $\mathbf{D}_1$. Suppose first that $\mathbf{x}$ is $\mathbf{z}$, so that $\mathbf{A}[\mathbf{z}|\mathbf{fa}_1 \ldots \mathbf{a}_n]$ is $\mathbf{A}$. By prenex operations, $\vdash_T \exists\mathbf{zD}_1 \wedge \exists\mathbf{zB}^* \leftrightarrow \exists\mathbf{z}(\mathbf{D}_1 \wedge \exists\mathbf{zB}^*)$. From this and $\vdash_T \exists\mathbf{zD}_1$, we obtain $\vdash_T \exists\mathbf{zB}^* \leftrightarrow \exists\mathbf{z}(\mathbf{D}_1 \wedge \exists\mathbf{zB}^*)$ by the tautology theorem, which is a variant of the expected result. Suppose that $\mathbf{x}$ and $\mathbf{z}$ are distinct. Then $\mathbf{A}[\mathbf{z}|\mathbf{fa}_1 \ldots \mathbf{a}_n]^*$ is a variant of $\exists\mathbf{xB}[\mathbf{z}|\mathbf{fa}_1 \ldots \mathbf{a}_n]^*$. By induction hypothesis, the variant theorem, and the equivalence theorem, $\vdash_T \mathbf{A}[\mathbf{z}|\mathbf{fa}_1 \ldots \mathbf{a}_n]^* \leftrightarrow \exists\mathbf{x}\exists\mathbf{z}(\mathbf{D}_1 \wedge \mathbf{B}^*)$. By prenex operations (exchanging beforehand $\exists\mathbf{x}$ and $\exists\mathbf{z}$ by virtue of ch. 1 §4.1 (iii) and the tautology theorem), we obtain $\vdash_T \mathbf{A}[\mathbf{z}|\mathbf{fa}_1 \ldots \mathbf{a}_n]^* \leftrightarrow \exists\mathbf{z}(\mathbf{D}_1 \wedge \exists\mathbf{xB}^*)$ which is as desired up to a variant. □

Theorem on Functional Definitions. With the notations of this paragraph, $\vdash_{T'} \mathbf{A} \leftrightarrow \mathbf{A}^*$. If moreover $\vdash_T \exists\mathbf{yD}$ and $\vdash_T \mathbf{D} \rightarrow \mathbf{D}[\mathbf{y}|\mathbf{y}'] \rightarrow \mathbf{y} = \mathbf{y}'$, then T' is a conservative extension of T.

Proof. To prove $\vdash_{T'} \mathbf{A} \leftrightarrow \mathbf{A}^*$, it suffices, by the equivalence theorem, to consider the case where $\mathbf{A}$ is atomic. We proceed by induction on the number of occurrences of $\mathbf{f}$ in $\mathbf{A}$. If $\mathbf{f}$ does not occur in $\mathbf{A}$, then $\mathbf{A}^*$ is $\mathbf{A}$ and hence $\vdash_{T'} \mathbf{A} \leftrightarrow \mathbf{A}^*$ by the tautology theorem. Suppose that $\mathbf{f}$ occurs in $\mathbf{A}$, and let $\mathbf{a}_1, \ldots, \mathbf{a}_n$, $\mathbf{B}$, and $\mathbf{D}'$ be as in the construction of $\mathbf{A}^*$. From the nonlogical axiom $\mathbf{y} = \mathbf{fx}_1 \ldots \mathbf{x}_n \leftrightarrow \mathbf{D}$, we obtain $\vdash_{T'} \mathbf{z} = \mathbf{fa}_1 \ldots \mathbf{a}_n \leftrightarrow \mathbf{D}'[\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{y}|\mathbf{a}_1, \ldots, \mathbf{a}_n, \mathbf{z}]$ by the version theorem. By the equivalence theorem, we get $\vdash_{T'} \exists\mathbf{z}(\mathbf{z} = \mathbf{fa}_1 \ldots \mathbf{a}_n \wedge \mathbf{B}^*) \leftrightarrow \mathbf{A}^*$. Since $\mathbf{f}$ occurs in $\mathbf{B}$ one less time than in $\mathbf{A}$, we have $\vdash_{T'} \mathbf{B} \leftrightarrow \mathbf{B}^*$ by the induction hypothesis; hence $\vdash_{T'} \exists\mathbf{z}(\mathbf{z} = \mathbf{fa}_1 \ldots \mathbf{a}_n \wedge \mathbf{B}) \leftrightarrow \mathbf{A}^*$ by the equivalence theorem. By the replacement theorem, we have $\vdash_{T'} \exists\mathbf{z}(\mathbf{z} = \mathbf{fa}_1 \ldots \mathbf{a}_n \wedge \mathbf{B}) \leftrightarrow \mathbf{B}[\mathbf{z}|\mathbf{fa}_1 \ldots \mathbf{a}_n]$, whence $\vdash_{T'} \mathbf{B}[\mathbf{z}|\mathbf{fa}_1 \ldots \mathbf{a}_n] \leftrightarrow \mathbf{A}^*$ by the tautology theorem, i.e., $\vdash_{T'} \mathbf{A} \leftrightarrow \mathbf{A}^*$.

To prove that T' is a conservative extension of T, it suffices to prove $\vdash_T \mathbf{A}^*$ for any theorem $\mathbf{A}$ of T'; for $\mathbf{A}^*$ is $\mathbf{A}$ for any formula $\mathbf{A}$ of T. We use tautological induction on theorems in T'. Suppose that $\mathbf{A}$ is a substitution axiom $\mathbf{B}[\mathbf{x}|\mathbf{a}] \rightarrow \exists\mathbf{xB}$. We prove $\vdash_T \mathbf{A}^*$ by induction on the number of occurrences of $\mathbf{f}$ in $\mathbf{a}$. If there are none, then $\mathbf{A}^*$ is a variant of the substitution axiom $\mathbf{B}^*[\mathbf{x}|\mathbf{a}] \leftrightarrow \exists\mathbf{xB}^*$, and hence is a theorem of T. If $\mathbf{f}$ occurs in $\mathbf{a}$, consider some occurrence of a term $\mathbf{fa}_1 \ldots \mathbf{a}_n$ in $\mathbf{a}$ such that $\mathbf{f}$ does not occur in $\mathbf{a}_1, \ldots, \mathbf{a}_n$, and let $\mathbf{b}$ be obtained from $\mathbf{a}$ by replacing that occurrence by a variable $\mathbf{z}$ not occurring in $\mathbf{A}$, so that $\mathbf{B}[\mathbf{x}|\mathbf{a}]$ is $\mathbf{B}[\mathbf{x}|\mathbf{b}][\mathbf{z}|\mathbf{fa}_1 \ldots \mathbf{a}_n]$. By the hypotheses on $\mathbf{D}$, we may apply Lemma 3 which yields

$$\vdash_T \mathbf{B}[\mathbf{x}|\mathbf{a}]^* \leftrightarrow \exists\mathbf{z}(\mathbf{D}'[\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{y}|\mathbf{a}_1, \ldots, \mathbf{a}_n, \mathbf{z}] \wedge \mathbf{B}[\mathbf{x}|\mathbf{b}]^*) \tag{5}$$

for some suitable variant $\mathbf{D}'$ of $\mathbf{D}$. Since $\mathbf{f}$ occurs in $\mathbf{b}$ one less time than in $\mathbf{a}$, we have $\vdash_T \mathbf{B}[\mathbf{x}|\mathbf{b}]^* \rightarrow \exists\mathbf{xB}^*$ by the induction hypothesis, whence $\vdash_T \mathbf{D}'[\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{y}|\mathbf{a}_1, \ldots, \mathbf{a}_n, \mathbf{z}] \wedge \mathbf{B}[\mathbf{x}|\mathbf{b}]^* \rightarrow \exists\mathbf{xB}^*$ by the tautology theorem. Using the $\exists$-introduction rule and (5), we obtain $\vdash_T \mathbf{B}[\mathbf{x}|\mathbf{a}]^* \rightarrow \exists\mathbf{xB}^*$, whence $\vdash_T \mathbf{A}^*$ by the variant theorem.

If $\mathbf{A}$ is an identity axiom, then $\mathbf{f}$ does not occur in $\mathbf{A}$; hence $\mathbf{A}^*$ is $\mathbf{A}$ and $\vdash_T \mathbf{A}^*$. Similarly, if $\mathbf{A}$ is an equality axiom in which $\mathbf{f}$ does not occur, then $\mathbf{A}^*$ is $\mathbf{A}$ and $\vdash_T \mathbf{A}^*$. Suppose that $\mathbf{A}$ is $\mathbf{y}_1 = \mathbf{y}_1' \rightarrow \cdots \rightarrow \mathbf{y}_n = \mathbf{y}_n' \rightarrow \mathbf{fy}_1 \ldots \mathbf{y}_n = \mathbf{fy}_1' \ldots \mathbf{y}_n'$. Then $\mathbf{A}^*$ is

$$\exists\mathbf{z}'(\mathbf{D}_1 \wedge \exists\mathbf{z}(\mathbf{D}_2 \wedge (\mathbf{y}_1 = \mathbf{y}_1' \rightarrow \cdots \rightarrow \mathbf{y}_n = \mathbf{y}_n' \rightarrow \mathbf{z} = \mathbf{z}'))),$$

where $\mathbf{D}_1$ is $\mathbf{D}'[\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{y}|\mathbf{y}_1', \ldots, \mathbf{y}_n', \mathbf{z}']$ and $\mathbf{D}_2$ is $\mathbf{D}''[\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{y}|\mathbf{y}_1, \ldots, \mathbf{y}_n, \mathbf{z}]$, for some $\mathbf{z}$ and $\mathbf{z}'$ distinct and not occurring in $\mathbf{A}$ and some variants $\mathbf{D}'$ and $\mathbf{D}''$ of $\mathbf{D}$. Note that $\mathbf{z}$ is not free in $\mathbf{D}_1$. From the hypothesis (2) and the version theorem,

$$\vdash_T \mathbf{D}''[\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{y}|\mathbf{y}_1', \ldots, \mathbf{y}_n', \mathbf{z}] \rightarrow \mathbf{D}_1 \rightarrow \mathbf{z} = \mathbf{z}', \tag{6}$$

and by the equality theorem,

$$\vdash_T \mathbf{y}_1 = \mathbf{y}_1' \rightarrow \cdots \rightarrow \mathbf{y}_n = \mathbf{y}_n' \rightarrow \mathbf{D}_2 \leftrightarrow \mathbf{D}''[\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{y}|\mathbf{y}_1', \ldots, \mathbf{y}_n', \mathbf{z}]. \tag{7}$$

A tautological consequence of (6) and (7) is

$$\mathbf{D}_1 \wedge \mathbf{D}_2 \rightarrow \mathbf{D}_1 \wedge \mathbf{D}_2 \wedge (\mathbf{y}_1 = \mathbf{y}_1' \rightarrow \cdots \rightarrow \mathbf{y}_n = \mathbf{y}_n' \rightarrow \mathbf{z} = \mathbf{z}'). \tag{8}$$

From the hypothesis (1) and the version theorem, we have $\vdash_T \exists\mathbf{z}'\mathbf{D}_1$ and $\vdash_T \exists\mathbf{zD}_2$, from which we infer $\vdash_T \exists\mathbf{z}'\exists\mathbf{z}(\mathbf{D}_1 \wedge \mathbf{D}_2)$ by the tautology theorem and prenex operations. Thus we obtain $\vdash_T \mathbf{A}^*$ from (8) by the distribution rule, the detachment rule, and prenex operations. If $\mathbf{A}$ is a nonlogical axiom of T, then $\mathbf{f}$ does not occur in $\mathbf{A}$; hence $\mathbf{A}^*$ is $\mathbf{A}$ and $\vdash_T \mathbf{A}^*$. If $\mathbf{A}$ is $\mathbf{y} = \mathbf{fx}_1 \ldots \mathbf{x}_n \leftrightarrow \mathbf{D}$, then $\mathbf{A}^*$ is $\exists\mathbf{z}(\mathbf{D}'[\mathbf{y}|\mathbf{z}] \wedge \mathbf{y} = \mathbf{z}) \leftrightarrow \mathbf{D}$, which is a theorem of T by the replacement theorem, the tautology theorem and the equivalence theorem, and the variant theorem.

Suppose that $\mathbf{A}$ is a tautological consequence of $\mathbf{B}_1, \dots, \mathbf{B}_k$. Then by Proposition 2 of ch. I §3.3, $\mathbf{A}^*$ is a tautological consequence of variants of $\mathbf{B}_1^*, \dots, \mathbf{B}_k^*$, and hence $\vdash_T \mathbf{A}^*$ by the variant theorem, the tautology theorem, and the induction hypothesis. Suppose that $\mathbf{A}$ is inferred from $\mathbf{B} \to \mathbf{C}$ by the $\exists$-introduction rule with the variable $\mathbf{x}$. Then $\mathbf{A}^*$ is $\exists \mathbf{x}\mathbf{B}' \to \mathbf{C}'$ where $\mathbf{B}'$ and $\mathbf{C}'$ are variants of $\mathbf{B}^*$ and $\mathbf{C}^*$. The induction hypothesis is $\vdash_T \mathbf{B}^* \to \mathbf{C}^*$. Since $\mathbf{x}$ is not free in $\mathbf{C}^*$, $\mathbf{A}$ is a theorem of T by the $\exists$-introduction rule and the variant theorem. □

The formula of (1) is called the *existence condition* for $\mathbf{y}$ in $\mathbf{D}$, and that of (2), for $\mathbf{y}'$ not free in $\mathbf{D}$, a *uniqueness condition* for $\mathbf{y}$ in $\mathbf{D}$. The following criteria are often useful.

PROPOSITION 1. If $\mathbf{D}$ as in this paragraph is of the form $\mathbf{y} = \mathbf{a}$ where $\mathbf{y}$ does not occur in $\mathbf{a}$, then existence and uniqueness conditions for $\mathbf{y}$ in $\mathbf{D}$ are theorems of T.

Proof. We have $\vdash_T \mathbf{a} = \mathbf{a}$ as an instance of an identity axiom. Hence $\vdash_T \exists \mathbf{y}(\mathbf{y} = \mathbf{a})$ by the substitution axioms and the detachment rule. This proves that the existence condition for $\mathbf{y}$ in $\mathbf{D}$ is a theorem of T. From the equality axioms, the symmetry theorem, and the equivalence theorem, we have $\vdash_T \mathbf{y} = \mathbf{a} \to \mathbf{y}' = \mathbf{a} \to \mathbf{a} = \mathbf{a} \to \mathbf{y} = \mathbf{y}'$. A tautological consequence of this and $\mathbf{a} = \mathbf{a}$ is $\mathbf{y} = \mathbf{a} \to \mathbf{y}' = \mathbf{a} \to \mathbf{y} = \mathbf{y}'$, which is a desired uniqueness condition for $\mathbf{y}$ in $\mathbf{D}$. □

PROPOSITION 2. Suppose that $\mathbf{D}$ as in this paragraph is $(\mathbf{A}_1 \wedge \mathbf{B}_1) \vee \cdots \vee (\mathbf{A}_n \wedge \mathbf{B}_n)$ where $\mathbf{y}$ is not free in $\mathbf{A}_1, \dots, \mathbf{A}_n$. If $\vdash_T \mathbf{A}_1 \vee \cdots \vee \mathbf{A}_n$ and $\vdash_T \mathbf{A}_i \to \exists \mathbf{y}\mathbf{B}_i$ for each i, then $\vdash_T \exists \mathbf{y}\mathbf{D}$. If $\vdash_T \mathbf{A}_i \to \neg \mathbf{A}_j$ whenever $i \neq j$ and if $\vdash_T \mathbf{A}_i \to \mathbf{B}_i \to \mathbf{B}_i[\mathbf{y}|\mathbf{y}'] \to \mathbf{y} = \mathbf{y}'$ for each i, then $\vdash_T \mathbf{D} \to \mathbf{D}[\mathbf{y}|\mathbf{y}'] \to \mathbf{y} = \mathbf{y}'$.

Proof. A tautological consequence of $\mathbf{A}_1 \vee \cdots \vee \mathbf{A}_n$ and all the $\mathbf{A}_i \to \exists \mathbf{y}\mathbf{B}_i$ is $\exists \mathbf{y}\mathbf{B}_1 \vee \cdots \vee \exists \mathbf{y}\mathbf{B}_n$, whence $\vdash_T \exists \mathbf{y}\mathbf{D}$ by (vi) of ch. I §4.1. The formula $\mathbf{D} \to \mathbf{D}[\mathbf{y}|\mathbf{y}'] \to \mathbf{y} = \mathbf{y}'$ is a tautological consequence of all the formulae $\vdash_T \mathbf{A}_i \to \neg \mathbf{A}_j$ and $\mathbf{A}_i \to \mathbf{B}_i \to \mathbf{B}_i[\mathbf{y}|\mathbf{y}'] \to \mathbf{y} = \mathbf{y}'$. □

The adjunction of a new symbol with a legit defining axiom to a first-order theory is sometimes called a *definition*. Definitions in the form of Proposition 1 are then called *explicit definitions*. By the equality theorem, $\vdash_{T'} \mathbf{y} = \mathbf{f}\mathbf{x}_1 \dots \mathbf{x}_n \leftrightarrow \mathbf{y} = \mathbf{a}$ if and only if $\vdash_{T'} \mathbf{f}\mathbf{x}_1 \dots \mathbf{x}_n = \mathbf{a}$, so in order not to encumber the notations, the defining axiom for an explicit definition is usually written $\mathbf{f}\mathbf{x}_1 \dots \mathbf{x}_n = \mathbf{a}$ instead of $\mathbf{y} = \mathbf{f}\mathbf{x}_1 \dots \mathbf{x}_n \leftrightarrow \mathbf{y} = \mathbf{a}$. Definitions in the form of Proposition 2 are called *definitions by cases*.

2.3 Extensions by definitions. Let T be a first-order theory. A first-order theory T' is called an *extension by definitions* of T if there are first-order theories $T_0, \dots, T_n$ such that T_0 is T, T_n is T', and for each i, one of the following holds.

(i) T_i is obtained from T_{i-1} by the adjunction of an n-ary predicate symbol $\mathbf{p}$ and a new axiom $\mathbf{p}\mathbf{x}_1 \dots \mathbf{x}_n \leftrightarrow \mathbf{D}$ where $\mathbf{x}_1, \dots, \mathbf{x}_n$ are distinct and include the variables free in $\mathbf{D}$.

(ii) T_i is obtained from T_{i-1} by the adjunction of an n-ary function symbol $\mathbf{f}$ and a new axiom $\mathbf{y} = \mathbf{f}\mathbf{x}_1 \dots \mathbf{x}_n \leftrightarrow \mathbf{D}$ where $\mathbf{x}_1, \dots, \mathbf{x}_n, \mathbf{y}$ are distinct and include the variables free in $\mathbf{D}$, $\vdash_{T_{i-1}} \exists \mathbf{y}\mathbf{D}$, and $\vdash_{T_{i-1}} \mathbf{D} \to \mathbf{D}[\mathbf{y}|\mathbf{y}'] \to \mathbf{y} = \mathbf{y}'$ for some variable $\mathbf{y}'$ distinct from $\mathbf{x}_1, \dots, \mathbf{x}_n, \mathbf{y}$.

If $\mathbf{A}$ is a formula of T', we can build a formula $\mathbf{A}^*$ of T, called a *translation of* $\mathbf{A}$ *into* T, by successive applications of the constructions of the previous paragraphs, and we have the following result:

THEOREM. Let T' be an extension by definitions of a first-order theory T, $\mathbf{A}$ a formula of T', and $\mathbf{A}^*$ and $\mathbf{A}^\circ$ translations of $\mathbf{A}$ into T. Then

(i) T' is a conservative extension of T;

(ii) $\vdash_T \mathbf{A}^*$ if and only if $\vdash_{T'} \mathbf{A}$; and

(iii) $\vdash_T \mathbf{A}^* \leftrightarrow \mathbf{A}^\circ$.

Proof. The fact that T' is a conservative extension of T follows from the transitivity of conservative extensions. By the theorems on definitions and the tautology theorem, we have

$$\vdash_{T'} \mathbf{A} \leftrightarrow \mathbf{A}^* \quad \text{and} \quad \vdash_{T'} \mathbf{A} \leftrightarrow \mathbf{A}^\circ. \tag{9}$$

Suppose $\vdash_T \mathbf{A}^*$. Then $\vdash_{T'} \mathbf{A}^*$, whence $\vdash_{T'} \mathbf{A}$ by (9) and the tautology theorem. Conversely, suppose $\vdash_{T'} \mathbf{A}$. Then $\vdash_{T'} \mathbf{A}^*$ by (9) and the tautology theorem, whence $\vdash_T \mathbf{A}^*$ by (i). This proves (ii). Finally, we have $\vdash_{T'} \mathbf{A}^* \leftrightarrow \mathbf{A}^\circ$ by (9) and the tautology theorem, whence (iii) from (i). □

2.4 Definitions in practice. Extensions by definitions provide a completely formal way of defining functions and predicates in first-order theories. When actually working in a first-order theory T, however, it becomes quickly laborious to keep track of all the extensions by definitions we have introduced so far. Fortunately, this is not necessary if we agree that whenever a symbol has been introduced with a defining axiom, thereby forming an extension by definitions of T, any subsequent occurrence of that symbol must be understood as being taken in a suitable extension by definitions with the same defining axiom. Explicitely, assume we extend T to T' by defining a symbol $\mathbf{s}$ with a certain defining axiom. If $\mathbf{B}$ is a formula of T', we should write $\vdash_{T'} \mathbf{B}$ to mean that $\mathbf{B}$ is a theorem of T'. Now since the defining axiom for $\mathbf{s}$ has been fixed, there is no confusion in writing $\vdash_T \mathbf{B}$ instead. This can be taken to mean that $\mathbf{B}$ is a theorem in any extension by definitions of T in which $\mathbf{s}$ is defined. Since any formula of any extension by definitions of T can be translated in any other extension by definitions of T, any theorem derived in an extension by definitions of T may be viewed as a theorem of any extension by definitions of T.

If T' is an extension by definitions of T, a symbol of T' is called a *defined symbol* of T, and a formula of T' a *defined formula* of T. These definitions are only useful when the defining axioms for the symbols have been fixed, otherwise any symbol or formula would satisfy them. Thus if we introduce a function symbol $\mathbf{f}$ with a new axiom, we often say that $\mathbf{f}$ is a defined symbol to mean that the associated axiom is a valid defining axiom, i.e., that existence and uniqueness conditions can be derived.

In practice, we often define infinitely many new symbols in a first-order theory T. Since in any given context at most finitely many of them can appear, any theorem of the first-order theory obtained from T by the definitions of all these symbols is a theorem of some extension by definitions of T.

§3 Interpretations

3.1 Interpretations. Let L and M be first-order languages. An *interpretation I of L in M* consists of:

(i) a unary predicate symbol of M, abbreviated by U_I;

(ii) for each n-ary function symbol $\mathbf{f}$ of L, an n-ary function symbol of M, abbreviated by $\mathbf{f}_I$;

(iii) for each n-ary predicate symbol $\mathbf{p}$ of L, an n-ary predicate symbol of M, abbreviated by $\mathbf{p}_I$.[†]

The predicate symbol U_I is called the *universe* of the interpretation I.

Let I be an interpretation of L in M. If $\mathbf{a}$ is a term of L, we abbreviate by $\mathbf{a}_I$ the term of M obtained from $\mathbf{a}$ by replacing each function symbol $\mathbf{f}$ by $\mathbf{f}_I$. The term $\mathbf{a}_I$ is called the *interpretation of* $\mathbf{a}$ *by I*. For any formula $\mathbf{A}$ of L, we define a formula $\mathbf{A}_I$ of M by induction on the length of $\mathbf{A}$. If $\mathbf{A}$ is $\mathbf{p}\mathbf{a}_1 \ldots \mathbf{a}_n$, then $\mathbf{A}_I$ is $\mathbf{p}_I(\mathbf{a}_1)_I \ldots (\mathbf{a}_n)_I$. If $\mathbf{A}$ is $\mathbf{B} \vee \mathbf{C}$, then $\mathbf{A}_I$ is $\mathbf{B}_I \vee \mathbf{C}_I$. If $\mathbf{A}$ is $\neg\mathbf{B}$, then $\mathbf{A}_I$ is $\neg\mathbf{B}_I$. If $\mathbf{A}$ is $\exists\mathbf{x}\mathbf{B}$, then $\mathbf{A}_I$ is $\exists\mathbf{x}(\mathrm{U}_I\mathbf{x} \wedge \mathbf{B}_I)$. If $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are the variables free in $\mathbf{A}$ in reverse alphabetical order, we let $\mathbf{A}^I$ abbreviate $\mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathbf{A}_I$, and we call $\mathbf{A}^I$ the *interpretation of* $\mathbf{A}$ *by I*.

Let L be a first-order language and U a first-order theory. An interpretation of L in $L(U)$ is an *interpretation of L in U* if $\vdash_U \exists x \mathrm{U}_I x$ and $\vdash_U \mathrm{U}_I x_1 \to \cdots \to \mathrm{U}_I x_n \to \mathrm{U}_I \mathbf{f}_I x_1 \ldots x_n$ for each n-ary function symbol $\mathbf{f}$ of L.

Let T and U be first-order theories. An interpretation of $L(T)$ in U is an *interpretation of T in U* if the interpretations by I of the identity axioms, equality axioms, and nonlogical axioms of T are theorems of U. Note that if $=_I$ is $=$, then the interpretations by I of identity and equality axioms are tautological consequences of identity and equality axioms of U, so in that case the first two conditions are automatically satisfied.

3.2 The interpretation theorem. In this paragraph we prove a finitary version of that result which in model theory is often called the soundness of first-order logic.

Lemma 1. Let L be a first-order language, U a first-order theory, and I an interpretation of L in U. If $\mathbf{x}_1, \ldots, \mathbf{x}_n$ include the variables occurring in a term $\mathbf{a}$ of L, in any order, then $\vdash_U \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathrm{U}_I\mathbf{a}_I$.

[†] To be completely general (and such generality is relevant), one should allow an interpretation to have *parameters*. Specifically, an interpretation with k parameters ($k \geq 0$) would consist of a $(k+1)$-ary U_I and $(k+n)$-ary $\mathbf{f}_I$'s and $\mathbf{p}_I$'s, as well as a k-ary predicate symbol Ω_I acting as the "parameter space". These parameters introduce no essential difficulty in the contents of this section, but they complicate the exposition considerably, which is why we have restricted it to the 0-parameter case.

Proof. We prove the lemma by induction on the length of $\mathbf{a}$. If $\mathbf{a}$ is a variable, then the result is a tautology. Suppose that $\mathbf{a}$ is $\mathbf{f}\mathbf{b}_1 \dots \mathbf{b}_n$. Then $\vdash_U \mathrm{U}_I(\mathbf{b}_1)_I \to \cdots \to \mathrm{U}_I(\mathbf{b}_n)_I \to \mathrm{U}_I\mathbf{a}_I$ by definition of an interpretation of L in U and the substitution rule. By induction hypothesis, $\mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathrm{U}_I(\mathbf{b}_i)_I$ for each i. The conclusion is a tautological consequence of the above formulae. □

LEMMA 2. Let T and U be first-order theories, let I be an interpretation of T in U, and let $\mathbf{A}$ be a formula of $L(T)$. If $\mathbf{x}_1, \dots, \mathbf{x}_n$ include the variables free in $\mathbf{A}$ and if $\vdash_U \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathbf{A}_I$, then $\vdash_U \mathbf{A}^I$.

Proof. The formula $\mathrm{U}_I\mathbf{y}_1 \to \cdots \to \mathrm{U}_I\mathbf{y}_k \to \mathbf{A}^I$, where $\mathbf{y}_1, \dots, \mathbf{y}_k$ include the variables among $\mathbf{x}_1, \dots, \mathbf{x}_n$ which are not free in $\mathbf{A}$, is a tautological consequence of $\mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathbf{A}_I$. Since $\mathbf{y}_1, \dots, \mathbf{y}_k$ are not free in $\mathbf{A}^I$, we get $\vdash_U \exists\mathbf{y}_1\mathrm{U}_I\mathbf{y}_1 \to \cdots \to \mathrm{U}_I\mathbf{y}_k \to \mathbf{A}^I$ by the $\exists$-introduction rule, and since $\vdash_U \exists x \mathrm{U}_I x$ we obtain $\vdash_U \mathrm{U}_I\mathbf{y}_2 \to \cdots \to \mathrm{U}_I\mathbf{y}_k \to \mathbf{A}^I$ by the substitution rule and the detachment rule. We repeat this derivation $k-1$ more times and we obtain $\vdash_U \mathbf{A}^I$ as desired. □

INTERPRETATION THEOREM. Let T and U be first-order theories. If I is an interpretation of T in U, then the interpretation by I of any theorem of T is a theorem of U.

Proof. Let $\mathbf{A}$ be a theorem of T. We prove $\vdash_U \mathbf{A}^I$ by tautological induction on theorems in T. Suppose that $\mathbf{A}$ is $\mathbf{B}[\mathbf{x}|\mathbf{a}] \to \exists\mathbf{x}\mathbf{B}$, and let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be the variables free in $\mathbf{A}$. By Lemma 1, we have $\vdash_U \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathrm{U}_I\mathbf{a}_I$. By the substitution theorem, $\vdash_U \mathrm{U}_I\mathbf{a}_I \wedge \mathbf{B}_I[\mathbf{x}|\mathbf{a}_I] \to \exists\mathbf{x}(\mathrm{U}_I\mathbf{x} \wedge \mathbf{B}_I)$. From these we obtain $\vdash_U \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathbf{A}_I$ by the tautology theorem, whence $\vdash_U \mathbf{A}^I$ by Lemma 2. If $\mathbf{A}$ is an identity or equality axiom, then $\vdash_U \mathbf{A}^I$ by definition of an interpretation of T in U.

Suppose that $\mathbf{A}$ is a tautological consequence of $\mathbf{B}_1, \dots, \mathbf{B}_n$. By Proposition 2 of ch. I §3.3, $\mathbf{A}_I$ is a tautological consequence of $(\mathbf{B}_1)_I, \dots, (\mathbf{B}_n)_I$. If $\mathbf{x}_1, \dots, \mathbf{x}_n$ denote the variables free in $\mathbf{A}, \mathbf{B}_1, \dots, \mathbf{B}_n$, then $\mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathbf{A}_I$ is a tautological consequence of $\mathbf{B}_1^I, \dots, \mathbf{B}_n^I$; hence $\vdash_U \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathbf{A}_I$ by the induction hypothesis and the tautology theorem. By Lemma 2, we obtain $\vdash_U \mathbf{A}^I$. Finally, suppose that $\mathbf{A}$ is inferred from $\mathbf{B} \to \mathbf{C}$ by the $\exists$-introduction rule with the variable $\mathbf{x}$. Denote by $\mathbf{x}_1, \dots, \mathbf{x}_n$ the variables free in $\mathbf{A}$. Since $\mathbf{x}$ is not free in $\mathbf{C}$, $\mathbf{x}$ is not among $\mathbf{x}_1, \dots, \mathbf{x}_n$. By the induction hypothesis and the tautology theorem, we have $\vdash_U \mathrm{U}_I\mathbf{x} \wedge \mathbf{B}_I \to \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathbf{C}_I$, whence $\exists\mathbf{x}(\mathrm{U}_I\mathbf{x} \wedge \mathbf{B}_I) \to \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathbf{C}_I$ by the $\exists$-introduction rule. By the latter and the tautology theorem, we obtain $\vdash_U \mathbf{A}^I$. □

The following corollary will also be referred to as the interpretation theorem.

COROLLARY. Let T and U be first-order theories. Suppose that there is an interpretation of an extension of T in a conservative extension of U. If T is inconsistent, then U is inconsistent.

Proof. Let T' be an extension of T and U' a conservative extension of U with an interpretation I of T' in U'. Suppose that T is inconsistent. Then T' is inconsistent. Let $\mathbf{A}$ be $\forall x(x = x)$. By the interpretation theorem, $\mathbf{A}^I$ and $(\neg\mathbf{A})^I$ are theorems of U'. But since $\mathbf{A}$ is closed, $(\neg\mathbf{A})^I$ is $\neg\mathbf{A}^I$. By a remark in §1.1, it follows that U' is inconsistent. Hence U is inconsistent. □

3.3 Interpretations and definitions. Let T and U be first-order theories and let I be an interpretation of $L(T)$ in U. Suppose that there exists a constant $\mathbf{e}$ in an extension by definitions U'' of U. For any extension by definitions T' of T, we shall define an extension U' of U. We first do this in the case where T' is obtained from T by the adjunction of one new symbol and one new nonlogical axiom (if T' is T, let U' be U). If T' is obtained from T by the adjunction of an n-ary predicate symbol $\mathbf{p}$ and the nonlogical axiom $\mathbf{p}\mathbf{x}_1 \dots \mathbf{x}_n \leftrightarrow \mathbf{D}$, we let U' be obtained from U by the adjunction of a new n-ary predicate symbol $\mathbf{p}'$ and the nonlogical axiom $\mathbf{p}'\mathbf{x}_1 \dots \mathbf{x}_n \leftrightarrow \mathbf{D}_I$. If T' is obtained from T by the adjunction of an n-ary function symbol $\mathbf{f}$ and the nonlogical axiom $\mathbf{y} = \mathbf{f}\mathbf{x}_1 \dots \mathbf{x}_n \leftrightarrow \mathbf{D}$, we let U' be obtained from U by the adjunction of a new n-ary function symbol $\mathbf{f}'$ and as a new nonlogical axiom a translation of

$$\mathbf{y} = \mathbf{f}'\mathbf{x}_1 \dots \mathbf{x}_n \leftrightarrow ((\mathrm{U}_I\mathbf{x}_1 \wedge \cdots \wedge \mathrm{U}_I\mathbf{x}_n) \wedge \mathrm{U}_I\mathbf{y} \wedge \mathbf{D}_I) \vee \neg(\mathrm{U}_I\mathbf{x}_1 \wedge \cdots \wedge \mathrm{U}_I\mathbf{x}_n) \wedge \mathbf{y} = \mathbf{e}$$

into U. Finally, if T' is any extension by definitions of T, we let U' be obtained from U by repeated applications of the above constructions.

We define an interpretation I' of $L(T')$ in $L(U')$ by letting $\mathrm{U}_{I'}$ be U_I and $\mathbf{s}_{I'}$ be $\mathbf{s}_I$ or $\mathbf{s}'$ according to $\mathbf{s}$ being a symbol of $L(T)$ or not. An interpretation I' defined in this way is called an *extension of I to $L(T')$*.

INTERPRETATION EXTENSION THEOREM. With the notations of this paragraph, I' is an interpretation of $L(T')$ in U'. Moreover, if I is an interpretation of T in U, then U' is an extension by definitions of U and I' is an interpretation of T' in U'.

Proof. It suffices to consider the special case where T' is obtained from T by the definition of a single symbol. Since $\vdash_U \exists x \mathrm{U}_I x$, we have $\vdash_{U'} \exists x \mathrm{U}_{I'} x$. Suppose that the defined symbol is the function symbol $\mathbf{f}$. From the defining axiom of $\mathbf{f}'$ and the equality theorem, we have $\vdash_{U''} (\mathrm{U}_I\mathbf{x}_1 \wedge \cdots \wedge \mathrm{U}_I\mathbf{x}_n \wedge \mathrm{U}_I\mathbf{f}'\mathbf{x}_1 \ldots \mathbf{x}_n \wedge \mathbf{D}_I) \vee \neg(\mathrm{U}_I\mathbf{x}_1 \wedge \cdots \wedge \mathrm{U}_I\mathbf{x}_n) \wedge \mathbf{f}'\mathbf{x}_1 \ldots \mathbf{x}_n = \mathbf{e}$. A tautological consequence of an instance of this formula is $\mathrm{U}_{I'}x_1 \to \cdots \to \mathrm{U}_{I'}x_n \to \mathrm{U}_{I'}\mathbf{f}'x_1 \ldots x_n$, which proves that I' is an interpretation of $L(T')$ in U'.

We now assume that I is an interpretation of T in U. To prove that U' is an extension by definitions of U, we must verify, in case the new symbol is a function symbol $\mathbf{f}$ with defining axiom $\mathbf{y} = \mathbf{f}\mathbf{x}_1 \ldots \mathbf{x}_n \leftrightarrow \mathbf{D}$, that existence and uniqueness conditions for $\mathbf{y}$ in $(\mathrm{U}_I\mathbf{x}_1 \wedge \cdots \wedge \mathrm{U}_I\mathbf{x}_n \wedge \mathrm{U}_I\mathbf{y} \wedge \mathbf{D}_I) \vee \neg(\mathrm{U}_I\mathbf{x}_1 \wedge \cdots \wedge \mathrm{U}_I\mathbf{x}_n) \wedge \mathbf{y} = \mathbf{e}$ are theorems of U'' (for it is easy to see that existence and uniqueness conditions for a given translation are translations of existence and uniqueness conditions; hence the former will be theorems of U). By the Propositions 1 and 2 of §2.2, it will suffice to prove

$$\vdash_U \mathrm{U}_I\mathbf{x}_1 \wedge \cdots \wedge \mathrm{U}_I\mathbf{x}_n \to \exists \mathbf{y}(\mathrm{U}_I\mathbf{y} \wedge \mathbf{D}_I) \text{ and} \tag{1}$$

$$\vdash_U \mathrm{U}_I\mathbf{x}_1 \wedge \cdots \wedge \mathrm{U}_I\mathbf{x}_n \to \mathrm{U}_I\mathbf{y} \wedge \mathbf{D}_I \to \mathrm{U}_I\mathbf{y}' \wedge \mathbf{D}_I[\mathbf{y}|\mathbf{y}'] \to \mathbf{y} = \mathbf{y}' \tag{2}$$

for some suitable $\mathbf{y}'$. The interpretation of the existence condition for $\mathbf{y}$ in $\mathbf{D}$ is $\mathrm{U}_I\mathbf{x}_1' \to \cdots \to \mathrm{U}_I\mathbf{x}_k' \to \exists \mathbf{y}(\mathrm{U}_I\mathbf{y} \wedge \mathbf{D}_I)$, where $\mathbf{x}_1', \ldots, \mathbf{x}_k'$ are the variables free in $\exists \mathbf{y}\mathbf{D}$ in reverse alphabetical order. Since I is an interpretation of T in U, this formula is a theorem of U by the interpretation theorem. Since $\mathbf{x}_1', \ldots, \mathbf{x}_k'$ are among $\mathbf{x}_1, \ldots, \mathbf{x}_n$, we obtain (1) by the tautology theorem. Similarly, the interpretation of a uniqueness condition for $\mathbf{y}$ in $\mathbf{D}$ is $\mathrm{U}_I\mathbf{x}_1' \to \cdots \to \mathrm{U}_I\mathbf{x}_k' \to \mathbf{D}_I \wedge \mathbf{D}_I[\mathbf{y}|\mathbf{y}'] \to \mathbf{y} = \mathbf{y}'$, where $\mathbf{x}_1', \ldots, \mathbf{x}_k'$ are the variables free in $\mathbf{D} \wedge \mathbf{D}[\mathbf{y}|\mathbf{y}']$ in reverse alphabetical order, of which (2) is a tautological consequence.

It remains to prove that the interpretation by I' of the new nonlogical axiom of T' is a theorem of U'. Suppose that this axiom is $\mathbf{p}\mathbf{x}_1 \ldots \mathbf{x}_n \leftrightarrow \mathbf{D}$. Its interpretation by I' is $\mathrm{U}_I\mathbf{x}_1' \to \cdots \to \mathrm{U}_I\mathbf{x}_n' \to \mathbf{p}'\mathbf{x}_1 \ldots \mathbf{x}_n \leftrightarrow \mathbf{D}_I$, where $\mathbf{x}_1', \ldots, \mathbf{x}_n'$ are $\mathbf{x}_1, \ldots, \mathbf{x}_n$ in reverse alphabetical order. This is tautological consequence of the defining axiom of $\mathbf{p}'$. Suppose that the new nonlogical axiom of T' is $\mathbf{y} = \mathbf{f}\mathbf{x}_1 \ldots \mathbf{x}_n \leftrightarrow \mathbf{D}$. Its interpretation by I' is $\mathrm{U}_I\mathbf{x}_1' \to \cdots \to \mathrm{U}_I\mathbf{x}_{n+1}' \to \mathbf{y} = \mathbf{f}'\mathbf{x}_1 \ldots \mathbf{x}_n \leftrightarrow \mathbf{D}_I$, which is again a tautological consequence of the defining axiom of $\mathbf{f}'$. □

Remark. Following the considerations in §2.4, if I is an interpretation of T in U and if there is a constant in some extension by definitions of U, then in practice we fix such a constant and we continue to write I for any extension I' of I to an extension by definitions T' of T. This is possible if moreover we agree that for any defined symbol $\mathbf{s}$ of T (to which, we recall, a defining axiom is assigned), we use the same symbol $\mathbf{s}'$ in forming U' for any extension by definitions T' of T in which $\mathbf{s}$ is defined.

3.4 Isomorphisms of interpretations. Let L be a first-order language, U a first-order theory, and I and J interpretations of L in U. A unary function symbol $\mathbf{g}$ of U is called an *isomorphism* from I to J if

(i) $\vdash_U \mathrm{U}_J y \leftrightarrow \exists x(\mathrm{U}_I x \wedge y = \mathbf{g}x)$;

(ii) for each n-ary function symbol $\mathbf{f}$ of L, $\vdash_U \mathrm{U}_I x_1 \to \cdots \to \mathrm{U}_I x_n \to \mathbf{g}\mathbf{f}_I x_1 \ldots x_n = \mathbf{f}_J \mathbf{g}x_1 \ldots \mathbf{g}x_n$;

(iii) for each n-ary predicate symbol $\mathbf{p}$ of L, $\vdash_U \mathrm{U}_I x_1 \to \cdots \to \mathrm{U}_I x_n \to \mathbf{p}_I x_1 \ldots x_n \leftrightarrow \mathbf{p}_J \mathbf{g}x_1 \ldots \mathbf{g}x_n$.

ISOMORPHISM EXTENSION THEOREM. Let L be a first-order language, U a first-order theory, and I and J interpretations of L in U. Suppose that some unary function symbol $\mathbf{g}$ of U is an isomorphism from I to J. For any term $\mathbf{a}$ of L, if $\mathbf{x}_1, \ldots, \mathbf{x}_n$ include the variables occurring in $\mathbf{a}$, then $\vdash_U \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathbf{g}\mathbf{a}_I = \mathbf{a}_J[\mathbf{x}_1, \ldots, \mathbf{x}_n|\mathbf{g}\mathbf{x}_1, \ldots, \mathbf{g}\mathbf{x}_n]$. For any formula $\mathbf{A}$ of L, if $\mathbf{x}_1, \ldots, \mathbf{x}_n$ include the variables free in $\mathbf{A}$, then $\vdash_U \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathbf{A}_I \leftrightarrow \mathbf{A}_J[\mathbf{x}_1, \ldots, \mathbf{x}_n|\mathbf{g}\mathbf{x}_1, \ldots, \mathbf{g}\mathbf{x}_n]$.

Proof. We prove the first assertion by induction on the length of $\mathbf{a}$. If $\mathbf{a}$ is a variable, the result is a tautological consequence of an instance of an identity axiom. Suppose that $\mathbf{a}$ is $\mathbf{f}\mathbf{b}_1 \ldots \mathbf{b}_k$. By Lemma 1 of §3.2, $\vdash_U \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathrm{U}_I(\mathbf{b}_i)_I$ for each i. As a tautological consequence of these formulae and of $\mathrm{U}_I(\mathbf{b}_1)_I \to \cdots \to \mathrm{U}_I(\mathbf{b}_k)_I \to \mathbf{g}\mathbf{a}_I = \mathbf{f}_J\mathbf{g}(\mathbf{b}_1)_I \ldots \mathbf{g}(\mathbf{b}_k)_I$, which is an instance of (ii), we obtain

$$\vdash_U \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathbf{g}\mathbf{a}_I = \mathbf{f}_J\mathbf{g}(\mathbf{b}_1)_I \ldots \mathbf{g}(\mathbf{b}_k)_I. \tag{3}$$

By induction hypothesis, $\vdash_U \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathbf{g}(\mathbf{b}_i)_I = (\mathbf{b}_i)_J[\mathbf{x}_1,\ldots,\mathbf{x}_n|\mathbf{g}\mathbf{x}_1,\ldots,\mathbf{g}\mathbf{x}_n]$. From these and (3), we obtain the desired result by the equality theorem and the tautology theorem.

We now turn to the proof of the second assertion, and we proceed by induction on the length of $\mathbf{A}$. Suppose that $\mathbf{A}$ is $\mathbf{p}\mathbf{a}_1\ldots\mathbf{a}_k$. By (iii) and the substitution rule, $\vdash_U \mathrm{U}_I(\mathbf{a}_1)_I \to \cdots \to \mathrm{U}_I(\mathbf{a}_k)_I \to \mathbf{p}_I(\mathbf{a}_1)_I\ldots(\mathbf{a}_k)_I \leftrightarrow \mathbf{p}_J\mathbf{g}(\mathbf{a}_1)_I\ldots\mathbf{g}(\mathbf{a}_k)_I$. Using as above Lemma 1 of §3.2 and the tautology theorem, we obtain

$$\vdash_U \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathbf{p}_I(\mathbf{a}_1)_I\ldots(\mathbf{a}_k)_I \leftrightarrow \mathbf{p}_J\mathbf{g}(\mathbf{a}_1)_I\ldots\mathbf{g}(\mathbf{a}_k)_I. \tag{4}$$

Then the desired result follows from the first result and (4) using the equality theorem and the tautology theorem. Suppose that $\mathbf{A}$ is $\mathbf{B}\vee\mathbf{C}$ or $\neg\mathbf{B}$. In both cases, the result follows from the induction hypothesis and the tautology theorem. Finally, suppose that $\mathbf{A}$ is $\exists\mathbf{x}\mathbf{B}$. By the induction hypothesis,

$$\vdash_U \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathrm{U}_I\mathbf{x} \to \mathbf{B}_I \leftrightarrow \mathbf{B}_J[\mathbf{x}_1,\ldots,\mathbf{x}_n,\mathbf{x}|\mathbf{g}\mathbf{x}_1,\ldots,\mathbf{g}\mathbf{x}_n,\mathbf{g}\mathbf{x}],$$

whence

$$\vdash_U \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \exists\mathbf{x}(\mathrm{U}_I\mathbf{x}\wedge\mathbf{B}_I) \leftrightarrow \exists\mathbf{x}(\mathrm{U}_I\mathbf{x}\wedge\mathbf{B}_J[\mathbf{x}_1,\ldots,\mathbf{x}_n,\mathbf{x}|\mathbf{g}\mathbf{x}_1,\ldots,\mathbf{g}\mathbf{x}_n,\mathbf{g}\mathbf{x}])$$

by the tautology theorem, the deduction theorem, and the distribution rule. Now $\exists\mathbf{x}(\mathrm{U}_I\mathbf{x}\wedge\mathbf{B}_I)$ is $\mathbf{A}_I$, so by the equivalence theorem it remains to show that

$$\vdash_U \exists\mathbf{x}(\mathrm{U}_I\mathbf{x}\wedge\mathbf{B}_J[\mathbf{x}_1,\ldots,\mathbf{x}_n,\mathbf{x}|\mathbf{g}\mathbf{x}_1,\ldots,\mathbf{g}\mathbf{x}_n,\mathbf{g}\mathbf{x}]) \leftrightarrow \mathbf{A}_J[\mathbf{x}_1,\ldots,\mathbf{x}_n|\mathbf{g}\mathbf{x}_1,\ldots,\mathbf{g}\mathbf{x}_n]. \tag{5}$$

Let $\mathbf{z}$ be distinct from $\mathbf{x}_1$, …, $\mathbf{x}_n$, and $\mathbf{x}$, and let $\mathbf{B}'$ be $\mathbf{B}_J[\mathbf{x}_1,\ldots,\mathbf{x}_n|\mathbf{g}\mathbf{x}_1,\ldots,\mathbf{g}\mathbf{x}_n]$. By the replacement theorem and the equivalence theorem,

$$\vdash_U \exists\mathbf{x}(\mathrm{U}_I\mathbf{x}\wedge\mathbf{B}'[\mathbf{x}|\mathbf{g}\mathbf{x}]) \leftrightarrow \exists\mathbf{x}(\mathrm{U}_I\mathbf{x}\wedge\exists\mathbf{z}(\mathbf{z}=\mathbf{g}\mathbf{x}\wedge\mathbf{B}'[\mathbf{x}|\mathbf{z}])). \tag{6}$$

By prenex operations and ch. I §4.1 (iii),

$$\vdash_U \exists\mathbf{x}(\mathrm{U}_I\mathbf{x}\wedge\exists\mathbf{z}(\mathbf{z}=\mathbf{g}\mathbf{x}\wedge\mathbf{B}'[\mathbf{x}|\mathbf{z}])) \leftrightarrow \exists\mathbf{z}(\exists\mathbf{x}(\mathrm{U}_I\mathbf{x}\wedge\mathbf{z}=\mathbf{g}\mathbf{x})\wedge\mathbf{B}'[\mathbf{x}|\mathbf{z}]). \tag{7}$$

By a version of (i) and the equivalence theorem,

$$\vdash_U \exists\mathbf{z}(\exists\mathbf{x}(\mathrm{U}_I\mathbf{x}\wedge\mathbf{z}=\mathbf{g}\mathbf{x})\wedge\mathbf{B}'[\mathbf{x}|\mathbf{z}]) \leftrightarrow \exists\mathbf{z}(\mathrm{U}_J\mathbf{z}\wedge\mathbf{B}'[\mathbf{x}|\mathbf{z}]). \tag{8}$$

By (6), (7), (8), and the tautology theorem, $\vdash_U \exists\mathbf{x}(\mathrm{U}_I\mathbf{x}\wedge\mathbf{B}'[\mathbf{x}|\mathbf{g}\mathbf{x}]) \leftrightarrow \exists\mathbf{z}(\mathrm{U}_J\mathbf{z}\wedge\mathbf{B}'[\mathbf{x}|\mathbf{z}])$, which is a variant of (5). ☐

Let I and J be interpretations of L in U. We say that I is *isomorphic to J in U* when there is an isomorphism from I to J in an extension by definitions of U. It can be proved without difficulty that this is an equivalence relation among the interpretations of L in U, but we shall not use this fact.

3.5 Inner interpretations and absoluteness. Let L be a first-order language, U a first-order theory, and I an interpretation of L in U. An n-ary function symbol $\mathbf{f}$ of U is said to be *I-invariant* if $\vdash_U \mathrm{U}_I x_1 \to \cdots \to \mathrm{U}_I x_n \to \mathrm{U}_I\mathbf{f}x_1\ldots x_n$. By definition of an interpretation of L in U, $\mathbf{f}_I$ is I-invariant for any function symbol $\mathbf{f}$ of L.

In this paragraph we shall discuss a special kind of interpretations often encountered in practice, namely interpretations of a first-order language L in a first-order language M which is an extension of L. We call such interpretations *inner interpretations* if moreover $=_I$ is $=$. In this setting new questions arise, for it is possible to compare in M designators of L with their interpretations. An even more special case is that of an inner interpretation I of L in M such that $\mathbf{s}_I$ is $\mathbf{s}$ for any nonlogical symbol $\mathbf{s}$ of L. Such an interpretation I, which is completely defined by its universe, is called *simple*. If $\mathbf{q}$ is a unary predicate symbol of M, the simple interpretation of L in M whose universe is $\mathbf{q}$ is called the simple interpretation *defined by* $\mathbf{q}$, or simply, by abuse, the *simple interpretation* $\mathbf{q}$.

Let L be a first-order language, U a first-order theory such that $L(U)$ is an extension of L, and I an interpretation of L in U such that $=_I$ is $=$. A term $\mathbf{a}$ of L is *absolute for I* if $\vdash_U \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathbf{a}=\mathbf{a}_I$, where $\mathbf{x}_1,\ldots,\mathbf{x}_n$ are the variables occurring in $\mathbf{a}$ (the order is irrelevant for the definition by the tautology theorem). A formula $\mathbf{A}$ of L with free variables $\mathbf{x}_1$, …, $\mathbf{x}_n$ is said to be *absolute for I* if $\vdash_U \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathbf{A} \leftrightarrow \mathbf{A}_I$. Thus $\mathbf{a}$ is absolute for I if and only if $\mathbf{y}=\mathbf{a}$ is for some $\mathbf{y}$ not occurring in $\mathbf{a}$. We also say that an n-ary function symbol $\mathbf{f}$ (resp. an n-ary predicate symbol $\mathbf{p}$) of L is absolute for I if the term $\mathbf{f}x_1 \dots x_n$ (resp. the formula $\mathbf{p}x_1 \dots x_n$) is absolute for I. Thus = is absolute for I. A formula $\mathbf{A}$ of L is said to be *complete in* $\mathbf{y}_1, \dots, \mathbf{y}_m$ *for* I if $\vdash_U \mathrm{U}_I\mathbf{x}_1 \to \dots \to \mathrm{U}_I\mathbf{x}_n \to \mathbf{A} \to \mathrm{U}_I\mathbf{y}_1 \wedge \dots \wedge \mathrm{U}_I\mathbf{y}_m$ where $\mathbf{x}_1, \dots, \mathbf{x}_n$ are the variables other than $\mathbf{y}_1, \dots, \mathbf{y}_m$ free in $\mathbf{A}$. For example, $x = y$ is complete in x for I and complete in y for I, but it is complete in x, y for I if and only if $\vdash_U \mathrm{U}_I x$.

Lemma 1. If $\mathbf{f}$ is absolute for I, then $\mathbf{f}$ is I-invariant.

Proof. We have $\vdash_U \mathrm{U}_I x_1 \to \dots \to \mathrm{U}_I x_n \to \mathbf{f}x_1 \dots x_n = \mathbf{f}_I x_1 \dots x_n$ by absoluteness of $\mathbf{f}$ and $\vdash_U \mathrm{U}_I x_1 \to \dots \to \mathrm{U}_I x_n \to \mathrm{U}_I \mathbf{f}_I x_1 \dots x_n$ by definition of an interpretation of L in U. Thus the result follows from the tautology theorem and a version of a predicative equality axiom. □

Lemma 2. If all the nonlogical symbols occurring in $\mathbf{a}$ are absolute for I, then $\mathbf{a}$ is absolute for I.

Proof. By induction on the length of $\mathbf{a}$. The result is a tautological consequence of an identity axiom if $\mathbf{a}$ is a variable. If $\mathbf{a}$ is $\mathbf{f}\mathbf{b}_1 \dots \mathbf{b}_n$ where $\mathbf{f}$ is absolute for I, then by induction hypothesis $\mathbf{b}_1, \dots, \mathbf{b}_n$ are absolute for I. Hence $\mathbf{a}$ is absolute for I by the equality theorem and the tautology theorem. □

Lemma 3. If $\mathbf{A}$ and $\mathbf{B}$ are absolute for I, then $\mathbf{A} \vee \mathbf{B}$, $\mathbf{A} \wedge \mathbf{B}$, $\neg\mathbf{B}$, $\mathbf{A} \to \mathbf{B}$, and $\mathbf{A} \leftrightarrow \mathbf{B}$ are absolute for I.

Proof. It suffices to consider $\mathbf{A} \vee \mathbf{B}$ and $\neg\mathbf{B}$. The condition of absoluteness of $\mathbf{A} \vee \mathbf{B}$ is a tautological consequence of that of $\mathbf{A}$ and that of $\mathbf{B}$. Similarly for $\neg\mathbf{A}$. □

Lemma 4. If $\mathbf{A}$ is open and all the nonlogical symbols occurring in $\mathbf{A}$ are absolute for I, then $\mathbf{A}$ is absolute for I.

Proof. We proceed by induction on the length of $\mathbf{A}$. If $\mathbf{A}$ is $\mathbf{p}\mathbf{a}_1 \dots \mathbf{a}_n$ where $\mathbf{p}$ is absolute for I, then by Lemma 2 $\mathbf{a}_1, \dots, \mathbf{a}_n$ are absolute for I, so $\mathbf{A}$ is absolute for I by the equality theorem and the tautology theorem. If $\mathbf{A}$ is $\mathbf{B} \vee \mathbf{C}$ or $\neg\mathbf{B}$, the result follows from the induction hypothesis and Lemma 3. □

Lemma 5. If $\mathbf{A}$ is absolute for I and complete in $\mathbf{x}$ for I, then $\exists\mathbf{x}\mathbf{A}$ is absolute for I. If $\mathbf{A}$ is absolute for I and if $\neg\mathbf{A}$ is complete in $\mathbf{x}$ for I, then $\forall\mathbf{x}\mathbf{A}$ is absolute for I.

Proof. Let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be the variables free in $\exists\mathbf{x}\mathbf{A}$. We let U' be obtained from U by the adjunction of n new constants $\mathbf{e}_1, \dots, \mathbf{e}_n$ and new axioms $\mathrm{U}_I\mathbf{e}_1, \dots, \mathrm{U}_I\mathbf{e}_n$. We let $\mathbf{A}'$ abbreviate $\mathbf{A}[\mathbf{x}_1, \dots, \mathbf{x}_n | \mathbf{e}_1, \dots, \mathbf{e}_n]$. Then for the first assertion it will suffice, by the deduction theorem, to prove $\vdash_{U'} \exists\mathbf{x}\mathbf{A}' \leftrightarrow \exists\mathbf{x}(\mathrm{U}_I\mathbf{x} \wedge (\mathbf{A}')_I)$. By the hypotheses, the substitution rule, and the detachment rule, we have $\vdash_{U'} \mathbf{A}' \leftrightarrow (\mathbf{A}')_I$ and $\vdash_{U'} \mathbf{A}' \to \mathrm{U}_I\mathbf{x}$, of whom $\mathbf{A}' \leftrightarrow \mathrm{U}_I\mathbf{x} \wedge (\mathbf{A}')_I$ is a tautological consequence. Thus $\vdash_{U'} \exists\mathbf{x}\mathbf{A}' \leftrightarrow \exists\mathbf{x}(\mathrm{U}_I\mathbf{x} \wedge (\mathbf{A}')_I)$ by the distribution rule. The second assertion follows from the first one by Lemma 3. □

In applications, proofs of absoluteness of formulae are thus reduced to proofs of completeness of formulae. Not much more can be said on completeness in the present general setting, for it depends heavily on the nonlogical axioms of U and the definition of I.

Lemma 6. Let $\mathbf{A}$ and $\mathbf{B}$ be formulae of L. If $\mathbf{A}$ and $\mathbf{B}$ are complete in $\mathbf{y}_1, \dots, \mathbf{y}_m$ for I, then $\mathbf{A} \vee \mathbf{B}$ is complete in $\mathbf{y}_1, \dots, \mathbf{y}_m$ for I. If $\mathbf{A}$ is complete in $\mathbf{y}_1, \dots, \mathbf{y}_m$ for I and if $\mathbf{B}$ is complete in $\mathbf{y}'_1, \dots, \mathbf{y}'_k$ for I, then $\mathbf{A} \wedge \mathbf{B}$ is complete in $\mathbf{y}_1, \dots, \mathbf{y}_m, \mathbf{y}'_1, \dots, \mathbf{y}'_k$ for I. If $\mathbf{x}$ is distinct from $\mathbf{y}_1, \dots, \mathbf{y}_m$ and if $\mathbf{A}$ is complete in $\mathbf{x}, \mathbf{y}_1, \dots, \mathbf{y}_m$ for I, then $\exists\mathbf{x}\mathbf{A}$ and $\forall\mathbf{x}\mathbf{A}$ are complete in $\mathbf{y}_1, \dots, \mathbf{y}_m$ for I.

Proof. The first two assertions are applications of the tautology theorem. For the last assertion, let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be the variables other than $\mathbf{x}, \mathbf{y}_1, \dots, \mathbf{y}_m$ free in $\mathbf{A}$. Then $\vdash_U \mathrm{U}_I\mathbf{x}_1 \wedge \dots \wedge \mathrm{U}_I\mathbf{x}_n \wedge \mathbf{A} \to \mathrm{U}_I\mathbf{y}_1 \wedge \dots \wedge \mathrm{U}_I\mathbf{y}_m$ by the hypothesis and the tautology theorem, whence $\vdash_U \mathrm{U}_I\mathbf{x}_1 \to \dots \to \mathrm{U}_I\mathbf{x}_n \to \exists\mathbf{x}\mathbf{A} \to \mathrm{U}_I\mathbf{y}_1 \wedge \dots \wedge \mathrm{U}_I\mathbf{y}_m$ by the $\exists$-introduction rule, prenex operations, and the tautology theorem. The proof for $\forall\mathbf{x}\mathbf{A}$ uses the substitution theorem instead of the $\exists$-introduction rule. □

In the case of a simple interpretation, we have the following very useful criterion.

Lemma 7. Assume that I is simple, and let Γ be a collection of formulae of L such that any subformula of a formula in Γ is in Γ. Then for the formulae of Γ to be absolute for I, it suffices that for any instantiation $\exists\mathbf{y}\mathbf{B}$ of Γ with free variables among $\mathbf{x}_1, \dots, \mathbf{x}_n$, $\vdash_U \mathrm{U}_I\mathbf{x}_1 \to \dots \to \mathrm{U}_I\mathbf{x}_n \to \exists\mathbf{y}\mathbf{B} \to \exists\mathbf{y}(\mathrm{U}_I\mathbf{y} \wedge \mathbf{B})$.

Proof. We proceed by induction on the length of $\mathbf{A}$ in Γ. If $\mathbf{A}$ is atomic, then $\mathbf{A}_I$ is $\mathbf{A}$ because I is simple, and the claim is a tautology. If $\mathbf{A}$ is $\mathbf{B} \vee \mathbf{C}$ or $\neg\mathbf{B}$, then the absoluteness of $\mathbf{A}$ follows tautologically from the induction hypothesis. Suppose that $\mathbf{A}$ is $\exists\mathbf{y}\mathbf{B}$. Then $\mathbf{B}$ is in Γ and by induction hypothesis and the tautology theorem $\vdash_U \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathrm{U}_I\mathbf{y} \to \mathbf{B} \leftrightarrow \mathbf{B}_I$. Using the distribution rule and the tautology theorem with the deduction theorem, we find $\vdash_U \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \exists\mathbf{y}(\mathrm{U}_I\mathbf{y} \wedge \mathbf{B}) \leftrightarrow \exists\mathbf{y}(\mathrm{U}_I\mathbf{y} \wedge \mathbf{B}_I)$. Now $\mathbf{A}_I$ is exactly $\exists\mathbf{y}(\mathrm{U}_I\mathbf{y} \wedge \mathbf{B}_I)$, so by the tautology theorem it will suffice to prove $\vdash_U \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathbf{A} \leftrightarrow \exists\mathbf{y}(\mathrm{U}_I\mathbf{y} \wedge \mathbf{B})$. The implication from right to left is obtained by the tautology theorem and the distribution rule using the deduction theorem. The other implication was assumed. □

We now suppose that U is an extension of T and that I is an interpretation of T in U such that $=_I$ is $=$.

LEMMA 8. If $\vdash_T \mathbf{a} = \mathbf{b}$ and $\mathbf{a}$ is absolute for I, then $\mathbf{b}$ is absolute for I. If $\vdash_T \mathbf{A} \leftrightarrow \mathbf{B}$ and $\mathbf{A}$ is absolute for I, then $\mathbf{B}$ is absolute for I.

Proof. Since I is an interpretation of T, we have $\vdash_U \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathbf{a}_I = \mathbf{b}_I$ (resp. $\vdash_U \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathbf{A}_I \leftrightarrow \mathbf{B}_I$) where $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are the variable occurring in $\mathbf{a}$ and $\mathbf{b}$ (resp. free in $\mathbf{A}$ and $\mathbf{B}$). The result follows by the equality theorem and the tautology theorem (resp. the tautology theorem). □

§4 Herbrand–Skolem theory

4.1 Skolem and Henkin theories. A first-order theory T is called a *Skolem theory* if for every instantiation $\exists\mathbf{x}\mathbf{A}$ of T there exist a term $\mathbf{a}$ of T and a variant $\mathbf{A}'$ of $\mathbf{A}$ such that

(i) the variables occurring in $\mathbf{a}$ occur freely in $\exists\mathbf{x}\mathbf{A}$ and

(ii) $\vdash_T \exists\mathbf{x}\mathbf{A} \to \mathbf{A}'[\mathbf{x}|\mathbf{a}]$.

We say that T is a *Henkin theory* if the previous statement holds when $\exists\mathbf{x}\mathbf{A}$ is a closed instantiation (in this case, condition (i) simply states that $\mathbf{a}$ is closed). Thus a Skolem theory is a Henkin theory, but the converse need not be true. One of the goals of this section will be to prove that any first-order theory has a conservative Skolem extension. In this paragraph we shall only prove that any first-order theory has a conservative Henkin extension. As will be seen, the proof of this fact will rely on the theorem on constants. To prove the more general result, we shall need a generalization of the theorem on constants for function symbols of higher arity: this is the theorem on functional extensions of §4.4.

Let L be a first-order language. We define the *special constants of level* n, for $n \geq 1$, by induction on n. Let $\Gamma_0(L)$ denote the collection of closed instantiations of L. Suppose that, for some $n \geq 0$, the collection $\Gamma_n(L)$ has been described. For every formula $\exists\mathbf{x}\mathbf{A}$ in $\Gamma_n(L)$, we choose a new constant called the *special constant for* $\exists\mathbf{x}\mathbf{A}$; the special constants of level $n+1$ are the special constants for the formulae of $\Gamma_n(L)$. We then let $\Gamma_{n+1}(L)$ consists of the closed instantiations not in $\Gamma_n(L)$ of the language obtained from L by adding the special constants of level at most $n+1$. The first-order language obtained from L by adding all the special constants is denoted by L_c. The *level* of a formula of L_c is 0 if it is a formula of L and is the greatest level of a special constant occurring in it otherwise. We use $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ as syntactical variables varying through special constants. If $\mathbf{i}$ is the special constant for $\exists\mathbf{x}\mathbf{A}$, the *special axiom for* $\mathbf{i}$ is the formula $\exists\mathbf{x}\mathbf{A} \to \mathbf{A}[\mathbf{x}|\mathbf{i}]$.

Let T be a first-order theory. We now describe a Henkin extension T_c of T. Its language is L_c and its nonlogical axioms are those of T and the special axioms for the special constants of L_c. It is obvious that T_c is a Henkin theory. A formula of $L(T_c)$ of the form $\forall\mathbf{x}(\mathbf{A} \leftrightarrow \mathbf{B}) \to \mathbf{i} = \mathbf{j}$, where $\mathbf{i}$ is the special constant for $\exists\mathbf{x}\mathbf{A}$ and $\mathbf{j}$ is the special constant for $\exists\mathbf{x}\mathbf{B}$, is called a *special equality axiom*. We form T_c' from T_c by adding as further nonlogical axioms all the special equality axioms.

LEMMA. Let T be a first-order theory and $\exists\mathbf{x}\mathbf{A}$ a closed formula of T which is a theorem of T. Then the first-order theory obtained from T by the adjunction of a new constant $\mathbf{e}$ and of the axiom $\mathbf{A}[\mathbf{x}|\mathbf{e}]$ is a conservative extension of T.

Proof. Let T' be the extension of T to be proved conservative, and let $\mathbf{B}$ be a formula of T which is a theorem of T'. If $\mathbf{y}$ is a variable not occurring in $\mathbf{A}$ or $\mathbf{B}$, then $\vdash_T \mathbf{A}[\mathbf{x}|\mathbf{y}] \to \mathbf{B}$ by the deduction theorem. By the $\exists$-introduction rule, $\vdash_T \exists\mathbf{y}\mathbf{A}[\mathbf{x}|\mathbf{y}] \to \mathbf{B}$. But $\vdash_T \exists\mathbf{y}\mathbf{A}[\mathbf{x}|\mathbf{y}]$ by the hypothesis, the variant theorem, and the tautology theorem. Hence $\vdash_T \mathbf{B}$ by the detachement rule. □

THEOREM. T_c and T_c' are conservative extensions of T.

Proof. Since T'_c is an extension of T_c, it suffices to prove that T'_c is a conservative extension of T. We first observe that any formula of the form

$$\forall\mathbf{x}(\mathbf{A} \leftrightarrow \mathbf{A}) \to \mathbf{a} = \mathbf{a} \text{ or} \tag{1}$$
$$(\forall\mathbf{x}(\mathbf{A} \leftrightarrow \mathbf{B}) \to \mathbf{a} = \mathbf{b}) \to \forall\mathbf{x}(\mathbf{B} \leftrightarrow \mathbf{A}) \to \mathbf{b} = \mathbf{a} \tag{2}$$

is derivable without nonlogical axioms: (1) is a tautological consequence of an instance of an identity axiom, while (2) is obtained from the tautology $(\mathbf{B} \leftrightarrow \mathbf{A}) \to (\mathbf{A} \leftrightarrow \mathbf{B})$ using the distribution rule, the symmetry theorem, and the tautology theorem.

Let $\mathbf{A}$ be a formula of T which is a theorem of T'_c; then $\mathbf{A}$ is a theorem of some first-order theory obtained from T by the adjunction of finitely many special constants $\mathbf{i}_1, \dots, \mathbf{i}_n$, their special axioms, and the special equality axioms whose right-hand sides are of the form $\mathbf{i}_r = \mathbf{i}_s$. Order these special constants so that the level of $\mathbf{i}_r$ is at least the level of $\mathbf{i}_s$ whenever $s < r$. Then if $\exists\mathbf{x}_k\mathbf{B}_k$ is the formula for which $\mathbf{i}_k$ is the special axiom, $\mathbf{i}_l$ does not occur in $\mathbf{B}_k$ for all $l \geq k$. We choose once and for all a variable $\mathbf{x}_{rs}$ which does not occur in $\exists\mathbf{x}_1\mathbf{B}_1, \dots, \exists\mathbf{x}_n\mathbf{B}_n$, for every pair of indices r, s. For $0 \leq k \leq n$, we designate by T_k the first-order theory whose language is obtained from $L(T)$ by the adjunction of the constants $\mathbf{i}_1, \dots, \mathbf{i}_k$ and whose nonlogical axioms are: the nonlogical axioms of T; the special axioms of $L(T_k)$; and the formulae of $L(T_k)$ of the form $\forall\mathbf{x}_{rs}(\mathbf{B}_r[\mathbf{x}_r|\mathbf{x}_{rs}] \leftrightarrow \mathbf{B}_s[\mathbf{x}_s|\mathbf{x}_{rs}]) \to \mathbf{i}_r = \mathbf{i}_s$ for $s < r$. Then T_0 is T and by the variant theorem, (1), (2), and the tautology theorem, every special equality axiom of $L(T_k)$ is a theorem of T_k. In particular, $\mathbf{A}$ is a theorem of T_n, so it will suffice to prove that for each $k < n$, T_{k+1} is a conservative extension of T_k. Fix $k < n$, and let $\mathbf{i}$, $\mathbf{x}$, $\mathbf{B}$, and $\mathbf{y}_r$ be $\mathbf{i}_{k+1}$, $\mathbf{x}_{k+1}$, $\mathbf{B}_{k+1}$, and $\mathbf{x}_{k+1,r}$. We form a first-order theory U from T_k by the adjunction of the constant $\mathbf{i}$ and of the single axiom

$$(\exists\mathbf{x}\mathbf{B} \to \mathbf{B}[\mathbf{x}|\mathbf{i}]) \wedge (\forall\mathbf{y}_1(\mathbf{B}[\mathbf{x}|\mathbf{y}_1] \leftrightarrow \mathbf{B}_1[\mathbf{x}_1|\mathbf{y}_1]) \to \mathbf{i} = \mathbf{i}_1) \wedge \cdots \wedge (\forall\mathbf{y}_k(\mathbf{B}[\mathbf{x}|\mathbf{y}_k] \leftrightarrow \mathbf{B}_k[\mathbf{x}_k|\mathbf{y}_k]) \to \mathbf{i} = \mathbf{i}_k)$$

which is just the conjunction of the nonlogical axioms which must be added to T_k in order to obtain T_{k+1}. Let $\mathbf{C}[\mathbf{z}|\mathbf{i}]$ be the above formula. By the tautology theorem, U is equivalent to T_{k+1}. Hence, by the lemma, it will suffice to prove $\vdash_{T_k} \exists\mathbf{z}\mathbf{C}$. We first prove

$$\vdash_{T_k} \forall\mathbf{y}_r(\mathbf{B}[\mathbf{x}|\mathbf{y}_r] \leftrightarrow \mathbf{B}_r[\mathbf{x}_r|\mathbf{y}_r]) \to \exists\mathbf{z}\mathbf{C} \tag{3}$$

for $1 \leq r \leq k$. By the deduction theorem, it will suffice to prove that $\exists\mathbf{z}\mathbf{C}$ is a theorem of the theory T' obtained from T_k by the adjunction of the axiom $\forall\mathbf{y}_r(\mathbf{B}[\mathbf{x}|\mathbf{y}_r] \leftrightarrow \mathbf{B}_r[\mathbf{x}_r|\mathbf{y}_r])$. We have $\vdash_{T'} \mathbf{B}[\mathbf{x}|\mathbf{y}_r] \leftrightarrow \mathbf{B}_r[\mathbf{x}_r|\mathbf{y}_r]$ by the closure theorem, whence $\vdash_{T'} \exists\mathbf{x}\mathbf{B} \leftrightarrow \exists\mathbf{x}_r\mathbf{B}_r$ by the distribution rule and the variant theorem and $\vdash_{T'} \mathbf{B}[\mathbf{x}|\mathbf{i}_r] \leftrightarrow \mathbf{B}_r[\mathbf{x}_r|\mathbf{i}_r]$ by the substitution rule. From these and the special axiom for $\mathbf{i}_r$, we obtain

$$\vdash_{T'} \exists\mathbf{x}\mathbf{B} \to \mathbf{B}[\mathbf{x}|\mathbf{i}_r] \tag{4}$$

by the tautology theorem. Let $1 \leq s \leq k$. From $\vdash_{T'} \mathbf{B}[\mathbf{x}|\mathbf{y}_r] \leftrightarrow \mathbf{B}_r[\mathbf{x}_r|\mathbf{y}_r]$ we also obtain $\vdash_{T'} \forall\mathbf{y}_s(\mathbf{B}[\mathbf{x}|\mathbf{y}_s] \leftrightarrow \mathbf{B}_s[\mathbf{x}_s|\mathbf{y}_s]) \leftrightarrow \forall\mathbf{x}_{rs}(\mathbf{B}_r[\mathbf{x}_r|\mathbf{x}_{rs}] \leftrightarrow \mathbf{B}_s[\mathbf{x}_s|\mathbf{x}_{rs}])$ by the tautology theorem, the distribution rule, and the variant theorem, whence

$$\vdash_{T'} \forall\mathbf{y}_s(\mathbf{B}[\mathbf{x}|\mathbf{y}_s] \leftrightarrow \mathbf{B}_s[\mathbf{x}_s|\mathbf{y}_s]) \to \mathbf{i}_r = \mathbf{i}_s \tag{5}$$

by the axioms of $L(T_k)$ and the tautology theorem. From (4) and (5) by the tautology theorem, $\vdash_{T'} \mathbf{C}[\mathbf{z}|\mathbf{i}_r]$, whence $\vdash_{T'} \exists\mathbf{z}\mathbf{C}$ by the substitution axioms. This proves (3). Next we prove

$$\vdash_{T_k} \neg(\forall\mathbf{y}_1(\mathbf{B}[\mathbf{x}|\mathbf{y}_1] \leftrightarrow \mathbf{B}_1[\mathbf{x}_1|\mathbf{y}_1]) \vee \cdots \vee \forall\mathbf{y}_k(\mathbf{B}[\mathbf{x}|\mathbf{y}_k] \leftrightarrow \mathbf{B}_k[\mathbf{x}_k|\mathbf{y}_k])) \to \exists\mathbf{z}\mathbf{C}. \tag{6}$$

Let $\mathbf{w}$ be a variable not occurring in the above formula. By the variant theorem and prenex operations, $\vdash_{T_k} \exists\mathbf{z}(\exists\mathbf{x}\mathbf{B} \to \mathbf{B}[\mathbf{x}|\mathbf{z}])$. Hence by the tautology theorem and the $\forall$-introduction rule,

$$\begin{aligned}\vdash_{T_k} \neg(\forall\mathbf{y}_1(\mathbf{B}[\mathbf{x}|\mathbf{y}_1] \leftrightarrow \mathbf{B}_1[\mathbf{x}_1|\mathbf{y}_1]) \vee \cdots \vee \forall\mathbf{y}_k(\mathbf{B}[\mathbf{x}|\mathbf{y}_k] \leftrightarrow \mathbf{B}_k[\mathbf{x}_k|\mathbf{y}_k])) \to \exists\mathbf{z}(\exists\mathbf{x}\mathbf{B} \to \mathbf{B}[\mathbf{x}|\mathbf{z}])\\ \wedge \forall\mathbf{w}((\forall\mathbf{y}_1(\mathbf{B}[\mathbf{x}|\mathbf{y}_1] \leftrightarrow \mathbf{B}_1[\mathbf{x}_1|\mathbf{y}_1]) \to \mathbf{i} = \mathbf{i}_1) \wedge \cdots \wedge (\forall\mathbf{y}_k(\mathbf{B}[\mathbf{x}|\mathbf{y}_k] \leftrightarrow \mathbf{B}_k[\mathbf{x}_k|\mathbf{y}_k]) \to \mathbf{i} = \mathbf{i}_k)),\end{aligned}$$

from which we obtain (6) by prenex operations, the substitution theorem, the distribution rule, and the tautology theorem. From (3) and (6) we obtain $\vdash_{T_k} \exists\mathbf{z}\mathbf{C}$ by the tautology theorem. ☐

This theorem is a syntactical *principle of choice*: if we imagine a formula $\mathbf{A}$ with one free variable $\mathbf{x}$ as representing the collection of all individuals $\mathbf{x}$ such that $\mathbf{A}$, then the special constants select a particular individual in each such collection. We shall see in §4.4 a considerable generalization of this conservativity result to instantiations $\exists\mathbf{x}\mathbf{A}$ with any number of free variables.

4.2 The consistency theorem. In this paragraph we shall prove a useful criterion for the consistency of first-order theories of a special kind, namely those all of whose nonlogical axioms are universal. If T is such a first-order theory and if T' has the same language as T and has as nonlogical axioms the matrices of the nonlogical axioms of T, then T and T' are equivalent by the closure theorem. Thus we shall not restrict the generality if we assume that the nonlogical axioms of T are open, and this will be technically convenient. A first-order theory whose nonlogical axioms are open will be called *open*.

Let T be a first-order theory. A formula of $L(T_c)$ is said to *belong* to the special constant $\mathbf{i}$ for $\exists\mathbf{x}\mathbf{A}$ if it is the special equality axiom for $\mathbf{i}$ or a closed substitution axiom of $L(T_c)$ of the form $\mathbf{A}[\mathbf{x}|\mathbf{a}] \to \exists\mathbf{x}\mathbf{A}$. We denote by $\Delta(T)$ the collection of formulae of $L(T_c)$ which either belong to some special constant or are closed instances in $L(T_c)$ of equality rules for $L(T)$ or of nonlogical axioms of T. The *rank* of the special constant for $\exists\mathbf{x}\mathbf{A}$ is the number of occurrences of $\exists$ in $\exists\mathbf{x}\mathbf{A}$. We let $\Delta_n(T)$ be the collection of formulae in $\Delta(T)$ which do not belong to special constants of rank $n+1$ or greater.

LEMMA. Let T be a first-order theory and $\mathbf{A}$ a theorem of T. Then any closed instance of $\mathbf{A}$ in $L(T_c)$ is a tautological consequence of formulae in $\Delta(T)$.

Proof. By tautological induction on theorems in T. Let $\mathbf{A}'$ be a closed instance of $\mathbf{A}$ in $L(T_c)$. If $\mathbf{A}$ is a substitution axiom, so is $\mathbf{A}'$, and then $\mathbf{A}'$ is in $\Delta(T)$ and is a tautological consequence of itself. If $\mathbf{A}$ is an identity axiom, equality axiom, or nonlogical axiom of T, then $\mathbf{A}'$ is in $\Delta(T)$ by definition. Suppose that $\mathbf{A}$ is a tautological consequence of $\mathbf{B}_1, \dots, \mathbf{B}_n$. By Proposition 2 of ch. I §3.3, $\mathbf{A}'$ is a tautological consequence of closed instances $\mathbf{B}'_1, \dots, \mathbf{B}'_n$ in $L(T_c)$ of $\mathbf{B}_1, \dots, \mathbf{B}_n$. By induction hypothesis, each $\mathbf{B}'_i$ is a tautological consequence of formulae in $\Delta(T)$, and hence so is $\mathbf{A}'$. Finally, suppose that $\mathbf{A}$ is inferred from $\mathbf{B} \to \mathbf{C}$ by the $\exists$-introduction rule with the variable $\mathbf{x}$. Then $\mathbf{A}'$ has the form $\exists\mathbf{x}\mathbf{B}' \to \mathbf{C}'$ where $\exists\mathbf{x}\mathbf{B}'$ and $\mathbf{C}'$ are closed instances in $L(T_c)$ of $\exists\mathbf{x}\mathbf{B}$ and $\mathbf{C}$. Let $\mathbf{i}$ be the special constant for $\exists\mathbf{x}\mathbf{B}'$. Then $\mathbf{B}'[\mathbf{x}|\mathbf{i}] \to \mathbf{C}'$ is a closed instance of $\mathbf{B} \to \mathbf{C}$, and so by induction hypothesis it is a tautological consequence of formulae in $\Delta(T)$. But then $\mathbf{A}'$ is a tautological consequence of these formulae and the formula $\exists\mathbf{x}\mathbf{B}' \to \mathbf{B}'[\mathbf{x}|\mathbf{i}]$ which is also in $\Delta(T)$. □

Let T be an open first-order theory. We let T^* (resp. T_c^*) be the formal system whose language is $L(T)$ (resp. $L(T_c)$) and whose rules of inference are the propositional rules for $L(T)$ (resp. for $L(T_c)$) and the instances in $L(T)$ (resp. the closed instances in $L(T_c)$) of the equality rules and nonlogical axioms of T. Note that the nonlogical axioms of T_c^* are exactly the formulae in $\Delta_0(T)$.

CONSISTENCY THEOREM. If T is an open first-order theory, the following statements are equivalent:

(i) T is inconsistent;

(ii) T_c^* is inconsistent;

(iii) T^* is inconsistent.

Proof. Suppose that T is inconsistent, and let us prove that T_c^* is inconsistent. Let F be the formal system whose language is $L(T_c)$ and whose rules of inference are the propositional rules. It will suffice to prove that for some formulae $\mathbf{A}_1, \dots, \mathbf{A}_n$ in $\Delta_0(T)$, $F[\mathbf{A}_1, \dots, \mathbf{A}_n]$ is inconsistent. Let $\mathbf{A}$ be a formula in $\Delta_0(T)$. By the lemma, $\neg\mathbf{A}$ is a tautological consequence of formulae $\mathbf{B}_1, \dots, \mathbf{B}_k$ in $\Delta(T)$. Then by the tautology theorem, $F[\mathbf{B}_1, \dots, \mathbf{B}_k, \mathbf{A}]$ is inconsistent. Since all of $\mathbf{B}_1, \dots, \mathbf{B}_k$, and $\mathbf{A}$ are in $\Delta_n(T)$ for some n, the proof will be complete if we can establish the following assertion: if $n \geq 0$ and if $F[\mathbf{A}_1, \dots, \mathbf{A}_k]$ is inconsistent for some formulae $\mathbf{A}_1, \dots, \mathbf{A}_k$ in $\Delta_{n+1}(T)$, then there are formulae $\mathbf{B}_1, \dots, \mathbf{B}_l$ in $\Delta_n(T)$ such that $F[\mathbf{B}_1, \dots, \mathbf{B}_l]$ is inconsistent. We prove the assertion by induction on the number of special constants of rank $n+1$ belonging to $\mathbf{A}_1, \dots, \mathbf{A}_k$. If there are none, then $\mathbf{A}_1, \dots, \mathbf{A}_k$ are already formulae of $\Delta_n(T)$ and there is nothing to prove. Otherwise, let these special constants be $\mathbf{i}_1, \dots, \mathbf{i}_r$, and $\mathbf{i}$ where the level of $\mathbf{i}$ is as great as the levels of $\mathbf{i}_1, \dots, \mathbf{i}_r$. Say $\mathbf{i}$ is the special constant for $\exists\mathbf{x}\mathbf{B}$. Let $\mathbf{C}_1, \dots, \mathbf{C}_p$ be the formulae among $\mathbf{A}_1, \dots, \mathbf{A}_k$ which are in $\Delta_n(T)$ or belong to one of $\mathbf{i}_1, \dots, \mathbf{i}_r$, and let $\mathbf{D}_1, \dots, \mathbf{D}_q$ be those which are substitution axioms belonging to $\mathbf{i}$. Then $\mathbf{D}_j$ is $\mathbf{B}[\mathbf{x}|\mathbf{a}_j] \to \exists\mathbf{x}\mathbf{B}$ for some closed term $\mathbf{a}_j$ of $L(T_c)$.

For $\mathbf{u}$ a designator of $L(T_c)$ and $1 \le j \le q$, we define a designator $\mathbf{u}^{(j)}$ by induction on the greatest level of a special constant in $\mathbf{u}$. If $\mathbf{u}$ contains no special constant, $\mathbf{u}^{(j)}$ is $\mathbf{u}$. If $\mathbf{u}$ is $\mathbf{i}$, then $\mathbf{u}^{(j)}$ is $\mathbf{a}_j$. If $\mathbf{u}$ is the special constant for $\exists\mathbf{y}\mathbf{C}$ and $\mathbf{u}$ is distinct from $\mathbf{i}$, then $\mathbf{u}^{(j)}$ is the special constant for $\exists\mathbf{y}\mathbf{C}^{(j)}$. Otherwise, $\mathbf{u}^{(j)}$ is the expression obtained from $\mathbf{u}$ by replacing every occurrence of a special constant $\mathbf{j}$ by $\mathbf{j}^{(j)}$. We now prove that for $1 \le i \le p$, if $\mathbf{C}_i$ is in $\Delta_n(T)$, so is $(\mathbf{C}_i)^{(j)}$, and if $\mathbf{C}_i$ belongs to one of $\mathbf{i}_1, \dots, \mathbf{i}_r$, so does $(\mathbf{C}_i)^{(j)}$. If $\mathbf{C}_i$ is a closed instance of a nonlogical axiom of T, so is $\mathbf{C}_i^{(j)}$, and if $\mathbf{C}_i$ belongs to a special constant $\mathbf{j}$ of rank at most n, then this special constant is not $\mathbf{i}$ and hence $\mathbf{C}_i^{(j)}$ belongs to $\mathbf{j}$. If $\mathbf{C}_i$ belongs to one of $\mathbf{i}_1, \dots,$ $\mathbf{i}_r$, then since the level of $\mathbf{i}$ is as great as the levels of $\mathbf{i}_1, \dots, \mathbf{i}_r$, $\mathbf{C}_i^{(j)}$ belongs to the same special constant. In view of this and the induction hypothesis, it will suffice to prove that

$$F[\mathbf{C}_1, \dots, \mathbf{C}_p, (\mathbf{C}_1)^{(1)}, \dots, (\mathbf{C}_1)^{(q)}, \dots, (\mathbf{C}_p)^{(1)}, \dots, (\mathbf{C}_p)^{(q)}] \tag{7}$$

is inconsistent.

We observe that $\exists\mathbf{x}\mathbf{B}$ does not occur in $\mathbf{C}_1, \dots, \mathbf{C}_p$. If $\mathbf{C}_i$ is an instance of a nonlogical axiom of T, then it is open by hypothesis. Since there are $n+1$ occurrences of $\exists$ in $\exists\mathbf{x}\mathbf{B}$, it clearly cannot occur in a formula belonging to a special constant of rank at most n. If $\exists\mathbf{x}\mathbf{B}$ occurs in a formula belonging to a special constant of rank exactly $n+1$, then this special constant must be $\mathbf{i}$, and hence it cannot occur in formulae belonging to $\mathbf{i}_1, \dots, \mathbf{i}_r$ either. Now if $\mathbf{A}'$ denotes the formula obtained from $\mathbf{A}$ by replacing each occurrence of $\exists\mathbf{x}\mathbf{B}$ by $\mathbf{B}[\mathbf{x}|\mathbf{i}]$, then by Proposition 2 of ch. I §3.3 and the tautology theorem, $F[\mathbf{A}'_1, \dots, \mathbf{A}'_k]$ is inconsistent. The formula $\mathbf{D}'_j$ is $\mathbf{B}[\mathbf{x}|\mathbf{a}_j] \to \mathbf{B}[\mathbf{x}|\mathbf{i}]$. If some $\mathbf{A}_i$ is the special axiom for $\mathbf{i}$, then $\mathbf{A}'_i$ is the tautology $\mathbf{B}[\mathbf{x}|\mathbf{i}] \to \mathbf{B}[\mathbf{x}|\mathbf{i}]$, so by the tautology theorem

$$F[\mathbf{C}_1, \dots, \mathbf{C}_p, \mathbf{B}[\mathbf{x}|\mathbf{a}_1] \to \mathbf{B}[\mathbf{x}|\mathbf{i}], \dots, \mathbf{B}[\mathbf{x}|\mathbf{a}_q] \to \mathbf{B}[\mathbf{x}|\mathbf{i}]] \tag{8}$$

is inconsistent. Hence $F[(\mathbf{C}_1)^{(j)}, \dots, (\mathbf{C}_p)^{(j)}, (\mathbf{B}[\mathbf{x}|\mathbf{a}_1] \to \mathbf{B}[\mathbf{x}|\mathbf{i}])^{(j)}, \dots, (\mathbf{B}[\mathbf{x}|\mathbf{a}_q] \to \mathbf{B}[\mathbf{x}|\mathbf{i}])^{(j)}]$ is inconsistent for $1 \le j \le q$ by the proposition of ch. I §3.3 and the tautology theorem. Noting that $(\mathbf{B}[\mathbf{x}|\mathbf{a}] \to \mathbf{B}[\mathbf{x}|\mathbf{i}])^{(j)}$ is $\mathbf{B}[\mathbf{x}|\mathbf{a}]^{(j)} \to \mathbf{B}[\mathbf{x}|\mathbf{a}_j]$,

$$F[(\mathbf{C}_1)^{(j)}, \dots, (\mathbf{C}_p)^{(j)}, \mathbf{B}[\mathbf{x}|\mathbf{a}_1]^{(j)} \to \mathbf{B}[\mathbf{x}|\mathbf{a}_j], \dots, \mathbf{B}[\mathbf{x}|\mathbf{a}_q]^{(j)} \to \mathbf{B}[\mathbf{x}|\mathbf{a}_j]] \tag{9}$$

is inconsistent for all j. Let V be a truth valuation on $L(T_c)$ assigning T to each $(\mathbf{C}_i)^{(j)}$. Then by the inconsistency of (9) and the tautology theorem, V must assign F to $\mathbf{B}[\mathbf{x}|\mathbf{a}_j]$ for all j. But then by the inconsistency of (8) and the tautology theorem, V must assign F to some $\mathbf{C}_i$. Thus there is no truth valuation V assigning T to all formulae in (7), and so by the tautology theorem (7) is inconsistent.

Next, suppose that T_c^* is inconsistent. By Proposition 2 of §1.1, there exists a tautology $\mathbf{A}$ which is a disjunction of negations of instances in $L(T_c)$ of equality rules and nonlogical axioms of T; let $\mathbf{i}_1, \dots, \mathbf{i}_n$ be the special constants occurring in $\mathbf{A}$, and let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be distinct variables not occurring in $\mathbf{A}$. Then by replacing each occurrence of $\mathbf{i}_i$ by $\mathbf{x}_i$ for all i, we obtain a formula $\mathbf{A}^*$ in T^*, which is again a tautology by Proposition 2 of ch. I §3.3. Clearly, $\mathbf{A}^*$ is a disjunction of negations of instances in $L(T)$ of equality rules and nonlogical axioms of T. By Proposition 2 of §1.1, T^* is inconsistent.

Finally, suppose that T^* is inconsistent. By the substitution rule, T is an extension of T^*. Hence $x = x$ and $x \ne x$ are theorems of T, which is therefore inconsistent by the tautology theorem. □

Corollary. Let T be an open first-order theory and $\mathbf{A}$ a closed existential formula of T. Then $\mathbf{A}$ is a theorem of T if and only if some disjunction of intances of the matrix of $\mathbf{A}$ is a tautological consequence of instances of equality rules for $L(T)$ and instances of nonlogical axioms of T.

Proof. Say $\mathbf{A}$ is $\exists\mathbf{x}_1 \dots \exists\mathbf{x}_n\mathbf{B}$ with $\mathbf{B}$ open. By proposition 1 of §1.1, $\mathbf{A}$ is a theorem of T if and only if $T[\neg\mathbf{A}]$ is inconsistent. By prenex operations and the closure theorem, this is the case if and only if $T[\neg\mathbf{B}]$ is inconsistent, and by the consistency theorem, this is the case if and only if $T[\neg\mathbf{B}]^*$ is inconsistent. By Proposition 2 of §1.1, this is in turn equivalent to the existence of a tautology which is a disjunction of negations of formulae of the following kind:

(i) instances of equality rules for $L(T)$;
(ii) instances of nonlogical axioms of T;
(iii) instances of $\neg\mathbf{B}$.

If $\neg\neg\mathbf{B}_1, \dots, \neg\neg\mathbf{B}_n$ are the negations of the formulae of type (iii) appearing in such a disjunction, then $\mathbf{B}_1 \vee \cdots \vee \mathbf{B}_n$ is a tautological consequences of formulae of types (i) and (ii). Conversely, if there are instances $\mathbf{B}_1, \dots, \mathbf{B}_n$ of the matrix of $\mathbf{A}$ whose disjunction is a tautological consequence of formulae $\mathbf{C}_1, \dots, \mathbf{C}_m$ of types (i) and (ii), then

$$\neg\neg\mathbf{B}_1 \vee \cdots \vee \neg\neg\mathbf{B}_n \vee \neg\mathbf{C}_1 \vee \cdots \vee \neg\mathbf{C}_m$$

is a tautology. □

4.3 Herbrand forms and Skolem forms. Let L be a first-order language and let $\mathbf{A}$ be a formula of L in prenex form. Then $\mathbf{A}$ can be written in the form

$$\exists\mathbf{x}_1 \dots \exists\mathbf{x}_{n_1} \forall\mathbf{y}_1 \dots \exists\mathbf{x}_{n_{k-1}+1} \dots \exists\mathbf{x}_{n_k} \forall\mathbf{y}_k \exists\mathbf{x}_{n_k+1} \dots \exists\mathbf{x}_{n_{k+1}} \mathbf{B} \tag{10}$$

where $\mathbf{B}$ is open and $0 \le n_1 \le \cdots \le n_{k+1}$. A *Herbrand form* of this prenex form is any formula of the form

$$\exists\mathbf{x}_1 \dots \exists\mathbf{x}_{n_{k+1}} \mathbf{B}[\mathbf{y}_1, \dots, \mathbf{y}_k | \mathbf{f}_1\mathbf{x}_1 \dots \mathbf{x}_{n_1}, \dots, \mathbf{f}_k\mathbf{x}_1 \dots \mathbf{x}_{n_k}] \tag{11}$$

in some extension of L, where the $\mathbf{f}_i$ are distinct and do not occur in $\mathbf{A}$. Thus, each $\mathbf{f}_i$ has index n_i. Dually, $\mathbf{A}$ can also be written in the form

$$\forall\mathbf{x}_1 \dots \forall\mathbf{x}_{n_1} \exists\mathbf{y}_1 \dots \forall\mathbf{x}_{n_{k-1}+1} \dots \forall\mathbf{x}_{n_k} \exists\mathbf{y}_k \forall\mathbf{x}_{n_k+1} \dots \forall\mathbf{x}_{n_{k+1}} \mathbf{B} \tag{12}$$

where $\mathbf{B}$ is open and $0 \le n_1 \le \cdots \le n_{k+1}$. A *Skolem form* of this prenex form is any formula of the form

$$\forall\mathbf{x}_1 \dots \forall\mathbf{x}_{n_{k+1}} \mathbf{B}[\mathbf{y}_1, \dots, \mathbf{y}_k | \mathbf{f}_1\mathbf{x}_1 \dots \mathbf{x}_{n_1}, \dots, \mathbf{f}_k\mathbf{x}_1 \dots \mathbf{x}_{n_k}] \tag{13}$$

in some extension of L, where the $\mathbf{f}_i$ are distinct and do not occur in $\mathbf{A}$. If $\mathbf{A}$ is an arbitrary formula of L, a Herbrand form (resp. a Skolem form) of $\mathbf{A}$ is defined to be a Herbrand form (resp. a Skolem form) of a prenex form of $\mathbf{A}$. Observe that Herbrand forms are existential formulae while Skolem forms are universal formulae.

THEOREM. Let T be an open first-order theory, $\mathbf{A}$ a closed formula of T and $\mathbf{A}^*$ a Herbrand form of $\mathbf{A}$. Let T' be obtained from T by the adjunction of the new function symbols in $\mathbf{A}^*$. Then $\vdash_T \mathbf{A}$ if and only if $\vdash_{T'} \mathbf{A}^*$.

Proof. By the theorem on prenex operations, we may suppose that $\mathbf{A}$ is in prenex form, say in the form (10), and that $\mathbf{A}^*$ has the form (11). Then $\vdash_{T'} \mathbf{A} \to \mathbf{A}^*$ by the tautology theorem and several applications of the substitution theorem and of the distribution rule. Now if $\vdash_T \mathbf{A}$, then $\vdash_{T'} \mathbf{A}$ and so $\vdash_{T'} \mathbf{A}^*$ by the detachement rule.

Before proving the converse we introduce some notations. For $1 \le i \le k$, let $\mathbf{A}_i$ be the subformula of (10) starting on the left from $\exists\mathbf{x}_{n_i+1}$, and let $\mathbf{B}_i$ be $\forall\mathbf{y}_i\mathbf{A}_i$. We also write $\mathbf{A}_0$ for $\mathbf{A}$ and $\mathbf{B}_{k+1}$ for $\mathbf{B}$. Let $\mathbf{a}_1, \dots, \mathbf{a}_{n_{k+1}}$ be closed terms of $L(T_c)$. For $1 \le i \le k$, we shall denote by $\mathbf{k}_i(\mathbf{a}_1, \dots, \mathbf{a}_{n_i})$ the special constant for the instantiation

$$\exists\mathbf{y}_i \neg\mathbf{A}_i[\mathbf{y}_1, \dots, \mathbf{y}_{i-1} | \mathbf{k}_1(\mathbf{a}_1, \dots, \mathbf{a}_{n_1}), \dots, \mathbf{k}_{i-1}(\mathbf{a}_1, \dots, \mathbf{a}_{n_{i-1}})][\mathbf{x}_1, \dots, \mathbf{x}_{n_i} | \mathbf{a}_1, \dots, \mathbf{a}_{n_i}].$$

We let $\mathbf{A}_i(\mathbf{a}_1, \dots, \mathbf{a}_{n_i})$ be the formula

$$\mathbf{A}_i[\mathbf{y}_1, \dots, \mathbf{y}_i | \mathbf{k}_1(\mathbf{a}_1, \dots, \mathbf{a}_{n_1}), \dots, \mathbf{k}_i(\mathbf{a}_1, \dots, \mathbf{a}_{n_i})][\mathbf{x}_1, \dots, \mathbf{x}_{n_i} | \mathbf{a}_1, \dots, \mathbf{a}_{n_i}]$$

(setting $n_0 = 0$), and similarly $\mathbf{B}_i(\mathbf{a}_1, \dots, \mathbf{a}_{n_i})$ is the formula

$$\mathbf{B}_i[\mathbf{y}_1, \dots, \mathbf{y}_{i-1} | \mathbf{k}_1(\mathbf{a}_1, \dots, \mathbf{a}_{n_1}), \dots, \mathbf{k}_{i-1}(\mathbf{a}_1, \dots, \mathbf{a}_{n_{i-1}})][\mathbf{x}_1, \dots, \mathbf{x}_{n_i} | \mathbf{a}_1, \dots, \mathbf{a}_{n_i}].$$

We claim that

$$\vdash_{T_c} \mathbf{B}(\mathbf{a}_1, \dots, \mathbf{a}_{n_{k+1}}) \to \mathbf{A}. \tag{14}$$

By the tautology theorem, it will suffice to prove

$$\vdash_{T_c} \mathbf{B}_i(\mathbf{a}_1, \dots, \mathbf{a}_{n_i}) \to \mathbf{A}_{i-1}(\mathbf{a}_1, \dots, \mathbf{a}_{n_{i-1}}) \text{ and}$$
$$\vdash_{T_c} \mathbf{A}_i(\mathbf{a}_1, \dots, \mathbf{a}_{n_i}) \to \mathbf{B}_i(\mathbf{a}_1, \dots, \mathbf{a}_{n_i}),$$

the former for $1 \le i \le k+1$ and the latter for $1 \le i \le k$. The former follows from $n_i - n_{i-1}$ substitution axioms and the tautology theorem, while the latter is a tautological consequence of the special axiom for $\mathbf{k}_i(\mathbf{a}_1, \dots, \mathbf{a}_{n_i})$.

Let $\mathbf{a}_1, \dots, \mathbf{a}_{n_k}, \mathbf{a}'_1, \dots, \mathbf{a}'_{n_k}$ be closed terms of $L(T_c)$. We claim that for all i

$$\vdash_{T'_c} \mathbf{a}_1 = \mathbf{a}'_1 \to \cdots \to \mathbf{a}_{n_i} = \mathbf{a}'_{n_i} \to \mathbf{k}_i(\mathbf{a}_1, \dots, \mathbf{a}_{n_i}) = \mathbf{k}_i(\mathbf{a}'_1, \dots, \mathbf{a}'_{n_i}). \tag{15}$$

We prove this fact by induction on i. Suppose it is proved for all $j < i$. Let $\exists \mathbf{y}_i \neg \mathbf{A}'_i(\mathbf{a}_1, \dots, \mathbf{a}_{n_i})$ be the formula for which $\mathbf{k}_i(\mathbf{a}_1, \dots, \mathbf{a}_{n_i})$ is the special constant. Then by the equality theorem, the tautology theorem, and the $\forall$-introduction rule,

$$\vdash_{T'_c} \mathbf{a}_1 = \mathbf{a}'_1 \to \cdots \to \mathbf{a}_{n_i} = \mathbf{a}'_{n_i} \to \forall \mathbf{y}_i (\neg \mathbf{A}'_i(\mathbf{a}_1, \dots, \mathbf{a}_{n_i}) \leftrightarrow \neg \mathbf{A}'_i(\mathbf{a}'_1, \dots, \mathbf{a}'_{n_i})),$$

and we obtain the desired result from the tautology theorem and the special equality axiom

$$\forall \mathbf{y}_i (\neg \mathbf{A}'_i(\mathbf{a}_1, \dots, \mathbf{a}_{n_i}) \leftrightarrow \neg \mathbf{A}'_i(\mathbf{a}'_1, \dots, \mathbf{a}'_{n_i})) \to \mathbf{k}_i(\mathbf{a}_1, \dots, \mathbf{a}_{n_i}) = \mathbf{k}_i(\mathbf{a}'_1, \dots, \mathbf{a}'_{n_i}).$$

Suppose that $\vdash_{T'} \mathbf{A}^*$. By the corollary to the consistency theorem, we can find in $L(T')$ a disjunction of the form

$$\begin{aligned} &\mathbf{B}[\mathbf{y}_1, \dots, \mathbf{y}_k | \mathbf{f}_1 \mathbf{x}_1 \dots \mathbf{x}_{n_1}, \dots, \mathbf{f}_k \mathbf{x}_1 \dots \mathbf{x}_{n_k}][\mathbf{x}_1, \dots, \mathbf{x}_{n_{k+1}} | \mathbf{a}_{1,1}, \dots, \mathbf{a}_{1,n_{k+1}}] \\ &\quad \vee \cdots \vee \mathbf{B}[\mathbf{y}_1, \dots, \mathbf{y}_k | \mathbf{f}_1 \mathbf{x}_1 \dots \mathbf{x}_{n_1}, \dots, \mathbf{f}_k \mathbf{x}_1 \dots \mathbf{x}_{n_k}][\mathbf{x}_1, \dots, \mathbf{x}_{n_{k+1}} | \mathbf{a}_{n,1}, \dots, \mathbf{a}_{n,n_{k+1}}] \end{aligned} \tag{16}$$

which is a tautological consequence of instances in $L(T')$ of equality rules for $L(T')$ and nonlogical axioms of T. Let $\mathbf{C}_1, \dots, \mathbf{C}_r$ be such instances. Then each $\mathbf{C}_l$ is an open formula because the nonlogical axioms of T are open. We now modify the formula (16) and the formulae $\mathbf{C}_1, \dots, \mathbf{C}_r$ simultaneously as follows. Choose a variable occurring in some $\mathbf{a}_{p,q}$ or some $\mathbf{C}_l$ and replace all its occurrences by some special constant. Continue to do so until there are no more variables left. Then, select an occurrence of a term of the form $\mathbf{f}_i \mathbf{a}_1 \dots \mathbf{a}_{n_i}$ in (16) or some $\mathbf{C}_l$ such that $\mathbf{f}_1, \dots, \mathbf{f}_k$ do not occur in $\mathbf{a}_1, \dots, \mathbf{a}_{n_i}$, and replace all such occurrences by $\mathbf{k}_i(\mathbf{a}_1, \dots, \mathbf{a}_{n_i})$. Continue to do so until there are no more occurrences of $\mathbf{f}_1, \dots, \mathbf{f}_k$. These modifications tranform (16) into a formula

$$\mathbf{B}(\mathbf{a}'_{1,1}, \dots, \mathbf{a}'_{1,n_{k+1}}) \vee \cdots \vee \mathbf{B}(\mathbf{a}'_{n,1}, \dots, \mathbf{a}'_{n,n_{k+1}}) \tag{17}$$

of $L(T_c)$, where the $\mathbf{a}'_{p,q}$ are closed terms, and they transform $\mathbf{C}_1, \dots, \mathbf{C}_r$ into formulae $\mathbf{C}'_1, \dots, \mathbf{C}'_r$ of $L(T_c)$. Since these transformations only affect atomic formulae, it follows from Proposition 2 of ch. I §3.3 that (17) is a tautological consequence of $\mathbf{C}'_1, \dots, \mathbf{C}'_r$. But the $\mathbf{C}'_l$ are again instances in $L(T_c)$ of equality rules for $L(T)$ and nonlogical axioms of T, unless $\mathbf{C}_l$ is an equality axiom of the form $\mathbf{x}_1 = \mathbf{y}_1 \to \cdots \mathbf{x}_{n_i} = \mathbf{y}_{n_i} \to \mathbf{f}_i \mathbf{x}_1 \dots \mathbf{x}_{n_i} = \mathbf{f}_i \mathbf{y}_1 \dots \mathbf{y}_{n_i}$. But then $\mathbf{C}'_l$ is a theorem of T'_c by (15). Thus $\mathbf{C}'_1, \dots, \mathbf{C}'_r$ are theorems of T'_c, so by the tautology theorem, (17) is a theorem of T'_c. By (14) and the tautology theorem, $\mathbf{A}$ is a theorem of T'_c and hence, by the theorem of §4.1, a theorem of T. ☐

This theorem and the following corollary are in fact true for an arbitrary first-order theory T, as we shall prove in §4.6 below.

COROLLARY 1. Let T be an open first-order theory, $\mathbf{A}$ a closed formula of T and $\mathbf{A}^\circ$ a Skolem form of $\mathbf{A}$. Let T' be obtained from T by the adjunction of the new function symbols in $\mathbf{A}^\circ$. Then $\vdash_T \neg \mathbf{A}$ if and only if $\vdash_{T'} \neg \mathbf{A}^\circ$.

Proof. We can obtain a Herbrand form of $\neg \mathbf{A}$ from $\neg \mathbf{A}^\circ$ using prenex operations. Thus, the corollary follows from the theorem, the theorem on prenex operations, and the tautology theorem. ☐

COROLLARY 2. Let T be an open first-order theory, $\mathbf{A}$ a closed formula of T and $\mathbf{A}^*$ a Herbrand form of $\mathbf{A}$. Let T' be obtained from T by the adjunction of the new function symbols in $\mathbf{A}^*$. Then $\vdash_T \mathbf{A}$ if and only if some disjunction of instances in $L(T')$ of the matrix of $\mathbf{A}^*$ is a tautological consequence of instances in $L(T')$ of equality rules for $L(T')$ and instances in $L(T')$ of nonlogical axioms of T'.

Proof. This follows at once from the theorem and the corollary to the consistency theorem. ☐

4.4 Functional extensions.

Theorem on Functional Extensions. Let T be a first-order theory, $\mathbf{A}$ a formula of T, and $\mathbf{x}_1, \dots, \mathbf{x}_n$, and $\mathbf{y}$ distinct variables including the variables free in $\mathbf{A}$. If $\vdash_T \exists\mathbf{y}\mathbf{A}$, then the first-order theory obtained from T by the adjunction of a new n-ary function symbol $\mathbf{f}$ and the axiom $\mathbf{A}[\mathbf{y}|\mathbf{f}\mathbf{x}_1, \dots, \mathbf{x}_n]$ is a conservative extension of T.

Proof. Let T' denote the extension of T to be proved conservative. Let T_0 be the first-order theory with language $L(T)$ and without nonlogical axioms, and let T_1 be obtained from T_0 by the adjunction of the n-ary function symbol $\mathbf{f}$. By the closure theorem, it suffices to prove that any closed formula of T which is a theorem of T' is a theorem of T. Let $\mathbf{B}$ be such a formula. By the reduction theorem, there are formulae $\mathbf{C}_1, \dots, \mathbf{C}_k$ among the closures of the nonlogical axioms of T such that $\forall\mathbf{x}_1 \dots \forall\mathbf{x}_n\mathbf{A}[\mathbf{y}|\mathbf{f}\mathbf{x}_1 \dots \mathbf{x}_n] \to \mathbf{C}_1 \to \dots \to \mathbf{C}_k \to \mathbf{B}$ is a theorem of T_1. Let $\mathbf{C}$ be the latter formula and let $\mathbf{D}$ be the formula $\forall\mathbf{x}_1 \dots \forall\mathbf{x}_n \exists\mathbf{y}\mathbf{A} \to \mathbf{C}_1 \to \dots \to \mathbf{C}_k \to \mathbf{B}$. Let $\mathbf{C}^*$ be a Herbrand form of $\mathbf{C}$, and let T_1' be obtained from T_1 by the adjunction of the new function symbols in $\mathbf{C}^*$. Then it is clear that $\mathbf{C}^*$ is also a Herbrand form of $\mathbf{D}$. By two applications of the theorem of §4.3, we obtain first $\vdash_{T_1'}\mathbf{C}^*$ and then $\vdash_{T_0}\mathbf{D}$. Then by the hypothesis, the closure theorem, and the detachement rule, $\vdash_T\mathbf{B}$. □

Remark. The theorem on functional extensions can be used to prove the second part of theorem on functional definitions almost instantly, by deriving the equivalence between the defining axiom of $\mathbf{f}$ and $\mathbf{D}[\mathbf{y}|\mathbf{f}\mathbf{x}_1, \dots, \mathbf{x}_n]$ using a uniqueness condition for $\mathbf{y}$ in $\mathbf{D}$. However, the direct proof of §2.2 produces a much shorter derivation of a translation of $\mathbf{A}$ from a given derivation of $\mathbf{A}$, and thus is practically preferrable.

We shall use this theorem to prove that any first-order theory has a conservative Skolem extension, using a construction similar to that of T_c. Let L be a first-order language. We define the *special function symbols of level n*, for $n \geq 1$, by induction on n. Let $\Gamma_0(L)$ denote the collection of instantiations of L. Suppose that, for some $n \geq 0$, the collection $\Gamma_n(L)$ has been described. For every formula $\exists\mathbf{x}\mathbf{A}$ in $\Gamma_n(L)$ and every choice of distinct variables $\mathbf{x}, \mathbf{x}_1, \dots, \mathbf{x}_n$ including the variables free in $\mathbf{A}$, we choose a new n-ary function symbol called the *special function symbol for* $\exists\mathbf{x}\mathbf{A}$ *and* $\mathbf{x}_1, \dots, \mathbf{x}_n$; the special function symbols of level $n+1$ are the special function symbols for all the formulae of $\Gamma_n(L)$ and all suitable families of variables. We then let $\Gamma_{n+1}(L)$ consists of the instantiations not in $\Gamma_n(L)$ of the language obtained from L by adding the special function symbols of level at most $n+1$. The first-order language obtained from L by adding all the special function symbols is denoted by L_f. If $\mathbf{f}$ is the special constant for $\exists\mathbf{x}\mathbf{A}$ and $\mathbf{x}_1, \dots, \mathbf{x}_n$, the *special axiom for* $\mathbf{f}$ is the formula $\exists\mathbf{x}\mathbf{A} \to \mathbf{A}'[\mathbf{x}|\mathbf{f}\mathbf{x}_1 \dots \mathbf{x}_n]$, where $\mathbf{A}'$ is a variant of $\mathbf{A}$, fixed once and for all, in which $\mathbf{f}\mathbf{x}_1 \dots \mathbf{x}_n$ is substitutible for $\mathbf{x}$. If now T is a first-order theory with language L, we let T_f be the first-order theory whose language is L_f and whose nonlogical axioms are those of T and the special axioms for the special function symbols of L_f. It is then obvious that T_f is a Skolem extension of T.

We form T_f' from T_f by adding as further nonlogical axioms all the formulae of the form

$$\forall\mathbf{x}(\mathbf{A} \leftrightarrow \mathbf{B}) \to \mathbf{f}\mathbf{x}_1 \dots \mathbf{x}_n = \mathbf{g}\mathbf{y}_1 \dots \mathbf{y}_k,$$

where $\mathbf{f}$ is the special function symbol for $\exists\mathbf{x}\mathbf{A}$ and $\mathbf{x}_1, \dots, \mathbf{x}_n$ and $\mathbf{g}$ is the special function symbol for $\exists\mathbf{x}\mathbf{B}$ and $\mathbf{y}_1, \dots, \mathbf{y}_k$.

Corollary. T_f and T_f' are conservative extensions of T.

Proof. Follow the proof of the theorem of §4.1, using the theorem on functional extensions instead of the lemma used there. The only difference is that in proving (3) using the deduction theorem we must first replace the free variables by new constants, and in the derivation we use the corresponding instances of the axioms of T_k rather than the axioms themselves. □

4.5 The ε-theorems. The goal of this paragraph is to indicate how two famous theorems of Hilbert can be recovered from the results of this section. As we shall see, they are merely reformulations of our own theorems. Given a first-order theory T, we define the *ε-terms* and *ε-formulae* of T by simultaneous induction as follows:

(i) variables are ε-terms;

(ii) if $\mathbf{a}_1, \dots, \mathbf{a}_n$ are ε-terms and $\mathbf{f}$ is an n-ary function symbol, then $\mathbf{f}\mathbf{a}_1 \dots \mathbf{a}_n$ is an ε-term;

(iii) if $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are ε-terms and $\mathbf{p}$ is an n-ary predicate symbol, then $\mathbf{p}\mathbf{a}_1 \ldots \mathbf{a}_n$ is an ε-formula;
(iv) if $\mathbf{A}$ and $\mathbf{B}$ are ε-formulae, then $\vee\mathbf{AB}$ is an ε-formula;
(v) if $\mathbf{A}$ is an ε-formula, then $\neg\mathbf{A}$ is an ε-formula;
(vi) if $\mathbf{A}$ is an ε-formula and $\mathbf{x}$ is a variable, then $\exists\mathbf{xA}$ is an ε-formula;
(vii) if $\mathbf{A}$ is an ε-formula and $\mathbf{x}$ is a variable, then $\varepsilon\mathbf{xA}$ is an ε-term.

If we specify that any occurrence of $\mathbf{x}$ within an occurrence of a term of the form $\varepsilon\mathbf{xA}$ is to be bound, we can define the notions of free and bound occurrences for ε-terms and ε-formulae in the obvious way, and we can then define instances and variants.

We let T_ε be the formal system whose alphabet is that of T together with the new symbol ε, whose formulae are the ε-formulae of T, and whose rules of inference are: the nonlogical axioms of T; if $\varepsilon\mathbf{xA}$ is substitutible for $\mathbf{x}$ in $\mathbf{A}$, infer $\exists\mathbf{xA} \to \mathbf{A}[\mathbf{x}|\varepsilon\mathbf{xA}]$; infer $\forall\mathbf{x}(\mathbf{A} \leftrightarrow \mathbf{B}) \to \varepsilon\mathbf{xA} = \varepsilon\mathbf{xB}$; and the rules (i)–(x) of ch. I §2.9 in which $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are now any ε-formulae. Clearly T_ε is an extension of T. Note that there is some redundancy in the language of T_ε, as we have $\vdash_{T_\varepsilon} \exists\mathbf{xA} \leftrightarrow \mathbf{A}'[\mathbf{x}|\varepsilon\mathbf{xA}]$ for a suitable variant $\mathbf{A}'$ of $\mathbf{A}$. We could thus have discarded the symbol ∃ entirely, modifying the rules of inference appropriately.

To any ε-term or ε-formula $\mathbf{u}$ of T, we associate a term or formula $\mathbf{u}^*$ of T'_{f} using recursion on (i)–(vii). If $\mathbf{u}$ is a variable, $\mathbf{u}^*$ is $\mathbf{u}$. If $\mathbf{u}$ is $\varepsilon\mathbf{xA}$, $\mathbf{u}^*$ is $\mathbf{f}\mathbf{x}_1 \ldots \mathbf{x}_n$ where $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are the variables free in $\mathbf{A}^*$ except $\mathbf{x}$ in alphabetical order and $\mathbf{f}$ is the special function symbol for $\exists\mathbf{xA}^*$ and $\mathbf{x}_1, \ldots, \mathbf{x}_n$. If $\mathbf{u}$ is obtained by (ii)–(vi), $\mathbf{u}^*$ is defined in the evident way. A straightforward induction on theorems in T_ε shows that for any formula $\mathbf{A}$ of T_ε, if $\vdash_{T_\varepsilon}\mathbf{A}$, then $\vdash_{T'_{\mathrm{f}}}\mathbf{A}^*$ (in fact, the converse is also easily proved). Since $\mathbf{A}^*$ is $\mathbf{A}$ if $\mathbf{A}$ is a formula of T, the corollary of §4.4 implies that T_ε is a conservative extension of T, a result known as the *second ε-theorem*.

The *first ε-theorem* is the following statement: if T is an open first-order theory and $\mathbf{A}$ is an open formula of T such that $\vdash_{T_\varepsilon}\mathbf{A}$, then $\mathbf{A}$ is a tautological consequence of instances of equality rules for $L(T)$ and nonlogical axioms of T. To see that it is true, apply first the second ε-theorem to deduce that $\mathbf{A}$ is a theorem of T. Then let $\mathbf{A}'$ be obtained from $\mathbf{A}$ by replacing the free variables by as many new constants, thereby forming an extension T' of T. By the substitution rule, $\vdash_{T'}\mathbf{A}'$. By the corollary to the consistency theorem, $\mathbf{A}'$ is a tautological consequence of instances in $L(T')$ of equality rules for $L(T')$ and nonlogical axioms of T. By Proposition 2 of ch. I §3.3 and the fact that replacing all occurrences of the constant $\mathbf{e}$ by a variable in the equality axiom $\mathbf{e} = \mathbf{e}$ yields an identity axiom, we deduce that $\mathbf{A}$ itself is a tautological consequence of instances of the contended kind.

4.6 The fundamental theorem of Herbrand–Skolem theory. Let T be a first-order theory. We denote by T° any first-order theory whose language is an extension of $L(T)$ and whose nonlogical axioms consist of the matrices of Skolem forms of all nonlogical axioms of T.

LEMMA. T° is a conservative extension of T.

Proof. Let us first prove that T° is an extension of T. By the substitution axioms, the distribution rule, the tautology theorem, and prenex operations, $\mathbf{A}^\circ \to \mathbf{A}$ is a theorem of T° for any formula $\mathbf{A}$ of T° with Skolem form $\mathbf{A}^\circ$ in T°. In particular, by the detachement rule and the closure theorem, the nonlogical axioms of T are theorems of T°.

To prove that T° is a conservative extension of T, we may suppose, by the reduction theorem and transitivity of conservative extensions, that T has only one nonlogical axiom $\mathbf{A}$. Let $\mathbf{A}^\circ$ be the nonlogical axiom of T°. Let $\mathbf{A}$ be written in the form (12) so that $\mathbf{A}^\circ$ is the matrix of (13). For $0 \leq i \leq k$, let $\mathbf{A}_i$ be the formula

$$\exists\mathbf{y}_{i+1}\forall\mathbf{x}_{n_{i+1}+1}\ldots\forall\mathbf{x}_{n_{i+2}}\ldots\exists\mathbf{y}_k\forall\mathbf{x}_{n_k+1}\ldots\forall\mathbf{x}_{n_{k+1}}\mathbf{B}[\mathbf{y}_1,\ldots,\mathbf{y}_i|\mathbf{f}_1\mathbf{x}_1\ldots\mathbf{x}_{n_1},\ldots,\mathbf{f}_i\mathbf{x}_1\ldots\mathbf{x}_{n_i}].$$

Then $\mathbf{A}_k$ is $\mathbf{A}^\circ$. Let T_i be the first-order theory obtained from T by the adjunction of the function symbols $\mathbf{f}_1, \ldots, \mathbf{f}_i$ and of the nonlogical axioms $\mathbf{A}_0, \ldots, \mathbf{A}_i$. By the functional extension theorem and the closure theorem, T_i is a conservative extension of T_{i-1}, for $1 \leq i \leq k$, and hence T_k is a conservative extension of T. By the proposition of ch. I §4.2, any formula of T which is a theorem of T° is a theorem of T_k, so T° is also a conservative extension of T. □

Let $\mathbf{A}$ be a closed formula of T, and let $\mathbf{A}^*$ be a Herbrand form of $\mathbf{A}$ in some extension L' of $L(T)$, chosen in such a way that there exists a first-order language L'° which is a common extension of L' and $L(T^\circ)$. Let T' (resp. T'°) be the first-order theory with language L' (resp. L'°) whose nonlogical axioms are those of T (resp. those of T°).

HERBRAND–SKOLEM THEOREM. With the notations of this paragraph, the following assertions are equivalent:

(i) $\mathbf{A}$ is a theorem of T;

(ii) $\mathbf{A}$ is a theorem of T°;

(iii) $\mathbf{A}^*$ is a theorem of T';

(iv) $\mathbf{A}^*$ is a theorem of T'°.

Proof. By the lemma, (i) and (ii) are equivalent and (iii) and (iv) are equivalent. But T° is an open first-order theory by definition, so by the theorem of §4.3 and the proposition of ch. I §4.2, (ii) and (iv) are equivalent. Hence, all four conditions are equivalent. □

This theorem reduces the problem of deriving a formula in a first-order theory to the problem of deriving an existential formula in an open first-order theory. Combining the Herbrand–Skolem theorem with the corollary to the consistency theorem, we also obtain a criterion for theoremhood in an arbitrary first-order theory T in terms of tautologies. This reduction of first-order logic to propositional logic was achieved by J. Herbrand and is generically known as *Herbrand's theorem*.

§5 Craig's interpolation lemma

5.1 Compatibility. We say that two first-order languages are *compatible* if they have a common first-order extension. If L_1 and L_2 are compatible, we can form the first-order language $L_1 \cup L_2$ (resp. $L_1 \cap L_2$; $L_1 - L_2$) as follows: a nonlogical symbol is an n-ary function symbol of $L_1 \cup L_2$ if and only if it is an n-ary function symbol of L_1 or of L_2 (resp. of L_1 and of L_2; of L_1 but not of L_2), and similarly for predicate symbols. If T_1 and T_2 are first-order theories such that $L(T_1)$ and $L(T_2)$ are compatible, we say that T_1 and T_2 are compatible and we define $T_1 \cup T_2$ to be the first-order theory with language $L(T_1) \cup L(T_2)$ and with nonlogical axioms those of T_1 together with those of T_2.

5.2 Eliminating function symbols. Two first-order theories T and T' are called *weakly equivalent* if some extension by definitions of T is equivalent to some extension by definitions of T'. In this paragraph we shall prove that any first-order theory is weakly equivalent to a first-order theory without function symbols.

Let L be a first-order language. We form a new first-order language $L^\$$ as follows: each n-ary predicate symbol of L is an n-ary predicate symbol of $L^\$$, and for each n-ary function symbol $\mathbf{f}$ of L, $L^\$$ has an $(n+1)$-ary predicate symbol $\mathbf{p_f}$. It is understood that $\mathbf{p_f}$ is distinct from all the symbols of L and is distinct from $\mathbf{p_g}$ if $\mathbf{f}$ is distinct from $\mathbf{g}$. Thus, $L^\$$ has no function symbols, and L and $L^\$$ are compatible. Let T be a first-order theory with language $L^\$$, and consider the first-order theory obtained from T by the adjunction of the n-ary function symbol $\mathbf{f}$ and the nonlogical axiom $y = \mathbf{f}x_1 \ldots x_n \leftrightarrow \mathbf{p_f} y x_1 \ldots x_n$, for every n-ary function symbol $\mathbf{f}$ of L. For $\mathbf{A}$ a formula of L, we shall denote by $\mathbf{A}^\$$ a translation of $\mathbf{A}$ into T, as defined in §2.2. Thus, $\mathbf{A}^\$$ is a formula of $L^\$$.

Let now T be a first-order theory with language L. We define $T^\$$ to be the first-order theory with language $L^\$$ whose nonlogical axioms are the formulae $\mathbf{A}^\$$ for each nonlogical axiom $\mathbf{A}$ of T and the formulae $\exists y \mathbf{p_f} y x_1 \ldots x_n$ and $\mathbf{p_f} y x_1 \ldots x_n \to \mathbf{p_f} y' x_1 \ldots x_n \to y = y'$ for each n-ary function symbol $\mathbf{f}$ of T.

THEOREM. With the notations of this paragraph, T and $T^\$$ are weakly equivalent.

Proof. Let U be obtained from T by the adjunction of the $(n+1)$-ary predicate symbols $\mathbf{p_f}$ and the axioms $y = \mathbf{f}x_1 \ldots x_n \leftrightarrow \mathbf{p_f} y x_1 \ldots x_n$ for each n-ary function symbol $\mathbf{f}$ of L. Let $U^\$$ be obtained from $T^\$$ by the adjunction of the n-ary function symbols $\mathbf{f}$ and the axioms $y = \mathbf{f}x_1 \ldots x_n \leftrightarrow \mathbf{p_f} y x_1 \ldots x_n$ for each n-ary function symbol $\mathbf{f}$ of L. Clearly U is equivalent to an extension by definitions of T and $U^\$$ is an extension by definitions of $T^\$$, so it will suffice to prove that U and $U^\$$ are equivalent. It is clear that U and $U^\$$ have the same language, namely $L \cup L^\$$.

Let $\mathbf{A}$ be a nonlogical axiom of U. If $\mathbf{A}$ is a nonlogical axiom of T, then $\mathbf{A}^\$$ is a nonlogical axiom of $T^\$$, and $\vdash_{U^\$} \mathbf{A}$ if and only if $\vdash_{T^\$} \mathbf{A}^\$$ by the theorem of §2.3. Thus $\vdash_{U^\$} \mathbf{A}$ in this case. If $\mathbf{A}$ is of the form $\mathbf{p_f} y x_1 \ldots x_n \leftrightarrow y = \mathbf{f}x_1 \ldots x_n$, then $\mathbf{A}$ is also a nonlogical axiom of $U^\$$. Thus, $U^\$$ is an extension of U.

Conversely, let $\mathbf{A}$ be a nonlogical axiom of $U^\$$. If $\mathbf{A}$ is $\mathbf{B}^\$$ for some nonlogical axiom $\mathbf{B}$ of T, the first part of the theorem on functional definitions shows that $\vdash_U \mathbf{B} \leftrightarrow \mathbf{A}$, and so $\vdash_U \mathbf{A}$. Suppose that $\mathbf{A}$ is $\exists y \mathbf{p_f} y x_1 \ldots x_n$. To prove $\vdash_U \mathbf{A}$, it will suffice, by the equivalence theorem, to prove $\vdash_U \exists y (y = \mathbf{f}x_1 \ldots x_n)$.

But this follows from an instance of an identity axiom, the substitution axioms, and the detachment rule. Similarly, if $\mathbf{A}$ is $\mathbf{p_f} y x_1 \dots x_n \to \mathbf{p_f} y' x_1 \dots x_n \to y = y'$, to prove $\vdash_U \mathbf{A}$, it suffices to prove

$$\vdash_U y = \mathbf{f} x_1 \dots x_n \to y' = \mathbf{f} x_1 \dots x_n \to y = y'.$$

This is true by the equality theorem, an instance of an identity axiom, and the tautology theorem. Thus, U is also an extension of $U^\$$, so that U and $U^\$$ are equivalent. □

Consider, more generally, two weakly equivalent first-order theories T and $T^\$$. For any formula $\mathbf{A}$ of T, let $\mathbf{A}^\$$ be a translation of $\mathbf{A}$ into $T^\$$, and for any formula $\mathbf{B}$ of $T^\$$, let $\mathbf{B}_\S$ be a translation of $\mathbf{B}$ into T.

Proposition. Let T and $T^\$$ be weakly equivalent first-order theories. Let $\mathbf{A}$ be a formula of T and $\mathbf{B}$ a formula of $T^\$$. Then

(i) $\vdash_T \mathbf{A}$ if and only if $\vdash_{T^\$} \mathbf{A}^\$$;
(ii) $\vdash_{T^\$} \mathbf{B}$ if and only if $\vdash_T \mathbf{B}_\S$;
(iii) $\vdash_T \mathbf{A} \leftrightarrow (\mathbf{A}^\$)_\S$;
(iv) $\vdash_{T^\$} \mathbf{B} \leftrightarrow (\mathbf{B}_\S)^\$$.

Proof. Note that it suffices to prove (i) and (iv). Let U and $U^\$$ be equivalent extensions by definitions of T and $T^\$$, respectively. By the conservativity of extensions by definitions, $\vdash_T \mathbf{A}$ if and only if $\vdash_U \mathbf{A}$, and by (ii) of the theorem of §2.3, $\vdash_{U^\$} \mathbf{A}$ if and only if $\vdash_{T^\$} \mathbf{A}^\$$. This proves (i). By the first part of the theorem on functional definitions, we have $\vdash_U \mathbf{B} \leftrightarrow \mathbf{B}_\S$ and $\vdash_{U^\$} \mathbf{B}_\S \leftrightarrow (\mathbf{B}_\S)^\$$, so $\vdash_{U^\$} \mathbf{B} \leftrightarrow (\mathbf{B}_\S)^\$$, whence (iv) by conservativity. □

5.3 Craig's interpolation lemma. In this paragraph we give a finitary proof of a famous result of first-order logic known as Craig's interpolation lemma: if T and U are compatible first-order theories and $\mathbf{A} \to \mathbf{B}$ is a theorem of $T \cup U$, where $\mathbf{A}$ is a formula of T and $\mathbf{B}$ of U, then we shall show how to find a formula $\mathbf{C}$ of $L(T) \cap L(U)$ such that $\mathbf{A} \to \mathbf{C}$ is a theorem of T and $\mathbf{C} \to \mathbf{B}$ is a theorem of U. Our strategy will be as follows. In a first step, we prove the result with "theorem" replaced by "tautology". In a second step we use Herbrand's theorem to generalize the first step to first-order theories with no function symbols. Finally, we prove the result for arbitrary first-order theories using the procedure described in §5.2.

Lemma 1. Let L and M be compatible first-order languages. Let $\mathbf{A}$ be a formula of L and $\mathbf{B}$ a formula of M. If $\mathbf{A} \to \mathbf{B}$ is a tautology, then there exists a formula $\mathbf{C}$ of $L \cap M$ such that $\mathbf{A} \to \mathbf{C}$ and $\mathbf{C} \to \mathbf{B}$ are tautologies. Moreover, $\mathbf{C}$ can be chosen so that the variables free in $\mathbf{C}$ are free in $\mathbf{A}$.

Proof. We prove the lemma by induction on the number of elementary formulae $\mathbf{D}$ such that $\mathbf{D}$ has an occurrence in $\mathbf{A}$ not happening within an occurrence of another elementary formula (we shall say that this is a *maximal* occurrence of $\mathbf{D}$) and such that nonlogical symbols of $L - M$ occur in $\mathbf{D}$. If there are none, $\mathbf{A}$ is already a formula of $L \cap M$ and we take $\mathbf{C}$ to be $\mathbf{A}$. Otherwise, choose an elementary formula $\mathbf{D}$ having a maximal occurrence in $\mathbf{A}$ and in which nonlogical symbols of $L - M$ occur, and let $\mathbf{A}_+$ (resp. $\mathbf{A}_-$) be obtained from $\mathbf{A}$ by replacing every maximal occurrence of $\mathbf{D}$ by $\exists x(x = x)$ (resp. $\neg\exists x(x = x)$). We claim that $\mathbf{A} \to \mathbf{A}_+ \vee \mathbf{A}_-$ and $\mathbf{A}_+ \vee \mathbf{A}_- \to \mathbf{B}$ are tautologies. To prove this, let V be a truth valuation on $L \cup M$. By a straighforward induction on the length of $\mathbf{A}$, we see that if $V(\mathbf{D})$ is $V(\exists x(x = x))$, then $V(\mathbf{A}) = V(\mathbf{A}_+)$, and otherwise $V(\mathbf{A}) = V(\mathbf{A}_-)$. It follows that $V(\mathbf{A} \to \mathbf{A}_+ \vee \mathbf{A}_-)$ is $\mathbf{T}$ in both cases. For the second formula, we have by Proposition 2 of ch. I §3.3 that both $\mathbf{A}_+ \to \mathbf{B}$ and $\mathbf{A}_- \to \mathbf{B}$ are tautologies, so $\mathbf{A}_+ \vee \mathbf{A}_- \to \mathbf{B}$ is a tautology. Note that the variables free in $\mathbf{A}_+ \vee \mathbf{A}_-$ are also free in $\mathbf{A}$. By induction hypothesis, there exists a formula $\mathbf{C}$ of $L \cap M$ whose free variables are free in $\mathbf{A}$ and such that $\mathbf{A}_+ \vee \mathbf{A}_- \to \mathbf{C}$ and $\mathbf{C} \to \mathbf{B}$ are tautologies. But then $\mathbf{A} \to \mathbf{C}$ and $\mathbf{C} \to \mathbf{B}$ are tautologies. □

Lemma 2. Let T and U be compatible first-order theories with no function symbols. Let $\mathbf{A}$ be a closed formula of T and $\mathbf{B}$ a closed formula of U such that $\vdash_{T \cup U} \mathbf{A} \to \mathbf{B}$. Then there exists a closed formula $\mathbf{C}$ of $L(T) \cap L(U)$ such that $\vdash_T \mathbf{A} \to \mathbf{C}$ and $\vdash_U \mathbf{C} \to \mathbf{B}$.

Proof. Choose once and for all a Skolem form $\forall \mathbf{x}_1 \dots \forall \mathbf{x}_n \mathbf{A}^\circ$ of $\mathbf{A}$, a Herbrand form $\mathbf{B}^*$ of $\mathbf{B}$, and Skolem forms of all the nonlogical axioms of $T \cup U$, and let L° be the first-order language obtained from $L(T) \cap L(U)$ by the adjunction of all the new function symbols introduced, as well as a new constant $\mathbf{e}$. Let T° be the first-order theory with language $L^\circ \cup L(T)$ whose nonlogical axioms are the matrices of the prescribed

Skolem forms of the nonlogical axioms of T, and define U° in the same way. Let also T' (resp. U') be the first-order theory obtained from T (resp. from U) by the adjunction of the function symbols of L°. By the detachment rule, $\mathbf{B}$ is a theorem of $T[\mathbf{A}] \cup U$, so by the Herbrand–Skolem theorem, $\mathbf{B}^*$ is a theorem of $T^\circ[\mathbf{A}^\circ] \cup U^\circ$. By the corollary to the consistency theorem, we can find instances $\mathbf{B}_1, \dots, \mathbf{B}_n$ of the matrix of $\mathbf{B}^*$ whose disjunction is a tautological consequence of instances of equality rules for $L(T^\circ[\mathbf{A}^\circ] \cup U^\circ)$ and instances of nonlogical axioms of $T^\circ[\mathbf{A}^\circ] \cup U^\circ$. Of those instances, let $\mathbf{C}_1, \dots, \mathbf{C}_l$ be the instances of the equality rules for $L(T^\circ[\mathbf{A}^\circ])$ and of the nonlogical axioms of $T^\circ[\mathbf{A}^\circ]$, and let $\mathbf{D}_1, \dots, \mathbf{D}_m$ be the instances of the other axioms. Then

$$(\mathbf{C}_1 \wedge \cdots \wedge \mathbf{C}_l) \to (\neg\mathbf{D}_1 \vee \cdots \vee \neg\mathbf{D}_m \vee \mathbf{B}_1 \vee \cdots \vee \mathbf{B}_n)$$

is a tautology. Substituting $\mathbf{e}$ for every free variable in the above formula and using Proposition 2 of ch. I §3.3, we can assume that the formula is closed. Using Lemma 1, we find a closed formula $\mathbf{C}$ of L° such that

$$(\mathbf{C}_1 \wedge \cdots \wedge \mathbf{C}_l) \to \mathbf{C} \text{ and } \mathbf{C} \to (\neg\mathbf{D}_1 \vee \cdots \vee \neg\mathbf{D}_m \vee \mathbf{B}_1 \vee \cdots \vee \mathbf{B}_n) \tag{1}$$

are tautologies. The first tautology in (1) tells us that $\mathbf{C}$ is a tautological consequence of instances of equality rules for $L(T^\circ[\mathbf{A}^\circ])$ and instances of nonlogical axioms of $T^\circ[\mathbf{A}^\circ]$. By the substitution rule and the tautology theorem, $\vdash_{T^\circ[\mathbf{A}^\circ]} \mathbf{C}$, so by the closure theorem and the deduction theorem, $\vdash_{T^\circ} \forall \mathbf{x}_1 \dots \forall \mathbf{x}_n \mathbf{A}^\circ \to \mathbf{C}$. Now $\forall \mathbf{x}_1 \dots \forall \mathbf{x}_n \mathbf{A}^\circ \to \mathbf{C}$ and $\mathbf{A} \to \mathbf{C}$ have a common Herbrand form, so $\vdash_{T'} \mathbf{A} \to \mathbf{C}$ by the Herbrand–Skolem theorem. The second tautology in (1) tells us that $\mathbf{B}_1 \vee \cdots \vee \mathbf{B}_n$ is a tautological consequence of $\mathbf{C}$, instances of equality rules for $L(U^\circ)$, and instances of nonlogical axioms of U°. By the corollary to the consistency theorem, $\vdash_{U^\circ[\mathbf{C}]} \mathbf{B}^*$, whence $\vdash_{U^\circ} \mathbf{C} \to \mathbf{B}^*$ by the deduction theorem. Since $\mathbf{C} \to \mathbf{B}^*$ and $\mathbf{C} \to \mathbf{B}$ have a common Herbrand form, we obtain $\vdash_{U'} \mathbf{C} \to \mathbf{B}$ by the Herbrand–Skolem theorem.

Now the proof of the proposition of ch. I §4.2 shows that if $\mathbf{x}$ is a variable not occurring in given derivations of $\mathbf{A} \to \mathbf{C}$ in T' and of $\mathbf{C} \to \mathbf{B}$ in U' and if $\mathbf{C}^*$ is obtained from $\mathbf{C}$ by replacing every occurrence of a term that is not a variable by $\mathbf{x}$, then $\vdash_T \mathbf{A} \to \mathbf{C}^*$ and $\vdash_U \mathbf{C}^* \to \mathbf{B}$. Then by the substitution axioms and the tautology theorem, we have $\vdash_T \mathbf{A} \to \exists \mathbf{x} \mathbf{C}^*$, and by the $\exists$-introduction rule, we have $\vdash_U \exists \mathbf{x} \mathbf{C}^* \to \mathbf{B}$. Thus, $\exists \mathbf{x} \mathbf{C}^*$ satisfies the conclusion of the lemma. □

Craig's Interpolation Lemma. Let T and U be compatible first-order theories. Let $\mathbf{A}$ be a formula of T and $\mathbf{B}$ a formula of U such that $\vdash_{T \cup U} \mathbf{A} \to \mathbf{B}$. Then there exists a formula $\mathbf{C}$ of $L(T) \cap L(U)$ such that $\vdash_T \mathbf{A} \to \mathbf{C}$ and $\vdash_U \mathbf{C} \to \mathbf{B}$. Moreover, $\mathbf{C}$ can be chosen so that the variables free in $\mathbf{C}$ are free in $\mathbf{A} \to \mathbf{B}$.

Proof. Suppose first that $\mathbf{A}$ and $\mathbf{B}$ are closed. Form $(T \cup U)^\$$ from $T \cup U$ as in §5.2. Form also $T^\$$ and $U^\$$ from T and U using the same symbols $\mathbf{p_f}$ that were used in forming $(T \cup U)^\$$. Then $(T \cup U)^\$$ is $T^\$ \cup U^\$$. By the proposition of §5.2, we have $\vdash_{T^\$ \cup U^\$} \mathbf{A}^\$ \to \mathbf{B}^\$$. By Lemma 2, we obtain a closed formula $\mathbf{C}$ of $L(T^\$) \cap L(U^\$)$ such that $\vdash_{T^\$} \mathbf{A}^\$ \to \mathbf{C}$ and $\vdash_{U^\$} \mathbf{C} \to \mathbf{B}^\$$. Then, again using the proposition of §5.2, $\vdash_T \mathbf{A} \to \mathbf{C}_\$$ and $\vdash_U \mathbf{C}_\$ \to \mathbf{B}$. Since $\mathbf{C}_\$$ is a closed formula of $L(T) \cap L(U)$, the theorem is proved in this case.

Suppose now that $\mathbf{A}$ and $\mathbf{B}$ are arbitrary. Let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be the free variables in $\mathbf{A} \to \mathbf{B}$, and let T' (resp. U') be obtained from T (resp. from U) by the adjunction of n new constants $\mathbf{e}_1, \dots, \mathbf{e}_n$. By the substitution rule, $\vdash_{T' \cup U'} \mathbf{A}[\mathbf{x}_1, \dots, \mathbf{x}_n | \mathbf{e}_1, \dots, \mathbf{e}_n] \to \mathbf{B}[\mathbf{x}_1, \dots, \mathbf{x}_n | \mathbf{e}_1, \dots, \mathbf{e}_n]$, so by the first part of the proof, we find a closed formula $\mathbf{C}$ of $L(T') \cap L(U')$ such that $\vdash_{T'} \mathbf{A}[\mathbf{x}_1, \dots, \mathbf{x}_n | \mathbf{e}_1, \dots, \mathbf{e}_n] \to \mathbf{C}$ and $\vdash_{U'} \mathbf{C} \to \mathbf{B}[\mathbf{x}_1, \dots, \mathbf{x}_n | \mathbf{e}_1, \dots, \mathbf{e}_n]$. Replacing $\mathbf{C}$ by a variant if necessary, we can assume that $\mathbf{C}$ has the form $\mathbf{C}'[\mathbf{x}_1, \dots, \mathbf{x}_n | \mathbf{e}_1, \dots, \mathbf{e}_n]$. Then by the theorem on constants, we obtain $\vdash_T \mathbf{A} \to \mathbf{C}'$ and $\vdash_U \mathbf{C}' \to \mathbf{B}$. □

The following is a reformulation of Craig's interpolation lemma that is useful for the study of consistency.

Joint Consistency Theorem. Let T and U be compatible first-order theories. Then $T \cup U$ is inconsistent if and only if there exists a closed formula $\mathbf{A}$ of $L(T) \cap L(U)$ such that $\vdash_T \mathbf{A}$ and $\vdash_U \neg\mathbf{A}$.

Proof. Suppose that $T \cup U$ is inconsistent. Applying Craig's interpolation lemma to the implication $\exists x(x = x) \to \neg\exists x(x = x)$, we find a closed formula $\mathbf{A}$ of $L(T) \cap L(U)$ such that $\vdash_T \exists x(x = x) \to \mathbf{A}$ and $\vdash_U \mathbf{A} \to \neg\exists x(x = x)$. Since $\vdash_T \exists x(x = x)$ and $\vdash_U \exists x(x = x)$ by the identity axioms, the substitution axioms, and the detachment rule, we obtain $\vdash_T \mathbf{A}$ and $\vdash_U \neg\mathbf{A}$ by the tautology theorem. Conversely, if $\vdash_T \mathbf{A}$ and $\vdash_U \neg\mathbf{A}$ for some formula $\mathbf{A}$, then $\mathbf{A}$ and $\neg\mathbf{A}$ are theorems of $T \cup U$, which is therefore inconsistent by the tautology theorem. $\square$

Chapter Three
The Incompleteness Theorem

§1 Number-theoretic functions and predicates

1.1 Recursive functions and predicates. In what follows, "function" and "predicate" are used exclusively for functions and predicates taking natural numbers as arguments and values. Also, a number is a natural number. In discussing natural numbers, we shall use the signs $\exists$, $\forall$, $\neg$, $\vee$, $\wedge$, $\rightarrow$, $\leftrightarrow$ with the usual meanings "for some", "for all", "not", "or", "and", "if … then", and "if and only if". The context will always distinguish them from the symbols and abbreviations of symbols of a formal system. We also use the expression μa to mean "the first (natural number) a such that". We use German letters to represent finite sequences of numbers, with conventions that will be apparent.

The following functions are called *initial functions*:

(i) the binary function $+$;

(ii) the binary function $\cdot$;

(iii) the binary function $\chi_<$ defined by $\chi_<(a,b) = \begin{cases} 0 & \text{if } a < b, \\ 1 & \text{otherwise;} \end{cases}$

(iv) for every n and every i with $n \geq 1$ and $1 \leq i \leq n$, the n-ary function π_i^n defined by $\pi_i^n(a_1, \ldots, a_n) = a_i$.

If g is a k-ary function and $h_1, \ldots, h_k$ are n-ary functions, the *composition* of $g, h_1, \ldots, h_k$ is the n-ary function f defined by $f(\mathfrak{a}) = g(h_1(\mathfrak{a}), \ldots, h_k(\mathfrak{a}))$. If g is an $(n+1)$-ary function and if $\forall\mathfrak{a}\exists a(g(\mathfrak{a}, a) = 0)$, the *minimization* of g is the n-ary function f defined by $f(\mathfrak{a}) = \mu a(g(\mathfrak{a}, a) = 0)$. A *recursive function* is one that is obtained from the initial functions through composition and minimization.

If p is a predicate, we define its *representing function* χ_p by

$$\chi_p(\mathfrak{a}) = \begin{cases} 0 & \text{if } p(\mathfrak{a}), \\ 1 & \text{otherwise.} \end{cases}$$

We then say that a predicate is *recursive* if its representing function is recursive.

We now describe general rules to decide whether certain functions or predicates are recursive.

(i) If p is a recursive k-ary predicate, if $h_1, \ldots, h_k$ are recursive n-ary functions, and if $q(\mathfrak{a}) \leftrightarrow p(h_1(\mathfrak{a}), \ldots, h_k(\mathfrak{a}))$, then q is recursive.

The representing function of q is the composition of $\chi_p, h_1, \ldots, h_k$, and hence is recursive.

(ii) If p is a recursive $(n+1)$-ary predicate such that $\forall\mathfrak{a}\exists a p(\mathfrak{a}, a)$ and if $f(\mathfrak{a}) = \mu a p(\mathfrak{a}, a)$, then f is recursive.

Indeed, f is the minimization of χ_p. In this situation, we say that f is the *minimization of p*.

We denote by c_k^n the n-ary function defined by $c_k^n(\mathfrak{a}) = k$.

(iii) For all n and k, c_k^n is recursive.

We prove this by induction on k. For $k = 0$, c_0^n is the minimization of π_{n+1}^{n+1}. If $k > 0$, then c_k^n is the minimization of p where $p(\mathfrak{a}, a) \leftrightarrow c_{k-1}^{n+1}(\mathfrak{a}, a) < \pi_{n+1}^{n+1}(\mathfrak{a}, a)$. But p is recursive by (i) and the induction hypothesis, so c_k^n is recursive by (ii).

(iv) If p and q are recursive n-ary predicates, then so are $\neg p$, $p \vee q$, $p \wedge q$, $p \rightarrow q$, and $p \leftrightarrow q$.

We have $\chi_{\neg p}(\mathfrak{a}) = \chi_<(c_0^n(\mathfrak{a}), \chi_p(\mathfrak{a}))$ and $\chi_{p \vee q}(\mathfrak{a}) = \chi_p(\mathfrak{a}) \cdot \chi_q(\mathfrak{a})$, so $\neg p$ and $p \vee q$ are recursive. Since $p \wedge q$ is $\neg(\neg p \vee \neg q)$, $p \rightarrow q$ is $\neg p \vee q$, and $p \leftrightarrow q$ is $(p \rightarrow q) \wedge (q \rightarrow p)$, the other cases follow.

We denote by $\top^n$ the n-ary predicate defined by $\top^n(\mathfrak{a}) \leftrightarrow 0 < 1$ and by $\bot^n$ the n-ary predicate $\neg\top^n$. Then by (i), (iii), and (iv),

(v) For all n, $\top^n$ and $\bot^n$ are recursive.

We say that predicates $r_1, \dots, r_k$ are *mutually exclusive* if $r_i(\mathfrak{a}) \to \neg r_j(\mathfrak{a})$ whenever $i \neq j$.

(vi) Let $g_1, \dots, g_{k+1}$ be recursive n-ary functions and $r_1, \dots, r_k$ mutually exclusive recursive n-ary predicates. If

$$f(\mathfrak{a}) = \begin{cases} g_1(\mathfrak{a}) & \text{if } r_1(\mathfrak{a}), \\ \vdots & \vdots \\ g_k(\mathfrak{a}) & \text{if } r_k(\mathfrak{a}), \\ g_{k+1}(\mathfrak{a}) & \text{otherwise,} \end{cases}$$

then f is recursive.

(vii) Let $q_1, \dots, q_{k+1}$ be recursive n-ary predicates and $r_1, \dots, r_k$ mutually exclusive recursive n-ary predicates. If

$$p(\mathfrak{a}) \leftrightarrow \begin{cases} q_1(\mathfrak{a}) & \text{if } r_1(\mathfrak{a}), \\ \vdots & \vdots \\ q_k(\mathfrak{a}) & \text{if } r_k(\mathfrak{a}), \\ q_{k+1}(\mathfrak{a}) & \text{otherwise,} \end{cases}$$

then p is recursive.

We have $f(\mathfrak{a}) = g_1(\mathfrak{a}) \cdot \chi_{r_1}(\mathfrak{a}) + \cdots + g_k(\mathfrak{a}) \cdot \chi_{r_k}(\mathfrak{a}) + g_{k+1}(\mathfrak{a}) \cdot \chi_{\neg r_1 \wedge \cdots \wedge \neg r_k}(\mathfrak{a})$ and $p(\mathfrak{a}) \leftrightarrow (r_1(\mathfrak{a}) \wedge q_1(\mathfrak{a})) \vee \cdots \vee (r_k(\mathfrak{a}) \wedge q_k(\mathfrak{a})) \vee (\neg r_1(\mathfrak{a}) \wedge \cdots \wedge \neg r_k(\mathfrak{a}) \wedge q_{k+1}(\mathfrak{a}))$.

(viii) The binary predicates $\leq$ and $=$ are recursive.

This follows from the relations $a \leq b \leftrightarrow \neg(\pi_2^2(a, b) < \pi_1^2(a, b))$ and $a = b \leftrightarrow a \leq b \wedge \pi_2^2(a, b) \leq \pi_1^2(a, b)$.

Since we do not deal with negative integers, we set $a - b$ to be 0 if $a < b$.

(ix) The binary function $-$ is recursive.

The function $-$ is the minimization of the ternary predicate p defined by $p(a, b, c) \leftrightarrow \pi_1^3(a, b, c) = \pi_2^3(a, b, c) + \pi_3^3(a, b, c) \vee \pi_1^3(a, b, c) < \pi_2^3(a, b, c)$.

We introduce some abbreviations. We write $\exists a_{<\cdots}$ and $\forall a_{<\cdots}$ to mean "for some $a < \cdots$" and "for all $a < \cdots$", respectively. We also write $\mu a_{<\cdots}$ for "the first a such that $a = \cdots$ or" (note that such an a always exists).

(x) If p is a recursive $(n+1)$-ary predicate and if $f(\mathfrak{a}, a) = \mu b_{<a} p(\mathfrak{a}, b)$, then f is recursive.

The function f is the minimization of the $(n+2)$-ary predicate q where $q(\mathfrak{a}, a, b) \leftrightarrow \pi_{n+2}^{n+2}(\mathfrak{a}, a, b) = \pi_{n+1}^{n+2}(\mathfrak{a}, a, b) \vee p(\pi_1^{n+2}(\mathfrak{a}, a, b), \dots, \pi_n^{n+2}(\mathfrak{a}, a, b), \pi_{n+2}^{n+2}(\mathfrak{a}, a, b))$. But q is recursive by our previous results, so f is recursive.

(xi) If p is a recursive $(n+1)$-ary predicate and if $q(\mathfrak{a}, a) \leftrightarrow \exists b_{<a} p(\mathfrak{a}, b)$ and $r(\mathfrak{a}, a) \leftrightarrow \forall b_{<a} p(\mathfrak{a}, b)$, then q and r are recursive.

We have $q(\mathfrak{a}, a) \leftrightarrow \mu b_{<a} p(\mathfrak{a}, b) < \pi_{n+1}^{n+1}(\mathfrak{a}, a)$ and $r(\mathfrak{a}, a) \leftrightarrow \neg\exists b_{<a} \neg p(\mathfrak{a}, b)$.

1.2 Coding functions. A binary function β is called a *coding function* if

(i) for all a and b, $\beta(a, b) \leq a - 1$;

(ii) for all $n \geq 1$ and for all $a_0, \dots, a_{n-1}$, there is a number a such that, for all $i < n$, $\beta(a, i) = a_i$.

Theorem. The function β defined by

$$\beta(a, i) = \mu b_{<a-1} \exists c_{<a} \exists d_{<a} (a = (c + d)(c + d) + c + 1 \wedge \exists e_{<a} (c = (((b + i)(b + i) + b + 2)d + 1)e))$$

is a recursive coding function.

We do not give the proof of this theorem here since it can be recovered from the derivation of a formalized version of this theorem which we shall give in full in chapter IV.

We now fix a coding function β. Thus if $a_1, \ldots, a_n$ are numbers, there exists a number a such that $\beta(a,0) = n$, $\beta(a,1) = a_1, \ldots, \beta(a,n) = a_n$; we denote the first such a by $\langle a_1, \ldots, a_n \rangle$ (the function β will always be fixed throughout a given discussion, so we shall not find it necessary to indicate it in the notation). As an n-ary function of $a_1, \ldots, a_n$, $\langle a_1, \ldots, a_n \rangle$ is called a *sequence function*. If β is recursive, the n-ary sequence function is recursive for every n, since $\langle a_1, \ldots, a_n \rangle = \mu a(\beta(a,0) = n \wedge \beta(a,1) = a_1 \wedge \cdots \wedge \beta(a,n) = a_n)$ and β is recursive. Note that $\beta(0,0) = 0$ because of (i), and hence $\langle\rangle = 0$. We abbreviate $\beta(a, i+1)$ to $(a)_i$ and we write $(a)_{i,j}$ instead of $((a)_i)_j$. We shall sometimes drop the parentheses when they are not needed: thus we may write $f(a,b)_i$ instead of $(f(a,b))_i$. We further define

(i) $\mathrm{len}(a) = \beta(a,0)$;

(ii) $\mathrm{sq}(a)$ if and only if there exists n and $a_1, \ldots, a_n$ such that $a = \langle a_1, \ldots, a_n \rangle$;

(iii) $a \in b$ if and only if $b = \langle a_1, \ldots, a_n \rangle$ and $a = a_i$ for some i;

(iv) $a * b = \begin{cases} \langle a_1, \ldots, a_n, b_1, \ldots, b_m \rangle & \text{if } a = \langle a_1, \ldots, a_n \rangle \text{ and } b = \langle b_1, \ldots, b_m \rangle, \\ 0 & \text{otherwise;} \end{cases}$

(v) $\mathrm{ini}(a,i) = \begin{cases} \langle a_1, \ldots, a_i \rangle & \text{if } a = \langle a_1, \ldots, a_n \rangle \text{ and } i \le n, \\ 0 & \text{otherwise;} \end{cases}$

(vi) $\mathrm{rmv}(a,i) = \begin{cases} \langle a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n \rangle & \text{if } a = \langle a_1, \ldots, a_n \rangle \text{ and } 1 \le i \le n \\ a & \text{otherwise.} \end{cases}$

In (vi), $\langle a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n \rangle$ is to be read as $\langle a_2, \ldots, a_n \rangle$ if $n \ge 2$ and $i = 1$, as $\langle a_1, \ldots, a_{n-1} \rangle$ if $n \ge 2$ and $i = n$, and as $\langle\rangle$ if $n = 1$ and $i = 1$.

Suppose that β is recursive. Then obviously (i) is recursive. The recursiveness of (ii)–(vi) follows from the relations

(ii) $\mathrm{sq}(a) \leftrightarrow \forall b_{<a} \exists i_{<\mathrm{len}(a)+1}(\beta(a,i) \ne \beta(b,i))$;

(iii) $a \in b \leftrightarrow \mathrm{sq}(b) \wedge \exists i_{<\mathrm{len}(b)}(a = (b)_i)$;

(iv) $a * b = \begin{cases} \mu c(\mathrm{len}(c) = \mathrm{len}(a) + \mathrm{len}(b) \wedge \forall i_{<\mathrm{len}(a)}((c)_i = (a)_i) \wedge \forall i_{<\mathrm{len}(b)}((c)_{i+\mathrm{len}(a)} = (b)_i)) \\ \quad \text{if } \mathrm{sq}(a) \wedge \mathrm{sq}(b), \\ 0 \quad \text{otherwise;} \end{cases}$

(v) $\mathrm{ini}(a,i) = \begin{cases} \mu b(\mathrm{len}(b) = i \wedge \forall j_{<i}((b)_j = (a)_j)) & \text{if } \mathrm{sq}(a) \wedge i \le \mathrm{len}(a), \\ 0 & \text{otherwise;} \end{cases}$

(vi) $\mathrm{rmv}(a,i) =$
$$\begin{cases} \mu b(\mathrm{len}(b) = \mathrm{len}(a) - 1 \wedge \forall c_{<\mathrm{len}(b)}((c+1 < i \wedge (b)_c = (a)_c) \vee (i \le c+1 \wedge (b)_c = (a)_{c+1}))) \\ \quad \text{if } \mathrm{sq}(a) \wedge i \ne 0 \wedge i \le \mathrm{len}(a) \wedge \mathrm{len}(a) \ne 0 \\ a \quad \text{otherwise.} \end{cases}$$

1.3 Recursion. Fix a recursive coding function β. If f is an $(n+1)$-ary function, we define an $(n+1)$-ary function $\bar{f}$ by $\bar{f}(0,\mathfrak{a}) = 0$ and, for $a \ge 1$, $\bar{f}(a,\mathfrak{a}) = \langle f(0,\mathfrak{a}), \ldots, f(a-1,\mathfrak{a}) \rangle$.

(i) The function f is recursive if and only if the function $\bar{f}$ is recursive.

This follows from the relations $\bar{f}(a,\mathfrak{a}) = \mu b(\mathrm{len}(b) = a \wedge \forall i_{<a}((b)_i = f(i,\mathfrak{a})))$ and $f(a,\mathfrak{a}) = \beta(\bar{f}(a+1,\mathfrak{a}), a+1)$.

(ii) If g is an $(n+2)$-ary recursive function and if $f(a,\mathfrak{a}) = g(\bar{f}(a,\mathfrak{a}), a, \mathfrak{a})$, then f is recursive.

By (i), we need only prove that $\bar{f}$ is recursive. But $\bar{f}(a,\mathfrak{a}) = \mu b(\mathrm{len}(b) = a \wedge \forall i_{<a}((b)_i = g(\mathrm{ini}(b,i), i, \mathfrak{a})))$, and so $\bar{f}$ is recursive.

1.4 Recursively enumerable predicates. An n-ary predicate p is *recursively enumerable* if there is a recursive $(n+1)$-ary predicate q such that $p(\mathfrak{a}) \leftrightarrow \exists a q(\mathfrak{a}, a)$.

Negation Lemma. Let p be an n-ary predicate. If p and $\neg p$ are recursively enumerable, then p is recursive.

Proof. There are recursive predicates q and r such that $p(\mathfrak{a}) \leftrightarrow \exists a q(\mathfrak{a}, a)$ and $\neg p(\mathfrak{a}) \leftrightarrow \exists a r(\mathfrak{a}, a)$. Then $\forall \mathfrak{a} \exists a(q(\mathfrak{a}, a) \vee r(\mathfrak{a}, a))$, so we can define a function f by $f(\mathfrak{a}) = \mu a(\chi_{q\vee r}(\mathfrak{a}, a) = 0)$; f is recursive being the minimization of the composition of $\cdot$, χ_q, and χ_r. Clearly $\chi_p(\mathfrak{a}) = \chi_q(\pi_1^n(\mathfrak{a}), \dots, \pi_n^n(\mathfrak{a}), f(\mathfrak{a}))$, so p is recursive. □

In this paragraph we examine briefly some properties of recursively enumerable predicates. We fix a recursive coding function β.

(i) Recursive predicates are recursively enumerable.

Let p be a recursive predicate. Define $q(\mathfrak{a}, a) \leftrightarrow p(\mathfrak{a})$. Then q is recursive and $p(\mathfrak{a}) \leftrightarrow \exists a q(\mathfrak{a}, a)$, so p is recursively enumerable.

(ii) If p is a recursively enumerable k-ary predicate, if $h_1, \dots, h_k$ are recursive n-ary functions, and if $q(\mathfrak{a}) \leftrightarrow p(h_1(\mathfrak{a}), \dots, h_k(\mathfrak{a}))$, then q is recursively enumerable.

There is a recursive $(k+1)$-ary predicate p' such that $p(\mathfrak{b}) \leftrightarrow \exists a p'(\mathfrak{b}, a)$. Define the $(n+1)$-ary predicate q' by $q'(\mathfrak{a}, a) \leftrightarrow p'(h_1(\mathfrak{a}), \dots, h_k(\mathfrak{a}), a)$. Then q' is recursive, and since $q(\mathfrak{a}) \leftrightarrow \exists a q'(\mathfrak{a}, a)$, q is recursively enumerable.

(iii) If p is a recursively enumerable $(n+1)$-ary predicate and if $q(\mathfrak{a}) \leftrightarrow \exists a p(\mathfrak{a}, a)$, then q is recursively enumerable.

Let p' be a recursive predicate such that $p(\mathfrak{a}, a) \leftrightarrow \exists b p'(\mathfrak{a}, a, b)$. Define q' by $q'(\mathfrak{a}, c) \leftrightarrow p'(\mathfrak{a}, (c)_0, (c)_1)$, so that q' is recursive. Clearly $\exists a \exists b p'(\mathfrak{a}, a, b) \leftrightarrow \exists c q'(\mathfrak{a}, c)$, and since $q(\mathfrak{a}) \leftrightarrow \exists a \exists b p'(\mathfrak{a}, a, b)$, q is recursively enumerable.

(iv) If p and q are recursively enumerable, then $p \vee q$ and $p \wedge q$ are recursively enumerable.

Let p', q' be recursive predicates such that $p(\mathfrak{a}) \leftrightarrow \exists a p'(\mathfrak{a}, a)$ and $q(\mathfrak{a}) \leftrightarrow \exists a q'(\mathfrak{a}, a)$. Then $p(\mathfrak{a}) \vee q(\mathfrak{a}) \leftrightarrow \exists a(p'(\mathfrak{a}, a) \vee q'(\mathfrak{a}, a))$ and $p(\mathfrak{a}) \wedge q(\mathfrak{a}) \leftrightarrow \exists a \exists b(p'(\mathfrak{a}, a) \wedge q'(\mathfrak{a}, b))$, which shows that $p \vee q$ is recursively enumerable and, using (iii), that $p \wedge q$ is recursively enumerable.

(v) If p is a recursively enumerable $(n+1)$-ary predicate and if $q(\mathfrak{a}, b) \leftrightarrow \forall a_{<b} p(\mathfrak{a}, a)$, then q is recursively enumerable.

Let r be a recursive predicate such that $p(\mathfrak{a}, a) \leftrightarrow \exists c r(\mathfrak{a}, a, c)$. Then $q(\mathfrak{a}, b) \leftrightarrow \exists c \forall a_{<b} r(\mathfrak{a}, a, (c)_a)$ and so q is recursively enumerable.

1.5 Graph of a function. If f is an n-ary function, the *graph* of f is the $(n+1)$-ary predicate $\mathfrak{G}f$ defined by $\mathfrak{G}f(\mathfrak{a}, b) \leftrightarrow f(\mathfrak{a}) = b$.

Proposition. Let f be a function. The following assertions are equivalent: f is recursive; $\mathfrak{G}f$ is recursive; $\mathfrak{G}f$ is recursively enumerable.

Proof. It is obvious that the first assertion implies the second, and the second implies the third by (i) of §1.4. Suppose that $\mathfrak{G}f$ is recursively enumerable. Then

$$\mathfrak{G}f(\mathfrak{a}, b) \leftrightarrow \exists c p(\mathfrak{a}, b, c) \tag{1}$$

for some recursive predicate p. In particular, $\forall \mathfrak{a} \exists c p(\mathfrak{a}, f(\mathfrak{a}), c)$ and hence $\forall \mathfrak{a} \exists c p(\mathfrak{a}, (c)_0, (c)_1)$. Thus we can define a recursive function g by $g(\mathfrak{a}) = \mu c p(\mathfrak{a}, (c)_0, (c)_1)$. We claim that $f(\mathfrak{a}) = g(\mathfrak{a})_0$, which will prove that f is recursive. Indeed we have $p(\mathfrak{a}, g(\mathfrak{a})_0, g(\mathfrak{a})_1)$, and so $\exists c p(\mathfrak{a}, g(\mathfrak{a})_0, c)$. The claim follows by (1). □

§2 Representability

2.1 Numerical languages. A first-order language L is *numerical* if it has the constant $\dot{0}$ and the unary function symbol S. A *numeral* of L is a term of L in which only $\dot{0}$ and S occur. If n is a natural number, we denote by $\dot{n}$ the numeral of L having n occurrences of S, also called the numeral *of* n. If n is 0, this notation coincides with the constant $\dot{0}$. If f is a function, $\dot{f}$ is the function which to $\mathfrak{a}$ associates the numeral of $f(\mathfrak{a})$. A *numerical instance* of a formula $\mathbf{A}$ of L is a closed formula of the form $\mathbf{A}[\mathbf{x}_1, \ldots, \mathbf{x}_n|\dot{a}_1, \ldots, \dot{a}_n]$. A first-order theory T is called *numerical* if $L(T)$ is numerical. If L and M are numerical first-order languages, an interpretation I of L in M is *numerical* if $=_I$ is $=$, $\dot{0}_I$ is $\dot{0}$, and S_I is S.

2.2 Representability of functions. In this section, T is a numerical first-order theory. Let f be an n-ary function, $\mathbf{a}$ a term of T, and $\mathbf{x}_1, \ldots, \mathbf{x}_n$ distinct variables. We say that $\mathbf{a}$ *with* $\mathbf{x}_1, \ldots, \mathbf{x}_n$ *represents* f *in* T if for any $a_1, \ldots, a_n$, $\vdash_T \mathbf{a}[\mathbf{x}_1, \ldots, \mathbf{x}_n|\dot{a}_1, \ldots, \dot{a}_n] = \dot{f}(a_1, \ldots, a_n)$. We simply say that $\mathbf{f}$ represents f in T if $\mathbf{f}x_1 \ldots x_n$ with $x_1, \ldots, x_n$ represents f in T.

Let f be an n-ary function, $\mathbf{A}$ a formula of T, and $\mathbf{x}_1, \ldots, \mathbf{x}_n$, and $\mathbf{y}$ distinct variables. We say that $\mathbf{A}$ *with* $\mathbf{x}_1, \ldots, \mathbf{x}_n$, $\mathbf{y}$ *represents* f *in* T if for any $a_1, \ldots, a_n$, $\vdash_T \mathbf{A}[\mathbf{x}_1, \ldots, \mathbf{x}_n|\dot{a}_1, \ldots, \dot{a}_n] \leftrightarrow \mathbf{y} = \dot{f}(a_1, \ldots, a_n)$. An n-ary function f is said to be *representable in* T if there is a formula $\mathbf{A}$ of T and distinct variables $\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{y}$ such that $\mathbf{A}$ with $\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{y}$ represents f in T.

Proposition. A function f is representable in T if and only if there exist an extension by definitions T' of T, a term $\mathbf{a}$ of T', and variables $\mathbf{x}_1, \ldots, \mathbf{x}_n$ such that $\mathbf{a}$ with $\mathbf{x}_1, \ldots, \mathbf{x}_n$ represents f in T'.

Proof. Suppose that $\mathbf{A}$ with $\mathbf{x}_1, \ldots, \mathbf{x}_n$, $\mathbf{y}$ represents f in T. Let $\mathbf{y}'$ be distinct from $\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{y}$ and not occurring in $\mathbf{A}$ and let $\mathbf{D}$ be the formula

$$(\exists \mathbf{y}\mathbf{A} \wedge \forall \mathbf{y}'(\mathbf{A} \to \mathbf{A}[\mathbf{y}|\mathbf{y}'] \to \mathbf{y} = \mathbf{y}') \wedge \mathbf{A}) \vee (\neg(\exists \mathbf{y}\mathbf{A} \wedge \forall \mathbf{y}'(\mathbf{A} \to \mathbf{A}[\mathbf{y}|\mathbf{y}'] \to \mathbf{y} = \mathbf{y}')) \wedge \mathbf{y} = \dot{0}).$$

We obtain T' from T by the adjunction of a new n-ary function symbol $\mathbf{f}$ and the axiom $\mathbf{y} = \mathbf{f}\mathbf{x}_1 \ldots \mathbf{x}_n \leftrightarrow \mathbf{D}$. It is easy to derive existence and uniqueness conditions for $\mathbf{y}$ in $\mathbf{D}$, so that T' is an extension by definitions of T. Clearly $\mathbf{f}\mathbf{x}_1 \ldots \mathbf{x}_n$ with $\mathbf{x}_1, \ldots, \mathbf{x}_n$ represents f in T'.

Conversely, suppose that $\mathbf{a}$ with $\mathbf{x}_1, \ldots, \mathbf{x}_n$ represents f in some extension by definitions T' of T. Let $\mathbf{A}$ be a translation of $\mathbf{y} = \mathbf{a}$ into T. Since $\vdash_{T'} \mathbf{y} = \mathbf{a}[\mathbf{x}_1, \ldots, \mathbf{x}_n|\dot{a}_1, \ldots, \dot{a}_n] \leftrightarrow \mathbf{y} = \dot{f}(a_1, \ldots, a_n)$ by the equality theorem, we have $\vdash_T \mathbf{A}[\mathbf{x}_1, \ldots, \mathbf{x}_n|\dot{a}_1, \ldots, \dot{a}_n] \leftrightarrow \mathbf{y} = \dot{f}(a_1, \ldots, a_n)$ by the theorem on definitions, so $\mathbf{A}$ with $\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{y}$ represents f in T. □

2.3 Representability of predicates. Let p be an n-ary predicate, $\mathbf{A}$ a formula of T, and $\mathbf{x}_1, \ldots, \mathbf{x}_n$ distinct variables. We say that $\mathbf{A}$ *with* $\mathbf{x}_1, \ldots, \mathbf{x}_n$ *represents* p *in* T if for all $a_1, \ldots, a_n$, $p(a_1, \ldots, a_n)$ implies $\vdash_T \mathbf{A}[\mathbf{x}_1, \ldots, \mathbf{x}_n|\dot{a}_1, \ldots, \dot{a}_n]$ and $\neg p(a_1, \ldots, a_n)$ implies $\vdash_T \neg\mathbf{A}[\mathbf{x}_1, \ldots, \mathbf{x}_n|\dot{a}_1, \ldots, \dot{a}_n]$. If only the former (resp. the latter) holds, we say that $\mathbf{A}$ with $\mathbf{x}_1, \ldots, \mathbf{x}_n$ *positively represents* (resp. *negatively represents*) p in T. We say that $\mathbf{p}$ represents p in T if $\mathbf{p}x_1 \ldots x_n$ with $x_1, \ldots, x_n$ represents p in T. An n-ary predicate p is said to be *representable in* T if there is a formula $\mathbf{A}$ of T and distinct variables $\mathbf{x}_1, \ldots, \mathbf{x}_n$ such that $\mathbf{A}$ with $\mathbf{x}_1, \ldots, \mathbf{x}_n$ represents p in T.

Proposition. If $\vdash_T \dot{0} \neq \dot{1}$, a predicate p is representable in T if and only if χ_p is representable in T.

Proof. Suppose that $\mathbf{A}$ with $\mathbf{x}_1, \ldots, \mathbf{x}_n$ represents p in T, and let $\mathbf{y}$ be distinct from $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Let $a_1, \ldots, a_n$ be natural numbers, and let $\mathbf{A}'$ abbreviate $\mathbf{A}[\mathbf{x}_1, \ldots, \mathbf{x}_n|\dot{a}_1, \ldots, \dot{a}_n]$. If $p(a_1, \ldots, a_n)$, then $\vdash_T \mathbf{A}'$, and if $\neg p(a_1, \ldots, a_n)$, then $\vdash_T \neg\mathbf{A}'$. In both cases, $\vdash_T (\mathbf{A}' \wedge \mathbf{y} = \dot{0}) \vee (\neg\mathbf{A}' \wedge \mathbf{y} = \dot{1}) \leftrightarrow \mathbf{y} = \dot{\chi}_p(a_1, \ldots, a_n)$ by the tautology theorem. So $(\mathbf{A} \wedge \mathbf{y} = \dot{0}) \vee (\neg\mathbf{A} \wedge \mathbf{y} = \dot{1})$ with $\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{y}$ represents χ_p.

Conversely, suppose that $\mathbf{A}$ with $\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{y}$ represents χ_p in T. If $p(a_1, \ldots, a_n)$, then $\chi_p(a_1, \ldots, a_n) = 0$ and so $\vdash_T \mathbf{A}[\mathbf{x}_1, \ldots, \mathbf{x}_n|\dot{a}_1, \ldots, \dot{a}_n] \leftrightarrow \mathbf{y} = \dot{0}$. By the substituion rule, $\vdash_T \mathbf{A}[\mathbf{y}|\dot{0}][\mathbf{x}_1, \ldots, \mathbf{x}_n|\dot{a}_1, \ldots, \dot{a}_n] \leftrightarrow \dot{0} = \dot{0}$, whence $\vdash_T \mathbf{A}[\mathbf{y}|\dot{0}][\mathbf{x}_1, \ldots, \mathbf{x}_n|\dot{a}_1, \ldots, \dot{a}_n]$ by the identity axioms, the substitution rule, and the tautology theorem. If $\neg p(a_1, \ldots, a_n)$, we find similarly $\vdash_T \mathbf{A}[\mathbf{y}|\dot{0}][\mathbf{x}_1, \ldots, \mathbf{x}_n|\dot{a}_1, \ldots, \dot{a}_n] \leftrightarrow \dot{0} = \dot{1}$, whence $\vdash_T \neg\mathbf{A}[\mathbf{y}|\dot{0}][\mathbf{x}_1, \ldots, \mathbf{x}_n|\dot{a}_1, \ldots, \dot{a}_n]$ by the hypothesis and the tautology theorem. Thus $\mathbf{A}[\mathbf{y}|\dot{0}]$ with $\mathbf{x}_1, \ldots, \mathbf{x}_n$ represents p in T. □

2.4 Representability and interpretations. We remark that if an n-ary function f is representable in T and if $\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{y}$ are distinct variables, then there is a formula $\mathbf{A}$ of T which with $\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{y}$ represents f in T and in which no variable other than $\mathbf{x}_1, \ldots, \mathbf{x}_n$, and $\mathbf{y}$ is free. For if $\mathbf{B}$ with $\mathbf{z}_1, \ldots, \mathbf{z}_n$, $\mathbf{w}$ represents f in T and if $\mathbf{x}'_1, \ldots, \mathbf{x}'_m$ are the variables free in $\mathbf{B}$ other than $\mathbf{z}_1, \ldots, \mathbf{z}_n$, and $\mathbf{w}$, then the formula

$\mathbf{B}'[\mathbf{x}_1', \ldots, \mathbf{x}_m', \mathbf{z}_1, \ldots, \mathbf{z}_n, \mathbf{w}|\dot{0}, \ldots, \dot{0}, \mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{y}]$, where $\mathbf{B}'$ is a variant of $\mathbf{B}$ in which $\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{y}$ are not bound, is as desired by the version theorem. A similar statement and its proof hold for predicates.

PROPOSITION. Let T and U be numerical first-order theories and let I be a numerical interpretation of T in U. Then any function or predicate representable in T is representable in U.

Proof. Let f be an n-ary function representable in T. By the remark preceding the proposition, we can find a formula $\mathbf{A}$ which with $\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{y}$ represents f in T and in which no variable other than $\mathbf{x}_1, \ldots, \mathbf{x}_n$, and $\mathbf{y}$ is free. Let $b = f(a_1, \ldots, a_n)$. By the interpretation theorem and the fact that $\dot{n}_I$ is $\dot{n}$ for all n,

$$\vdash_U \mathrm{U}_I\mathbf{y} \to \mathbf{A}_I[\mathbf{x}_1, \ldots, \mathbf{x}_n|\dot{a}_1, \ldots, \dot{a}_n] \leftrightarrow \mathbf{y} = \dot{b}. \tag{1}$$

But $\vdash_U \mathrm{U}_I\dot{b}$ because $\dot{b}_I$ is $\dot{b}$ and I is an interpretation of $L(T)$ in U, and hence $\vdash_U \mathbf{y} = \dot{b} \to \mathrm{U}_I\mathbf{y}$ by the equality theorem. From this and (1) we obtain $\vdash_U (\mathrm{U}_I\mathbf{y} \land \mathbf{A}_I[\mathbf{x}_1, \ldots, \mathbf{x}_n|\dot{a}_1, \ldots, \dot{a}_n]) \leftrightarrow \mathbf{y} = \dot{b}$ by the tautology theorem, so $\mathrm{U}_I\mathbf{y} \land \mathbf{A}_I$ with $\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{y}$ represents f in U.

Suppose that $\mathbf{A}$ with $\mathbf{x}_1, \ldots, \mathbf{x}_n$ represents a predicate p in T and that no variable other than $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is free in $\mathbf{A}$. Then $\mathbf{A}[\mathbf{x}_1, \ldots, \mathbf{x}_n|\dot{a}_1, \ldots, \dot{a}_n]^I$ is $\mathbf{A}_I[\mathbf{x}_1, \ldots, \mathbf{x}_n|\dot{a}_1, \ldots, \dot{a}_n]$ and $(\neg\mathbf{A}[\mathbf{x}_1, \ldots, \mathbf{x}_n|\dot{a}_1, \ldots, \dot{a}_n])^I$ is $\neg\mathbf{A}_I[\mathbf{x}_1, \ldots, \mathbf{x}_n|\dot{a}_1, \ldots, \dot{a}_n]$. It follows from the interpretation theorem that $\mathbf{A}_I$ with $\mathbf{x}_1, \ldots, \mathbf{x}_n$ represents p in U. □

§3 Arithmetizations

3.1 Arithmetizations. An *arithmetization* of a first-order language L is an effective mapping that assigns a natural number to every designator of L and to every sequence of formulae of L (including the empty sequence) in such a way that different numbers are assigned to different designators, and different numbers are assigned to different sequences of formulae. The number assigned to an object is called its *expression number*. By an "effective" mapping we mean not only that we can determine the expression number of a given designator or sequence of formulae, but also the following: if we are given a number, we can decide whether it is the expression number of a designator (resp. of a sequence of formulae) or not, and if it is, we can determine the designator (resp. the sequence of formulae) of which it is the expression number.†

A first-order language will be called *arithmetized* when it is endowed with an arithmetization. If $\mathbf{u}$ is a designator of an arithmetized first-order language, we write '$\mathbf{u}$' its expression number, and $\ulcorner\mathbf{u}\urcorner$ the numeral of '$\mathbf{u}$'. This notation can be ambiguous when we are dealing with several arithmetized languages, so we shall occasionally refine these conventions. We say that a first-order theory is arithmetized if its language is arithmetized.

If L is an arithmetized first-order language and T an arithmetized first-order theory, we define

(i) $\mathrm{vble}_L(a)$ if and only if a is the expression number of a variable;

(ii) $\mathrm{tm}_L(a)$ if and only if a is the expression number of a term of L;

(iii) $\mathrm{atfm}_L(a)$ if and only if a is the expression number of an atomic formula of L;

(iv) $\mathrm{fm}_L(a)$ if and only if a is the expression number of a formula of L;

(v) $\mathrm{des}_L(a)$ if and only if a is the expression number of a designator of L;

(vi) $\mathrm{occ}_L(a, b)$ if and only if a and b are expression numbers of designators of L and the designator with expression number b occurs in the designator with expression number a;

(vii) $\mathrm{fr}_L(a, b)$ if and only if $\mathrm{des}_L(a)$ and $\mathrm{vble}_L(b)$ and the variable with expression number b is free in the designator with expression number a;

(viii) $\mathrm{cl}_L(a)$ if and only if a is the expression number of a closed designator of L;

(ix) $\mathrm{subtl}_L(a, b, c)$ if and only if $\mathrm{des}_L(a)$ and $\mathrm{vble}_L(b)$ and $\mathrm{tm}_L(c)$ and the term with expression number c is substitutible for the variable with expression number b in the designator with expression number a;

(x) $\mathrm{pax}_L(a)$ if and only if a is the expression number of a propositional axiom for L;

(xi) $\mathrm{sax}_L(a)$ if and only if a is the expression number of a substitution axiom for L;

†This requirement that arithmetizations be effective is of course not used in any proof and hence is not strictly necessary. However, some proofs would lose their constructive character if they were applied to noneffective arithmetizations.

(xii) $\mathrm{iax}_L(a)$ if and only if a is the expression number of an identity axiom for L;

(xiii) $\mathrm{feax}_L(a)$ if and only if a is the expression number of a functional equality axiom for L;

(xiv) $\mathrm{peax}_L(a)$ if and only if a is the expression number of a predicative equality axiom for L;

(xv) $\mathrm{ax}_L(a)$ if and only if a is the expression number of a logical axiom for L;

(xvi) $\mathrm{ctr}_L(a, b)$ if and only if $\mathrm{fm}_L(a)$ and $\mathrm{fm}_L(b)$ and the formula with expression number a is the conclusion of a contraction rule whose premise has expression number b;

(xvii) $\mathrm{exp}_L(a, b)$ if and only if $\mathrm{fm}_L(a)$ and $\mathrm{fm}_L(b)$ and the formula with expression number a is the conclusion of an expansion rule whose premise has expression number b;

(xviii) $\mathrm{assoc}_L(a, b)$ if and only if $\mathrm{fm}_L(a)$ and $\mathrm{fm}_L(b)$ and the formula with expression number a is the conclusion of an associativity rule whose premise has expression number b;

(xix) $\mathrm{cut}_L(a, b, c)$ if and only if $\mathrm{fm}_L(a)$ and $\mathrm{fm}_L(b)$ and $\mathrm{fm}_L(c)$ and the formula with expression number a is the conclusion of a cut rule whose premises, in order, have expression numbers b and c;

(xx) $\mathrm{intr}_L(a, b)$ if and only if $\mathrm{fm}_L(a)$ and $\mathrm{fm}_L(b)$ and the formula with expression number a is the conclusion of an $\exists$-introduction rule whose premise has expression number b;

(xxi) $\mathrm{nlax}_T(a)$ if and only if a is the expression number of a nonlogical axiom of T;

(xxii) $\mathrm{der}_T(a, b)$ if and only if $\mathrm{fm}_{L(T)}(a)$ and b is the expression number of a derivation in T of the formula with expression number a;

(xxiii) $\mathrm{thm}_T(a)$ if and only if a is the expression number of a theorem of T.

3.2 Numerotations. We now describe a natural and highly applicable method to obtain arithmetizations of first-order languages. Let L be a first-order language. We assume given a mapping σ which to each symbol $\mathbf{s}$ of L associates a number $\sigma(\mathbf{s})$, called the *symbol number* of $\mathbf{s}$. We require that different symbols be associated to different numbers. Such a mapping is called a *numerotation* of L. We can then define

(i) $\mathrm{vr}_\sigma(n)$ is the symbol number of the $(n+1)$th variable in the alphabetical order;

(ii) $\mathrm{func}_\sigma(a, n)$ if and only if a is the symbol number of an n-ary function symbol of L;

(iii) $\mathrm{pred}_\sigma(a, n)$ if and only if a is the symbol number of an n-ary predicate symbol of L;

(iv) $\mathrm{sym}_\sigma(a)$ if and only if a is the symbol number of a symbol of L.

We call the numerotation σ *recursive* if vr_σ, func_σ, and pred_σ are recursive and if vr_σ is an increasing function, i.e., $\mathrm{vr}_\sigma(a) < \mathrm{vr}_\sigma(b)$ whenever $a < b$. When this is the case, sym_σ is recursively enumerable.

We now describe a method to obtain arithmetizations from numerotations. Let L be a first-order language, σ a numerotation of L, and β a coding function. We first show how to assign a number to every designator of L. We do this by induction on the length of the designator. If $\mathbf{u}$ is a designator of L, then by the formation theorem $\mathbf{u}$ can be written in one and only one way as $\mathbf{s}\mathbf{u}_1 \dots \mathbf{u}_n$ where $\mathbf{s}$ is a symbol of L of index n and $\mathbf{u}_1, \dots, \mathbf{u}_n$ are designators of L. We then define

$$`\mathbf{u}' = \langle \sigma(\mathbf{s}), `\mathbf{u}_1', \dots, `\mathbf{u}_n' \rangle.$$

If $\mathbf{A}_1, \dots, \mathbf{A}_n$ is a sequence of formulae of L, then the expression number of this sequence is the number $\langle `\mathbf{A}_1', \dots, `\mathbf{A}_n' \rangle$. It is obvious that this defines an arithmetization of L, i.e., that different designators (resp. different sequences of formulae) have different expression numbers.†

We say that a first-order language L is arithmetized *from* the numerotation σ *by* the coding function β if it is endowed with the arithmetization just described.

Theorem. Let L be a first-order language arithmetized from a recursive numerotation σ by a recursive coding function β. Then: the predicates (i)–(xx) of §3.1 are recursive; for each symbol $\mathbf{s}$ of L of index n there is an n-ary recursive function $f_{\mathbf{s}}$ such that for every designator of the form $\mathbf{s}\mathbf{u}_1 \dots \mathbf{u}_n$,

† To define numerotations themselves, one can draw inspiration from this method and the remark concerning variables in ch. I §2.1: thinking of variables as concatenations of two symbols x and $'$, if ς assigns injectively a number to x, $'$, and the symbols of L that are not variables, then, if $\mathbf{x}_n$ is the nth variable in the alphabetical order, the mapping σ defined by $\sigma(\mathbf{x}_1) = \langle \varsigma(x) \rangle$, $\sigma(\mathbf{x}_{n+1}) = \langle \varsigma('), \sigma(\mathbf{x}_n) \rangle$, and $\sigma(\mathbf{s}) = \langle \varsigma(\mathbf{s}) \rangle$ for $\mathbf{s}$ not a variable is a numerotation of L. Moreover, vr_σ is automatically increasing, so σ will be recursive provided that func_ς, pred_ς, and β are recursive.

$f_{\mathbf{s}}(\text{‘}\mathbf{u}_1\text{’}, \ldots, \text{‘}\mathbf{u}_n\text{’}) = \text{‘}\mathbf{s}\mathbf{u}_1 \ldots \mathbf{u}_n\text{’}$; there is a recursive function sub such that for every designator $\mathbf{u}$, distinct variables $\mathbf{x}_1, \ldots, \mathbf{x}_n$, and terms $\mathbf{a}_1, \ldots, \mathbf{a}_n$,

$$\mathrm{sub}(\text{‘}\mathbf{u}\text{’}, \langle \text{‘}\mathbf{x}_1\text{’}, \ldots, \text{‘}\mathbf{x}_n\text{’}\rangle, \langle \text{‘}\mathbf{a}_1\text{’}, \ldots, \text{‘}\mathbf{a}_n\text{’}\rangle) = \text{‘}\mathbf{u}[\mathbf{x}_1, \ldots, \mathbf{x}_n|\mathbf{a}_1, \ldots, \mathbf{a}_n]\text{’};$$

there is a recursive function clos such that for every formula $\mathbf{A}$, if $\mathbf{A}'$ is the closure of $\mathbf{A}$, $\mathrm{clos}(\text{‘}\mathbf{A}\text{’}) = \text{‘}\mathbf{A}'\text{’}$. If moreover L is numerical, then the function num defined by $\mathrm{num}(a) = \text{‘}\dot{a}\text{’}$ is recursive.

If T is a first-order theory with language L and if nlax_T is recursive (resp. recursively enumerable), then der_T is recursive (resp. recursively enumerable) and thm_T is recursively enumerable.

Proof. We define $f_{\mathbf{s}}(a_1, \ldots, a_n) = \langle \sigma(\mathbf{s}), a_1, \ldots, a_n\rangle$. Then $f_{\mathbf{s}}$ is recursive and has the desired property. Define $f(a,b) = \mu i((a)_0 \neq \sigma(\exists) \vee \neg((a)_1 \in b) \vee (i \neq 0 \wedge (a)_1 = \beta(b,i)))$. This is obviously well defined and recursive, and $f(\mathbf{u}, \langle \text{‘}\mathbf{x}_1\text{’}, \ldots, \text{‘}\mathbf{x}_n\text{’}\rangle)$ is the first i such that $\mathbf{x}_i$ is $\mathbf{x}$ if $\mathbf{u}$ is $\exists\mathbf{x}\mathbf{A}$ and such an i exists, and is 0 otherwise. We let $\mathrm{sub}(a,b,c)$ be $\beta(c, \mu i(\neg(a \in b) \vee a = (b)_i) + 1)$ if $\mathrm{vble}_L(a) \wedge a \in b$,

$$\mu d(\mathrm{len}(d) = \mathrm{len}(a) \wedge (d)_0 = (a)_0 \wedge \forall i_{<\mathrm{len}(a)-1}((d)_{i+1} = \mathrm{sub}((a)_{i+1}, \mathrm{rmv}(b, f(a,b)), \mathrm{rmv}(c, f(a,b)))))$$

if $\neg\,\mathrm{vble}_L(a)$, and a otherwise. It is easy to check that $\mathrm{sub}(\text{‘}\mathbf{u}\text{’}, \langle \text{‘}\mathbf{x}_1\text{’}, \ldots, \text{‘}\mathbf{x}_n\text{’}\rangle, \langle \text{‘}\mathbf{a}_1\text{’}, \ldots, \text{‘}\mathbf{a}_n\text{’}\rangle) = \text{‘}\mathbf{u}[\mathbf{x}_1, \ldots, \mathbf{x}_n|\mathbf{a}_1, \ldots, \mathbf{a}_n]\text{’}$. The recursiveness of sub, (i)–(xii), and (xv)–(xx) follows from the relations

(i) $\mathrm{vble}_L(a) \leftrightarrow a = \langle (a)_0\rangle \wedge \exists b_{<a}(\mathrm{vr}_\sigma(b) = (a)_0)$ (here we use that vr_σ is increasing);

(ii) $\mathrm{tm}_L(a) \leftrightarrow \mathrm{vble}_L(a) \vee (\mathrm{sq}(a) \wedge \mathrm{len}(a) \neq 0 \wedge \mathrm{func}_\sigma((a)_0, \mathrm{len}(a)-1) \wedge \forall b_{<\mathrm{len}(a)-1}(\mathrm{tm}_L((a)_{b+1})))$;

(iii) $\mathrm{atfm}_L(a) \leftrightarrow \mathrm{sq}(a) \wedge \mathrm{len}(a) \neq 0 \wedge \mathrm{pred}_\sigma((a)_0, \mathrm{len}(a)-1) \wedge \forall b_{<\mathrm{len}(a)-1}(\mathrm{tm}_L((a)_{b+1}))$;

(iv) $\mathrm{fm}_L(a) \leftrightarrow \mathrm{atfm}_L(a) \vee (a = \langle \sigma(\vee), (a)_1, (a)_2\rangle \wedge \mathrm{fm}_L((a)_1) \wedge \mathrm{fm}_L((a)_2)) \vee (a = \langle \sigma(\neg), (a)_1\rangle \wedge \mathrm{fm}_L((a)_1)) \vee (a = \langle \sigma(\exists), (a)_1, (a)_2\rangle \wedge \mathrm{vble}_L((a)_1) \wedge \mathrm{fm}_L((a)_2))$;

(v) $\mathrm{des}_L(a) \leftrightarrow \mathrm{tm}_L(a) \vee \mathrm{fm}_L(a)$;

(vi) $\mathrm{occ}_L(a,b) \leftrightarrow \mathrm{des}_L(a) \wedge \mathrm{des}_L(b) \wedge (a = b \vee \exists i_{<\mathrm{len}(a)-1}\,\mathrm{occ}_L((a)_{i+1}, b))$;

(vii) $\mathrm{fr}_L(a,b) \leftrightarrow \mathrm{des}_L(a) \wedge \mathrm{vble}_L(b) \wedge \begin{cases} \mathrm{occ}_L(a,b) & \text{if } \mathrm{tm}_L(a) \vee \mathrm{atfm}_L(a), \\ \mathrm{fr}_L((a)_1, b) \vee \mathrm{fr}_L((a)_2, b) & \text{if } a = \langle \sigma(\vee), (a)_1, (a)_2\rangle, \\ \mathrm{fr}_L((a)_1, b) & \text{if } a = \langle \sigma(\neg), (a)_1\rangle, \\ \mathrm{fr}_L((a)_2, b) \wedge (a)_1 \neq b & \text{otherwise;} \end{cases}$

(viii) $\mathrm{cl}_L(a) \leftrightarrow \mathrm{des}_L(a) \wedge \neg\exists b_{<a}\,\mathrm{fr}_L(a,b)$;

(ix) $\mathrm{subtl}_L(a,b,c) \leftrightarrow \mathrm{des}_L(a) \wedge \mathrm{vble}_L(b) \wedge \mathrm{tm}_L(c)$
$\wedge \begin{cases} \mathrm{subtl}_L((a)_1, b, c) \wedge \mathrm{subtl}_L((a)_2, b, c) & \text{if } a = \langle \sigma(\vee), (a)_1, (a)_2\rangle, \\ \mathrm{subtl}_L((a)_1, b) & \text{if } a = \langle \sigma(\neg), (a)_1\rangle, \\ \mathrm{subtl}_L((a)_2, b, c) \wedge (\neg\,\mathrm{fr}_L((a)_2, b) \vee \neg\,\mathrm{fr}_L(c, (a)_1)) & \text{if } a = \langle \sigma(\exists), (a)_1, (a)_2\rangle \wedge (a)_1 \neq b, \\ 0 = 0 & \text{otherwise;} \end{cases}$

(x) $\mathrm{pax}_L(a) \leftrightarrow \mathrm{fm}_L(a) \wedge a = \langle \sigma(\vee), \langle \sigma(\neg), (a)_2\rangle, (a)_2\rangle$;

(xi) $\mathrm{sax}_L(a) \leftrightarrow \mathrm{fm}_L(a) \wedge a = \langle \sigma(\vee), \langle \sigma(\neg), (a)_{1,1}\rangle, \langle \sigma(\exists), (a)_{2,1}, (a)_{2,2}\rangle\rangle$
$\wedge \exists b_{<a}(\mathrm{tm}_L(b) \wedge \mathrm{subtl}_L((a)_{2,2}, (a)_{2,1}, b) \wedge (a)_{1,1} = \mathrm{sub}((a)_{2,2}, (a)_{2,1}, b))$;

(xii) $\mathrm{iax}_L(a) \leftrightarrow a = \langle \sigma(=), (a)_1, (a)_1\rangle \wedge \mathrm{vble}_L((a)_1)$;

(xv) $\mathrm{ax}_L(a) \leftrightarrow \mathrm{pax}_L(a) \vee \mathrm{sax}_L(a) \vee \mathrm{iax}_L(a) \vee \mathrm{feax}_L(a) \vee \mathrm{peax}_L(a)$;

(xvi) $\mathrm{ctr}_L(a,b) \leftrightarrow \mathrm{fm}_L(a) \wedge b = \langle \sigma(\vee), a, a\rangle$;

(xvii) $\mathrm{exp}_L(a,b) \leftrightarrow \mathrm{fm}_L(a) \wedge a = \langle \sigma(\vee), (a)_1, b\rangle$;

(xviii) $\mathrm{assoc}_L(a,b) \leftrightarrow \mathrm{fm}_L(a) \wedge a = \langle \sigma(\vee), \langle \sigma(\vee), (a)_{1,1}, (a)_{1,2}\rangle, (a)_2\rangle$
$\wedge\, b = \langle \sigma(\vee), (a)_{1,1}, \langle \sigma(\vee), (a)_{1,2}, (a)_2\rangle\rangle$;

(xix) $\mathrm{cut}_L(a,b,c) \leftrightarrow \mathrm{fm}_L(b) \wedge \mathrm{fm}_L(c) \wedge b = \langle \sigma(\vee), (b)_1, (b)_2\rangle \wedge c = \langle \sigma(\vee), \langle \sigma(\neg), (b)_1\rangle, (c)_2\rangle$
$\wedge\, a = \langle \sigma(\vee), (b)_2, (c)_2\rangle$;

(xx) $\mathrm{intr}_L(a,b) \leftrightarrow \mathrm{fm}_L(a) \wedge a = \langle \sigma(\vee), \langle \sigma(\neg), \langle \sigma(\exists), (a)_{1,1,1}, (a)_{1,1,2}\rangle\rangle, (a)_2\rangle$
$\wedge\, b = \langle \sigma(\vee), \langle \sigma(\neg), (a)_{1,1,2}\rangle, (a)_2\rangle \wedge \neg\,\mathrm{fr}_L((a)_2, (a)_{1,1,1})$.

The predicates feax_L and peax_L require some more work. Let f be the binary function defined by $f(a, n) = (a)_{2,\ldots,2}$ with n occurrences of 2. Then f is recursive because

$$f(a, n) = \begin{cases} a & \text{if } n = 0, \\ f(a, n-1)_2 & \text{otherwise.} \end{cases}$$

Now if a is the expression number of an equality axiom for, say, an n-ary function symbol $\mathbf{f}$ and if $0 \le i \le n$, then $f(a, i)$ is the expression number of

$$\mathbf{x}_{i+1} = \mathbf{y}_{i+1} \to \cdots \to \mathbf{x}_n = \mathbf{y}_n \to \mathbf{f}\mathbf{x}_1 \ldots \mathbf{x}_n = \mathbf{f}\mathbf{y}_1 \ldots \mathbf{y}_n.$$

We thus have the relations

(xiii) $\mathrm{feax}_L(a) \leftrightarrow \exists n_{<a}(f(a,n) = \langle \sigma(=), f(a,n)_1, f(a,n)_2\rangle$
$\wedge \operatorname{len}(f(a,n)_1) = n+1 \wedge \operatorname{len}(f(a,n)_2) = n+1$
$\wedge \operatorname{func}_\sigma(f(a,n)_{1,0}, n) \wedge f(a,n)_{1,0} = f(a,n)_{2,0}$
$\wedge \forall i_{<n}(\mathrm{vble}_L(f(a,n)_{1,i+1}) \wedge \mathrm{vble}_L(f(a,n)_{2,i+1}))$
$\wedge \forall i_{<n}(f(a,i) = \langle \sigma(\vee), \langle \sigma(\neg), \langle \sigma(=), f(a,i)_{1,1,1}, f(a,i)_{1,1,2}\rangle\rangle, f(a,i+1)\rangle$
$\wedge f(a,i)_{1,1,1} = f(a,n)_{1,i+1} \wedge f(a,i)_{1,1,2} = f(a,n)_{2,i+1}))$ and

(xiv) $\mathrm{peax}_L(a) \leftrightarrow \exists n_{<a}(f(a,n) = \langle \sigma(\vee), \langle \sigma(\neg), f(a,n)_{1,1}\rangle, f(a,n)_2\rangle$
$\wedge \operatorname{len}(f(a,n)_{1,1}) = n+1 \wedge \operatorname{len}(f(a,n)_2) = n+1$
$\wedge \operatorname{pred}_\sigma(f(a,n)_{1,1,0}, n) \wedge f(a,n)_{1,1,0} = f(a,n)_{2,0}$
$\wedge \forall i_{<n}(\mathrm{vble}_L(f(a,n)_{1,1,i+1}) \wedge \mathrm{vble}_L(f(a,n)_{2,i+1}))$
$\wedge \forall i_{<n}(f(a,i) = \langle \sigma(\vee), \langle \sigma(\neg), \langle \sigma(=), f(a,i)_{1,1,1}, f(a,i)_{1,1,2}\rangle\rangle, f(a,i+1)\rangle$
$\wedge f(a,i)_{1,1,1} = f(a,n)_{1,1,i+1} \wedge f(a,i)_{1,1,2} = f(a,n)_{2,i+1}))$

showing that feax_L and peax_L are recursive.

We define a unary function g by

$$g(a) = \begin{cases} \langle \sigma(\neg), \langle \sigma(\exists), \mu b(\mathrm{fr}_L(a,b)), \langle \sigma(\neg), a\rangle\rangle\rangle & \text{if } \neg\,\mathrm{cl}_L(a), \\ a & \text{otherwise;} \end{cases}$$

Then g is recursive and $g(\text{‘}\mathbf{A}\text{’})$ is $\mathbf{A}$ if $\mathbf{A}$ is closed and $\forall \mathbf{x}\mathbf{A}$ otherwise, where $\mathbf{x}$ is the variable free in $\mathbf{A}$ which comes first in the alphabetical ordering of all the variables free in $\mathbf{A}$. We observe that there cannot be more than ‘$\mathbf{A}$’ variables free in $\mathbf{A}$ because each of them has a different nonzero expression number. We then define

$$h(a, n) = \begin{cases} a & \text{if } n = 0, \\ h(g(a), n-1) & \text{otherwise,} \end{cases}$$

and finally $\mathrm{clos}(a) = h(a, a)$. Then clos is recursive and $\mathrm{clos}(\text{‘}\mathbf{A}\text{’})$ is the expression number of the closure of $\mathbf{A}$, as desired.

If L is numerical, we have

$$\mathrm{num}(a) = \begin{cases} \langle \sigma(\dot{0})\rangle & \text{if } a = 0, \\ \langle \sigma(\mathrm{S}), \mathrm{num}(a-1)\rangle & \text{otherwise,} \end{cases}$$

and hence num is recursive.

Suppose that T is a first-order theory with language L. We observe that

(xxii) $\mathrm{der}_T(a,b) \leftrightarrow \mathrm{sq}(b) \wedge \operatorname{len}(b) \ne 0 \wedge a = \beta(b, \operatorname{len}(b)) \wedge \forall c_{<\operatorname{len}(b)}(\mathrm{ax}_L(c) \vee \mathrm{nlax}_T(c)$
$\vee \exists d_{<c}(\mathrm{ctr}_L((b)_c, (b)_d) \vee \exp_L((b)_c, (b)_d) \vee \mathrm{assoc}_L((b)_c, (b)_d) \vee \exists e_{<c}\,\mathrm{cut}((b)_c, (b)_d, (b)_e)$
$\vee \mathrm{intr}_L((b)_c, (b)_d)))$; and

(xxiii) $\mathrm{thm}_T(a) \leftrightarrow \exists b\,\mathrm{der}_T(a,b)$.

From these relations the remaining assertions of the theorem follow. □

There are yet more syntactical constructions which become recursive functions under suitable numerotations. For example, if L and L' are arithmetized from recursive numerotations σ and σ' by a recursive coding function, if I is an interpretation of L in L', and if there is a recursive function f such that for all variables, function symbols, and predicate symbols $\mathbf{s}$ of L, $f(\sigma(\mathbf{s})) = \sigma'(\mathbf{s}_I)$, then there is a recursive function which associates to the expression number of a designator of L the expression number of its interpretation by I.

§4 The incompleteness theorem

4.1 The diagonal lemma. Let L be an arithmetized numerical first-order language. A unary function f is a *diagonal function for L* if for some variable $\mathbf{x}_f$, $f(‘\mathbf{A}’) = ‘\mathbf{A}[\mathbf{x}_f|\ulcorner\mathbf{A}\urcorner]’$ for all $\mathbf{A}$. An arithmetized numerical first-order theory T will be called *diagonalizable* if some diagonal function for $L(T)$ is representable in T.

A unary function h is a *negation function for T* if for any formula $\mathbf{A}$ of T, if $\mathbf{A}'$ is the closure of $\mathbf{A}$, then $h(‘\mathbf{A}’)$ is the expression number of a formula $\mathbf{B}$ of T such that $\vdash_T \mathbf{B} \leftrightarrow \neg\mathbf{A}'$. When a negation function for T has been chosen, we define a unary predicate thm'_T by $\mathrm{thm}'_T(a) \leftrightarrow \exists b(\mathrm{der}_T(a,b) \wedge \forall c_{<b} \neg\, \mathrm{der}_T(h(a),c))$. Note that $\mathrm{thm}'_T(a) \to \mathrm{thm}_T(a)$.

> DIAGONAL LEMMA. Let T be an arithmetized numerical first-order theory, let f be a diagonal function for $L(T)$, and let h be a negation function for T. If the predicate p defined by $p(a) \leftrightarrow \neg\, \mathrm{thm}'_T(f(a))$ is representable in T, then T is inconsistent.

Proof. There is a formula $\mathbf{A}$ of T with expression number a such that $\mathbf{A}$ with $\mathbf{x}_f$ represents p in T. By definition of p, $\neg p(a)$ implies $\vdash_T \mathbf{A}[\mathbf{x}_f|\dot{a}]$. But if $p(a)$, then $\vdash_T \mathbf{A}[\mathbf{x}_f|\dot{a}]$ because $\mathbf{A}$ with $\mathbf{x}_f$ represents p in T, so in all cases $\vdash_T \mathbf{A}[\mathbf{x}_f|\dot{a}]$. Choose b such that $\mathrm{der}_T(f(a),b)$. If $\mathrm{der}_T(h(f(a)),c)$ for some $c < b$, then T is inconsistent by the closure theorem and the tautology theorem. Otherwise, $\neg p(a)$, and since $\mathbf{A}$ with $\mathbf{x}_f$ represents p in T, $\vdash_T \neg\mathbf{A}[\mathbf{x}_f|\dot{a}]$. By the tautology theorem, T is inconsistent in this case as well. □

Remark. The proof of the diagonal lemma is the only potentially nonconstructive proof on these pages. Indeed, we might not be able to decide whether $p(a)$ or $\neg p(a)$, in which case the proof does not actually produce a derivation of $\mathbf{A}[\mathbf{x}_f|\dot{a}]$ in T, but merely establishes that the nonexistence of such a derivation leads to a contradiction. However, in our only application of the diagonal lemma in §4.2, p will be recursive and, in particular, decidable.

4.2 Church's theorem and incompleteness. We give a first proof of the incompleteness theorem based on the diagonal lemma. We let T be an arithmetized numerical first-order theory. We must assume that the arithmetization of T is sufficiently well-behaved. Precisely, we shall require that

(i) the predicates $\mathrm{fm}_{L(T)}$ and der_T are recursive;
(ii) there is a recursive diagonal function f for $L(T)$;
(iii) there is a recursive negation function h for T.

These assumptions are met for example if the arithmetization of $L(T)$ comes from a recursive numerotation by a recursive coding function and if nlax_T is recursive. For then, with the notations of the theorem of §3.2, the function f defined by $f(a) = \mathrm{sub}(a, \langle ‘x’\rangle, \langle \mathrm{num}(a)\rangle)$ satisfies (ii), and the function h defined by $h(a) = f_\neg(\mathrm{clos}(a))$ satisfies (iii). Note that (i) and (iii) imply that thm'_T is recursively enumerable.

> CHURCH'S THEOREM. Suppose that every recursive unary predicate is representable in T. If thm'_T is recursive, then T is inconsistent.

Proof. Since thm'_T is recursive, the predicate p defined by $p(a) \leftrightarrow \neg\, \mathrm{thm}'_T(f(a))$ is recursive, and hence representable in T. By the diagonal lemma, T is inconsistent. □

> LEMMA. If T is complete, then thm'_T is recursive.

Proof. Indeed, we have $\neg\, \mathrm{thm}'_T(a) \leftrightarrow \neg\, \mathrm{fm}_{L(T)}(a) \vee \exists b(\mathrm{der}_T(h(a),b) \wedge \forall c_{<b}(\neg\, \mathrm{der}_T(a,c)))$, and so $\neg\, \mathrm{thm}'_T$ is recursively enumerable. By the negation lemma, thm'_T is recursive. □

From Church's theorem and the lemma, we obtain at once

> INCOMPLETENESS THEOREM 1. Suppose that every recursive unary predicate is representable in T. If T is complete, then it is inconsistent.

Note that the diagonal lemma, Church's theorem, and the above lemma can all be proved with thm_T in place of thm'_T. However, the proof of the incompleteness theorem obtained in this way is not constructive in most cases of interest, since the proof of the lemma would depend on the consistency of T.

4.3 The fixed point theorem. The above proof of the incompleteness theorem is constructive, but it is somewhat unsatisfying: if a first-order theory T satisfying the various hypotheses *is* consistent, we are unable to use it to find a closed formula $\mathbf{A}$ of T such that neither $\mathbf{A}$ nor $\neg\mathbf{A}$ are theorems of T. In the remaining of this section, we shall give a proof of the incompleteness theorem (with different hypotheses on T) by actually producing such a formula.

We let T be an arithmetized numerical first-order theory.

Fixed Point Theorem. Suppose that some diagonal function f for $L(T)$ is representable in T. For any formula $\mathbf{A}$ of T, we can find a formula $\mathbf{B}$ of T, whose free variables are those of $\mathbf{A}$ without $\mathbf{x}_f$, such that $\vdash_T \mathbf{B} \leftrightarrow \mathbf{A}[\mathbf{x}_f|\ulcorner\mathbf{B}\urcorner]$.

Proof. Let $\mathbf{x}$ be distinct from $\mathbf{x}_f$ and not occurring in $\mathbf{A}$. There is a formula $\mathbf{D}$ of T such that $\mathbf{D}$ with $\mathbf{x}_f, \mathbf{x}$ represents f in T and in which no variable other than $\mathbf{x}_f$ and $\mathbf{x}$ is free. Let a be the expression number of $\exists\mathbf{x}(\mathbf{D} \wedge \mathbf{A}[\mathbf{x}_f|\mathbf{x}])$, and let $\mathbf{B}$ be the formula $\exists\mathbf{x}(\mathbf{D}[\mathbf{x}_f|\dot{a}] \wedge \mathbf{A}[\mathbf{x}_f|\mathbf{x}])$. If b is the expression number of $\mathbf{B}$, then $b = f(a)$. Since $\mathbf{D}$ with $\mathbf{x}_f, \mathbf{x}$ represents f, we have

$$\vdash_T \mathbf{D}[\mathbf{x}_f|\dot{a}] \leftrightarrow \mathbf{x} = \dot{b}, \tag{1}$$

whence $\vdash_T \mathbf{D}[\mathbf{x}_f, \mathbf{x}|\dot{a}, \dot{b}]$ by the equality theorem. The formula $\mathbf{D}[\mathbf{x}_f, \mathbf{x}|\dot{a}, \dot{b}] \wedge \mathbf{A}[\mathbf{x}_f|\dot{b}] \to \mathbf{B}$ is a substitution axiom, so

$$\vdash_T \mathbf{A}[\mathbf{x}_f|\dot{b}] \to \mathbf{B} \tag{2}$$

by the tautology theorem. By the equality theorem, $\vdash_T \mathbf{x} = \dot{b} \to \mathbf{A}[\mathbf{x}_f|\mathbf{x}] \leftrightarrow \mathbf{A}[\mathbf{x}_f|\dot{b}]$. A tautological consequence of this and (1) is $\vdash_T \mathbf{D}[\mathbf{x}_f|\dot{a}] \wedge \mathbf{A}[\mathbf{x}_f|\mathbf{x}] \to \mathbf{A}[\mathbf{x}_f|\dot{b}]$. Hence

$$\vdash_T \mathbf{B} \to \mathbf{A}[\mathbf{x}_f|\dot{b}] \tag{3}$$

by the $\exists$-introduction rule. From (2) and (3), we obtain $\vdash_T \mathbf{B} \leftrightarrow \mathbf{A}[\mathbf{x}_f|\dot{b}]$ by the tautology theorem, which is the desired result. □

As an example, we prove an interesting corollary to the fixed point theorem. A *truth definition for T* is a formula $\mathbf{A}$ of T in which no variable other than x is free and such that for every closed formula $\mathbf{B}$ of T, $\vdash_T \mathbf{B} \leftrightarrow \mathbf{A}[x|\ulcorner\mathbf{B}\urcorner]$.

Theorem on truth definitions. Suppose that T is diagonalizable. If there exists a truth definition for T, then T is inconsistent.

Proof. Let $\mathbf{A}$ be a truth definition for T and let f be a diagonal function for $L(T)$ which is representable in T. By the fixed point theorem, we can find a closed formula $\mathbf{B}$ of T such that $\vdash_T \mathbf{B} \leftrightarrow \neg\mathbf{A}'[x|\mathbf{x}_f][\mathbf{x}_f|\ulcorner\mathbf{B}\urcorner]$, where $\mathbf{A}'$ is a variant of $\mathbf{A}$ in which $\mathbf{x}_f$ is substitutible for x. By the variant theorem, $\vdash_T \mathbf{B} \leftrightarrow \neg\mathbf{A}[x|\ulcorner\mathbf{B}\urcorner]$. But we also have $\vdash_T \mathbf{B} \leftrightarrow \mathbf{A}[x|\ulcorner\mathbf{B}\urcorner]$, so T is inconsistent by the tautology theorem. □

4.4 Rosser's formula and incompleteness. Let T be an arithmetized numerical first-order theory having the binary predicate symbol $<$ about which we make the following assumptions:

(i) T is diagonalizable;

(ii) some function $f_\neg$ such that $f_\neg(`\mathbf{A}') = `\neg\mathbf{A}'$ for all $\mathbf{A}$ is representable in T;

(iii) der_T is representable in T;

and for some formula $\mathbf{U}$ of T in which only x is free,

(iv) for all n, $\vdash_T \mathbf{U}[x|\dot{n}]$;

(v) for all n, $\vdash_T \mathbf{U} \to x < \mathrm{S}\dot{n} \to x = \dot{0} \vee \cdots \vee x = \dot{n}$;

(vi) for all n, $\vdash_T \mathbf{U} \to x < \mathrm{S}\dot{n} \vee \dot{n} < \mathrm{S}x$.

Intuitively, we intend the formula $\mathbf{U}$ to mean "x is a natural number".

Let f be a diagonal function for $L(T)$ which is representable in T, and let $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$, and $\mathbf{w}$ be distinct variables, distinct from $\mathbf{x}_f$ and not occurring in $\mathbf{U}$. By assumptions (ii) and (iii), there are formulae $\mathbf{C}$ and $\mathbf{D}$ of T such that $\mathbf{C}$ with $\mathbf{x}$, $\mathbf{y}$ represents $f_\neg$ in T and such that $\mathbf{D}$ with $\mathbf{y}$, $\mathbf{z}$ represents der_T in T. We may suppose that no variable other than $\mathbf{x}$ and $\mathbf{y}$ is free in $\mathbf{C}$, and no variable other than $\mathbf{y}$ and $\mathbf{z}$ is free in $\mathbf{D}$.

Taking variants if necessary, we may assume that $\mathbf{x}_f$ is substitutible for $\mathbf{x}$ in $\mathbf{C}$ and for $\mathbf{y}$ in $\mathbf{D}$, and that $\mathbf{w}$ is substitutible for $\mathbf{z}$ in $\mathbf{D}$. Applying the fixed point theorem to the formula

$$\forall\mathbf{y}\forall\mathbf{z}(\mathbf{U}[x|\mathbf{z}] \to \mathbf{C}[\mathbf{x}|\mathbf{x}_f] \to \mathbf{D}[\mathbf{y}|\mathbf{x}_f] \to \exists\mathbf{w}(\mathbf{U}[x|\mathbf{w}] \wedge \mathbf{w} < \mathrm{S}\mathbf{z} \wedge \mathbf{D}[\mathbf{z}|\mathbf{w}])),$$

we find a closed formula $\mathbf{R}$ of T such that

$$\vdash_T \mathbf{R} \leftrightarrow \forall\mathbf{y}\forall\mathbf{z}(\mathbf{U}[x|\mathbf{z}] \to \mathbf{C}[\mathbf{x}|\ulcorner\mathbf{R}\urcorner] \to \mathbf{D}[\mathbf{y}|\ulcorner\mathbf{R}\urcorner] \to \exists\mathbf{w}(\mathbf{U}[x|\mathbf{w}] \wedge \mathbf{w} < \mathrm{S}\mathbf{z} \wedge \mathbf{D}[\mathbf{z}|\mathbf{w}])). \tag{4}$$

Incompleteness Theorem 2. With the hypotheses of this paragraph, if either $\vdash_T \mathbf{R}$ or $\vdash_T \neg\mathbf{R}$, then T is inconsistent.

Proof. Assume first that $\vdash_T \mathbf{R}$ and let n be the expression number of a derivation of $\mathbf{R}$ in T, i.e., such that $\mathrm{der}_T(\text{‘}\mathbf{R}\text{’}, n)$. By (4), the tautology theorem, and the substitution theorem, we find $\vdash_T \mathbf{U}[x|\dot{n}] \to \mathbf{C}[\mathbf{x},\mathbf{y}|\ulcorner\mathbf{R}\urcorner,\ulcorner\neg\mathbf{R}\urcorner] \to \mathbf{D}[\mathbf{y},\mathbf{z}|\ulcorner\mathbf{R}\urcorner,\dot{n}] \to \exists\mathbf{w}(\mathbf{U}[x|\mathbf{w}] \wedge \mathbf{w} < \mathrm{S}\dot{n} \wedge \mathbf{D}[\mathbf{y},\mathbf{z}|\ulcorner\neg\mathbf{R}\urcorner,\mathbf{w}])$. But $\vdash_T \mathbf{U}[x|\dot{n}]$ by (iv), and $\vdash_T \mathbf{C}[\mathbf{x},\mathbf{y}|\ulcorner\mathbf{R}\urcorner,\ulcorner\neg\mathbf{R}\urcorner]$ and $\vdash_T \mathbf{D}[\mathbf{y},\mathbf{z}|\ulcorner\mathbf{R}\urcorner,\dot{n}]$ by representability, so $\vdash_T \exists\mathbf{w}(\mathbf{U}[x|\mathbf{w}]\wedge\mathbf{w} < \mathrm{S}\dot{n}\wedge\mathbf{D}[\mathbf{y},\mathbf{z}|\ulcorner\neg\mathbf{R}\urcorner,\mathbf{w}])$ by the detachment rule. By an instance of (v), the tautology theorem, and the distribution rule, $\vdash_T \exists\mathbf{w}((\mathbf{w} = \dot{0} \wedge \mathbf{D}[\mathbf{y},\mathbf{z}|\ulcorner\neg\mathbf{R}\urcorner,\mathbf{w}]) \vee \cdots \vee (\mathbf{w} = \dot{n} \wedge \mathbf{D}[\mathbf{y},\mathbf{z}|\ulcorner\neg\mathbf{R}\urcorner,\mathbf{w}]))$. Hence by $\exists$–$\vee$ distributivity and the replacement theorem,

$$\vdash_T \mathbf{D}[\mathbf{y},\mathbf{z}|\ulcorner\neg\mathbf{R}\urcorner,\dot{0}] \vee \cdots \vee \mathbf{D}[\mathbf{y},\mathbf{z}|\ulcorner\neg\mathbf{R}\urcorner,\dot{n}]. \tag{5}$$

We now consider two possibilities. If $\mathrm{der}_T(\text{‘}\neg\mathbf{R}\text{’}, k)$ for some k with $0 \le k \le n$, then $\vdash_T \neg\mathbf{R}$, in which case T is inconsistent. Otherwise, $\neg\,\mathrm{der}_T(\text{‘}\neg\mathbf{R}\text{’}, k)$ for all k with $0 \le k \le n$. By representability, $\vdash_T \neg\mathbf{D}[\mathbf{y},\mathbf{z}|\ulcorner\neg\mathbf{R}\urcorner,\dot{k}]$ for all k with $0 \le k \le n$. Together with (5), we conclude that T is inconsistent in this case as well.

Assume now that $\vdash_T \neg\mathbf{R}$ and let n be the expression number of a derivation of $\neg\mathbf{R}$ in T. By representability, $\vdash_T \mathbf{D}[\mathbf{y},\mathbf{z}|\ulcorner\neg\mathbf{R}\urcorner,\dot{n}]$, whence $\vdash_T \dot{n} < \mathrm{S}\mathbf{z} \to \exists\mathbf{w}(\mathbf{U}[x|\mathbf{w}] \wedge \mathbf{w} < \mathrm{S}\mathbf{z} \wedge \mathbf{D}[\mathbf{y},\mathbf{z}|\ulcorner\neg\mathbf{R}\urcorner,\mathbf{w}])$ by (iv), the tautology theorem, and the substitution axioms. From this by the equality theorem, $\vdash_T \dot{n} < \mathrm{S}\mathbf{z} \to \mathbf{y} = \ulcorner\neg\mathbf{R}\urcorner \to \exists\mathbf{w}(\mathbf{U}[x|\mathbf{w}] \wedge \mathbf{w} < \mathrm{S}\mathbf{z} \wedge \mathbf{D}[\mathbf{z}|\mathbf{w}])$ whence

$$\vdash_T \dot{n} < \mathrm{S}\mathbf{z} \to \mathbf{C}[\mathbf{x}|\ulcorner\mathbf{R}\urcorner] \to \exists\mathbf{w}(\mathbf{U}[x|\mathbf{w}] \wedge \mathbf{w} < \mathrm{S}\mathbf{z} \wedge \mathbf{D}[\mathbf{z}|\mathbf{w}]) \tag{6}$$

by representability and the equivalence theorem. As before, we consider two cases. If $\mathrm{der}_T(\text{‘}\mathbf{R}\text{’}, k)$ for some k with $0 \le k \le n$, then $\vdash_T \mathbf{R}$ and T is inconsistent. Otherwise, $\neg\,\mathrm{der}_T(\text{‘}\mathbf{R}\text{’}, k)$ for all k with $0 \le k \le n$. By representability and the equality theorem, $\vdash_T \mathbf{z} = \dot{k} \to \neg\mathbf{D}[\mathbf{y}|\ulcorner\mathbf{R}\urcorner]$ for all k with $0 \le k \le n$. Using an instance of (v) and the tautology theorem, we obtain

$$\vdash_T \mathbf{U}[x|\mathbf{z}] \to \mathbf{z} < \mathrm{S}\dot{n} \to \neg\mathbf{D}[\mathbf{y}|\ulcorner\mathbf{R}\urcorner]. \tag{7}$$

From (6), (7), and (vi) by the tautology theorem, we find $\vdash_T \mathbf{U}[x|\mathbf{z}] \to \mathbf{C}[\mathbf{x}|\ulcorner\mathbf{R}\urcorner] \to \mathbf{D}[\mathbf{y}|\ulcorner\mathbf{R}\urcorner] \to \exists\mathbf{w}(\mathbf{U}[x|\mathbf{w}] \wedge \mathbf{w} < \mathrm{S}\mathbf{z} \wedge \mathbf{D}[\mathbf{z}|\mathbf{w}])$, whence $\vdash_T \mathbf{R}$ by the generalization rule and (4). So T is inconsistent. ☐

Remark. This theorem does *not* imply that $T[\mathbf{R}]$ and $T[\neg\mathbf{R}]$ are inconsistent. In fact, in view of Proposition 1 of ch. II §1.1, it implies that if either $T[\mathbf{R}]$ or $T[\neg\mathbf{R}]$ is inconsistent, then T itself is inconsistent.

We delay to the next section the introduction of a large class of arithmetized first-order theories which satisfy the hypotheses of either version of the incompleteness theorem.

§5 Minimal arithmetic

5.1 Minimal arithmetic. We introduce a first-order theory called *minimal arithmetic* and denoted by N. The nonlogical symbols of $L(\mathrm{N})$ are the constant $\dot{0}$, the unary function symbol S, the binary function symbols $+$ and $\cdot$, and the binary predicate symbol $<$. In particular, $L(\mathrm{N})$ is numerical. We abbreviate the terms $+\mathbf{ab}$ and $\cdot\mathbf{ab}$ by $(\mathbf{a} + \mathbf{b})$ and $(\mathbf{a} \cdot \mathbf{b})$, respectively. As usual, we shall drop the parentheses whenever possible. The nonlogical axioms of N are

N1 $\mathrm{S}x \ne \dot{0}$;

N2 $\mathrm{S}x = \mathrm{S}y \to x = y$;

N3 $x + \dot{0} = x$;

N4 $x + \mathrm{S}y = \mathrm{S}(x + y)$;

N5 $x \cdot \dot{0} = \dot{0}$;
N6 $x \cdot \mathrm{S}y = (x \cdot y) + x$;
N7 $\neg(x < \dot{0})$;
N8 $x < \mathrm{S}y \leftrightarrow x < y \vee x = y$;
N9 $x < y \vee x = y \vee y < x$.

5.2 Properties of N. We intend to prove that N satisfies the hypotheses of (both versions of) the incompleteness theorem. We begin by investigating some basic properties of N.

(i) If $m \neq n$, then $\vdash_{\mathrm{N}} \dot{m} \neq \dot{n}$;
(ii) if $m + n = p$, then $\vdash_{\mathrm{N}} \dot{m} + \dot{n} = \dot{p}$;
(iii) if $mn = p$, then $\vdash_{\mathrm{N}} \dot{m} \cdot \dot{n} = \dot{p}$;
(iv) if $m < n$, then $\vdash_{\mathrm{N}} \dot{m} < \dot{n}$;
(v) if $m \geq n$, then $\vdash_{\mathrm{N}} \neg(\dot{m} < \dot{n})$;
(vi) $\vdash_{\mathrm{N}} x < \mathrm{S}y \vee y < \mathrm{S}x$;
(vii) $\vdash_{\mathrm{N}} x < \mathrm{S}\dot{n} \leftrightarrow x = \dot{0} \vee \cdots \vee x = \dot{n}$;
(viii) $\vdash_{\mathrm{N}} \exists \mathbf{x}(\mathbf{x} < \mathrm{S}\dot{n} \wedge \mathbf{A}) \leftrightarrow \mathbf{A}[\mathbf{x}|\dot{0}] \vee \cdots \vee \mathbf{A}[\mathbf{x}|\dot{n}]$;
(ix) $\vdash_{\mathrm{N}} \forall \mathbf{x}(\mathbf{x} < \mathrm{S}\dot{n} \to \mathbf{A}) \leftrightarrow \mathbf{A}[\mathbf{x}|\dot{0}] \wedge \cdots \wedge \mathbf{A}[\mathbf{x}|\dot{n}]$;
(x) if $\mathbf{a}_0$ is x and $\mathbf{a}_{n+1}$ is $\mathrm{S}\mathbf{a}_n$, then $\vdash_{\mathrm{N}} x + \dot{n} = \mathbf{a}_n$;
(xi) if $\mathbf{b}_0$ is $\dot{0}$ and $\mathbf{b}_{n+1}$ is $\mathbf{b}_n + x$, then $\vdash_{\mathrm{N}} x \cdot \dot{n} = \mathbf{b}_n$.

To prove (i) we may assume $n < m$ by the symmetry theorem. Using $k = m - n$ instances of N2 and the tautology theorem, $\vdash_{\mathrm{N}} \dot{m} = \dot{n} \to \dot{k} = \dot{0}$. But $\vdash_{\mathrm{N}} \dot{k} \neq \dot{0}$ by N1 and the substitution rule, so $\vdash_{\mathrm{N}} \dot{m} \neq \dot{n}$ by the tautology theorem.

We prove (ii) by induction on n. If $n = 0$, then $p = m$ and $\dot{m} + \dot{0} = \dot{m}$ is an instance of N3. Suppose that $\vdash_{\mathrm{N}} \dot{m} + (n \dot{-} 1) = p \dot{-} 1$. By the equality axioms, $\vdash_{\mathrm{N}} \mathrm{S}(\dot{m} + (n \dot{-} 1)) = \dot{p}$, whence $\vdash_{\mathrm{N}} \dot{m} + \dot{n} = \dot{p}$ by an instance of N4 and the equality theorem. The proof of (iii) is also by induction on n. If $n = 0$, then $p = 0$ and $\dot{m} \cdot \dot{0} = \dot{0}$ is an instance of N5. Suppose that $\vdash_{\mathrm{N}} \dot{m} \cdot (n \dot{-} 1) = p \dot{-} m$. By N6, $\vdash_{\mathrm{N}} \dot{m} \cdot \dot{n} = (\dot{m} \cdot (n \dot{-} 1)) + \dot{m}$, whence $\vdash_{\mathrm{N}} \dot{m} \cdot \dot{n} = (p \dot{-} m) + \dot{m}$ by the equality theorem. By (ii), $\vdash_{\mathrm{N}} (p \dot{-} m) + \dot{m} = \dot{p}$, so $\vdash_{\mathrm{N}} \dot{m} \cdot \dot{n} = \dot{p}$ by the equality theorem.

We prove (iv) by induction on n. Suppose that $m < n + 1$. If $m < n$ then $\vdash_{\mathrm{N}} \dot{m} < \dot{n}$ by induction hypothesis, whence $\vdash_{\mathrm{N}} \dot{m} < \mathrm{S}\dot{n}$ by an instance N8 and the tautology theorem. Otherwise $m = n$ and (iv) is a tautological consequence of $\vdash_{\mathrm{N}} \dot{n} = \dot{n}$ and an instance of N8.

If $n = 0$, (v) is an instance of N7. Suppose that $n \geq 1$ and that $m \geq n+1$. Then $m \geq n$ and so $\vdash_{\mathrm{N}} \neg(\dot{m} < \dot{n})$ by induction hypothesis. But also $m \neq n$ and hence $\vdash_{\mathrm{N}} \dot{m} \neq \dot{n}$ by (i). From these by an instance of N8 and the tautology theorem, $\vdash_{\mathrm{N}} \neg(\dot{m} < \mathrm{S}\dot{n})$.

The formula (vi) is a tautological consequence of N8 and N9. The implication from right to left in (vii) follows from (iv) with the tautology theorem. We prove the other implication by induction on n. If $n = 0$, the result is a tautological consequence of N7 and an instance of N8. Suppose that $n \geq 1$. By N8, $\vdash_{\mathrm{N}} x < \mathrm{S}\dot{n} \to x < \dot{n} \vee x = \dot{n}$ and so (vii) follows from the induction hypothesis and the tautology theorem.

By an instance of (vii), the tautology theorem, and the equivalence theorem, $\vdash_{\mathrm{N}} \exists \mathbf{x}(\mathbf{x} < \mathrm{S}\dot{n} \wedge \mathbf{A}) \leftrightarrow \exists \mathbf{x}((\mathbf{x} = \dot{0} \wedge \mathbf{A}) \vee \cdots \vee (\mathbf{x} = \dot{n} \wedge \mathbf{A}))$, whence $\vdash_{\mathrm{N}} \exists \mathbf{x}(\mathbf{x} < \mathrm{S}\dot{n} \wedge \mathbf{A}) \leftrightarrow \exists \mathbf{x}(\mathbf{x} = \dot{0} \wedge \mathbf{A}) \vee \cdots \vee \exists \mathbf{x}(\mathbf{x} = \dot{n} \wedge \mathbf{A})$ by $\exists$–$\vee$ distributivity. From this by the replacement theorem and the equivalence theorem, we obtain (viii). The proof of (ix) is identical using $\forall$–$\wedge$ distributivity and the other statement of the replacement theorem.

The last two assertions are proved by induction on n, being axioms of N if $n = 0$. By N4 and the substitution rule, $\vdash_{\mathrm{N}} x + \mathrm{S}\dot{n} = \mathrm{S}(x + \dot{n})$, so by induction hypothesis and the tautology theorem, $\vdash_{\mathrm{N}} x + \mathrm{S}\dot{n} = \mathrm{S}\mathbf{a}_n$, that is, $\vdash_{\mathrm{N}} x + \mathrm{S}\dot{n} = \mathbf{a}_{n+1}$. Similary, by N6 and the substitution rule, $\vdash_{\mathrm{N}} x \cdot \mathrm{S}\dot{n} = (x \cdot \dot{n}) + x$, so by induction hypothesis and the equality theorem, $\vdash_{\mathrm{N}} x \cdot \mathrm{S}\dot{n} = \mathbf{b}_{n+1}$.

5.3 The representability theorem.

Representability Theorem. All recursive functions and recursive predicates are representable in N.

Proof. By (i), $\vdash_{\mathrm{N}} \dot{0} \neq \dot{1}$. Thus by the proposition of §2.3, it suffices to show that every recursive function is representable in N. For this it will suffice to prove that initial functions are representable in N, and that the composition and minimization of functions representable in N are representable in N.† We start by proving that initial functions are representable in N. By (ii) and (iii), $x + y$ with x, y represents $+$ in N and $x \cdot y$ with x, y represents $\cdot$ in N. By the proposition of §2.2, $+$ and $\cdot$ are representable in N. By (iv) and (v), $x < y$ with x, y represents $<$ in N; since $\vdash_{\mathrm{N}} \dot{0} \neq \dot{1}$, $\chi_<$ is representable in N by the proposition of §2.3. Finally, x_i with $x_1, \dots, x_n$ represents π_i^n in N by the identity axioms and the substitution rule, so π_i^n is representable in N by the proposition of §2.2.

Suppose that f is the composition of $g, h_1, \dots, h_k$, and that $g, h_1, \dots, h_k$ are representable in N. Then there are formulae $\mathbf{B}, \mathbf{C}_1, \dots, \mathbf{C}_k$ such that $\mathbf{B}$ with $y_1, \dots, y_k, z$ represents g in N and $\mathbf{C}_i$ with $x_1, \dots, x_n, y_i$ represents h_i in N, for all i. We may assume that no variable other than $y_1, \dots, y_k, z$ is free in $\mathbf{B}$. Let $\mathbf{A}$ be

$$\exists y_1 \dots \exists y_k (\mathbf{C}_1 \wedge \cdots \wedge \mathbf{C}_k \wedge \mathbf{B}).$$

Suppose that $h_i(a_1, \dots, a_n) = c_i$ and that $f(a_1, \dots, a_n) = g(c_1, \dots, c_k) = b$. Then for all i, $\vdash_{\mathrm{N}} \mathbf{C}_i[x_1, \dots, x_n|\dot{a}_1, \dots, \dot{a}_n] \leftrightarrow y_i = \dot{c}_i$, whence $\vdash_{\mathrm{N}} \mathbf{A}[x_1, \dots, x_n|\dot{a}_1, \dots, \dot{a}_n] \leftrightarrow (y_1 = \dot{c}_1 \wedge \cdots \wedge y_k = \dot{c}_k \wedge \mathbf{B})$ by the equivalence theorem. By k uses of the replacement theorem, we obtain $\vdash_{\mathrm{N}} \mathbf{A}[x_1, \dots, x_n|\dot{a}_1, \dots, \dot{a}_n] \leftrightarrow \mathbf{B}[y_1, \dots, y_k|\dot{c}_1, \dots, \dot{c}_k]$. But $\vdash_{\mathrm{N}} \mathbf{B}[y_1, \dots, y_k|\dot{c}_1, \dots, \dot{c}_k] \leftrightarrow z = \dot{b}$ by representability and hence $\vdash_{\mathrm{N}} \mathbf{A}[x_1, \dots, x_n|\dot{a}_1, \dots, \dot{a}_n] \leftrightarrow z = \dot{b}$ by the tautology theorem. Thus, $\mathbf{A}$ with $x_1, \dots, x_n, z$ represents f in N.

Finally, suppose that f is the minimization of g and that g is representable in N. There is a formula $\mathbf{B}$ such that $\mathbf{B}$ with $x_1, \dots, x_n, y, z$ represents g in N. Let $\mathbf{A}$ be

$$\mathbf{B}[z|\dot{0}] \wedge \forall w (w < y \rightarrow \neg \mathbf{B}[y, z|w, \dot{0}]).$$

Suppose that $f(a_1, \dots, a_n) = b$. Then $g(a_1, \dots, a_n, b) = 0$ and $g(a_1, \dots, a_n, k) = c_k \neq 0$ for $k < b$. By representability, $\vdash_{\mathrm{N}} \mathbf{B}[x_1, \dots, x_n, y|\dot{a}_1, \dots, \dot{a}_n, \dot{b}] \leftrightarrow z = \dot{0}$, so

$$\vdash_{\mathrm{N}} \mathbf{B}[x_1, \dots, x_n, y, z|\dot{a}_1, \dots, \dot{a}_n, \dot{b}, \dot{0}] \tag{1}$$

by the equality theorem and the tautology theorem, and hence

$$\vdash_{\mathrm{N}} y = \dot{b} \rightarrow \mathbf{B}[x_1, \dots, x_n, z|\dot{a}_1, \dots, \dot{a}_n, \dot{0}] \tag{2}$$

by the equality theorem. Also by representability, $\vdash_{\mathrm{N}} \mathbf{B}[x_1, \dots, x_n, y|\dot{a}_1, \dots, \dot{a}_n, \dot{k}] \leftrightarrow z = \dot{c}_k$ if $k < b$. Since $\vdash_{\mathrm{N}} \dot{c}_k \neq \dot{0}$ by (i), $\vdash_{\mathrm{N}} y = \dot{k} \rightarrow \neg \mathbf{B}[x_1, \dots, x_n, z|\dot{a}_1, \dots, \dot{a}_n, \dot{0}]$ by the tautology theorem and the equality theorem. By (ix), this implies $\vdash_{\mathrm{N}} y < \dot{b} \rightarrow \neg \mathbf{B}[x_1, \dots, x_n, z|\dot{a}_1, \dots, \dot{a}_n, \dot{0}]$), whence

$$\vdash_{\mathrm{N}} \mathbf{B}[x_1, \dots, x_n, z|\dot{a}_1, \dots, \dot{a}_n, \dot{0}] \rightarrow \neg(y < \dot{b}) \tag{3}$$

by the tautology theorem, and

$$\vdash_{\mathrm{N}} y = \dot{b} \rightarrow \forall w (w < y \rightarrow \neg \mathbf{B}[x_1, \dots, x_n, y, z|\dot{a}_1, \dots, \dot{a}_n, w, \dot{0}]) \tag{4}$$

by the substitution rule, the generalization rule, and the equality theorem. By the substitution theorem, (1), and the tautology theorem, $\vdash_{\mathrm{N}} \forall w (w < y \rightarrow \neg \mathbf{B}[x_1, \dots, x_n, y, z|\dot{a}_1, \dots, \dot{a}_n, w, \dot{0}]) \rightarrow \neg(\dot{b} < y)$. From this, (3), and an instance of N9, we obtain

$$\vdash_{\mathrm{N}} \mathbf{A}[x_1, \dots, x_n|\dot{a}_1, \dots, \dot{a}_n] \rightarrow y = \dot{b} \tag{5}$$

by the tautology theorem. From (2), (4), and (5) we get $\vdash_{\mathrm{N}} \mathbf{A}[x_1, \dots, x_n|\dot{a}_1, \dots, \dot{a}_n] \leftrightarrow y = \dot{b}$ by the tautology theorem. So $\mathbf{A}$ with $x_1, \dots, x_n, y$ represents f in N. □

†This statement holds in fact for any extension T of N, as will be apparent from the proof. But we shall never use this stronger result.

5.4 The incompleteness of arithmetic. It is easy to devise a recursive numerotation for $L(\mathrm{N})$. For example, we can set the symbol number of the $(n+1)$th variable to be $n+9$, and assign the numbers 0 through 8 to the other symbols of $L(\mathrm{N})$. For any arithmetization of N, $\mathrm{nlax}_{\mathrm{N}}$ is recursive, because if $n_1, \dots, n_9$ are the expression numbers of N1–N9 then $\mathrm{nlax}_{\mathrm{N}}(a) \leftrightarrow a = n_1 \vee \cdots \vee a = n_9$. Thus N itself can be arithmetized so that it satisfies the hypotheses of both versions of the incompleteness theorem (taking $\mathbf{U}$ to be $x = x$ for the second version).

More generally, suppose that T is a first-order theory which is arithmetized from a recursive numerotation by a recursive coding function in such a way that nlax_T is recursive, and suppose given an interpretation I of N in T for which $=_I$ is $=$. Renaming the symbols of T if necessary, we may assume that $\dot{0}_I$ is $\dot{0}$, S_I is S, and $<_I$ is $<$. By the representability theorem and the proposition of §2.4, every recursive function or predicate is representable in T, and hence T satisfies the hypotheses of the first version of the incompleteness theorem. Taking $\mathbf{U}$ to be the formula $\mathrm{U}_I x$, we see that T is also subject to the second version of the incompleteness theorem.

Theorem. Let T be a first-order theory and I an interpretation of N in an extension by definitions of T such that $=_I$ is $=$. Assume that T is arithmetized from a recursive numerotation by a recursive coding function in such a way that nlax_T is recursive. If T is complete, then T is inconsistent.

Proof. Let σ be the given numerotation of T, β the given coding function, and T' the given extension by definitions of T. If T' is obtained from T by the adjunction of n defined symbols, we can devise a recursive numerotation σ' for T' by letting $\sigma'(\mathbf{s})$ be $\sigma(\mathbf{s}) + n$ if $\mathbf{s}$ is a symbol of T and by assigning the numbers 0 to $n-1$ to the new symbols. Let T' be arithmetized from σ' by β, and let $a_1, \dots, a_n$ be the expression numbers of the new nonlogical axioms of T'. Define recursive functions f and g by

$$f(b) = \mu a(\mathrm{len}(a) = \mathrm{len}(b) \wedge (a)_0 = (b)_0 + n \wedge \forall i_{<\mathrm{len}(b)-1}((a)_{i+1} = f((b)_{i+1}))),$$
$$g(a) = \mu b(\mathrm{len}(b) = \mathrm{len}(a) \wedge (b)_0 = (a)_0 - n \wedge \forall i_{<\mathrm{len}(a)-1}((b)_{i+1} = g((a)_{i+1}))).$$

Then

$$\mathrm{nlax}_{T'}(a) \leftrightarrow (\mathrm{nlax}_T(g(a)) \wedge a = f(g(a))) \vee a = a_1 \vee \cdots \vee a = a_n,$$

so $\mathrm{nlax}_{T'}$ is recursive. By the theorem on definitions, T' is also complete, so it is inconsistent by the preceding discussion. But T' is a conservative extension of T, so T is inconsistent. ☐

We end this section by noting that N is, in fact, consistent and, therefore, incomplete. Observe that all the nonlogical axioms of N are open, so by the corollary to the consistency theorem, $\dot{0} \neq \dot{0}$ is a theorem of N if and only if $\dot{0} \neq \dot{0}$ is a tautological consequence of instances of equality rules and nonlogical axioms of N. By Proposition 2 of ch. I §3.3, this will be the case if and only if $\dot{0} \neq \dot{0}$ is a tautological consequence of *closed* instances of equality rules and nonlogical axioms of N. Let $\mathbf{A}_1, \dots, \mathbf{A}_n$ be such closed instances. Define a truth valuation V on $L(\mathrm{N})$ as follows: if $\mathbf{A}$ is elementary, V assigns $\mathbf{T}$ to $\mathbf{A}$ if and only if $\mathbf{A}$ is $\dot{0} = \dot{0}$ or an atomic subformula of $\mathbf{A}_1, \dots, \mathbf{A}_n$ for which $\mathbf{A}_\nu$ (see ch. IV §1.4 for the definition of the predicate $\mathbf{A}_\nu$). By analyzing each axiom in turn, we see that $V(\mathbf{A}_i)$ is $\mathbf{T}$ for all i. But clearly $V(\dot{0} \neq \dot{0})$ is $\mathbf{F}$, and so $\dot{0} \neq \dot{0}$ is not a theorem of N.

Chapter Four
First-Order Number Theory

§1 Recursive extensions

1.1 Numerical realizations. Let L be a first-order language. A *numerical realization* α of L is the association of an n-ary function $\mathbf{f}_\alpha$ to each n-ary function symbol $\mathbf{f}$ of L and of two n-ary predicates $\mathbf{p}^+_\alpha$ and $\mathbf{p}^-_\alpha$ to each n-ary predicate symbol $\mathbf{p}$ of L, with the requirement that both $=^+_\alpha$ and $=^-_\alpha$ be $=$. When $\mathbf{p}^+_\alpha$ and $\mathbf{p}^-_\alpha$ are the same predicate, we say that α *decides* $\mathbf{p}$. We define a function or predicate $\mathbf{u}_\alpha$ for every designator $\mathbf{u}$ of L. We first define $\mathbf{u}^+_\alpha$ and $\mathbf{u}^-_\alpha$ simultaneously by induction on the length of $\mathbf{u}$. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be the variables free in $\mathbf{u}$ in alphabetical order. If $\mathbf{u}$ is a variable, set $\mathbf{u}^\pm_\alpha(a) = a$. If $\mathbf{u}$ is $\mathbf{f}\mathbf{a}_1 \ldots \mathbf{a}_n$ and if $\mathbf{x}_{j_i(1)}, \ldots, \mathbf{x}_{j_i(k_i)}$ are the variables occurring in $\mathbf{a}_i$ in alphabetical order, set

$$\mathbf{u}^\pm_\alpha(a_1, \ldots, a_n) = \mathbf{f}_\alpha((\mathbf{a}_1)^\pm_\alpha(a_{j_1(1)}, \ldots, a_{j_1(k_1)}), \ldots, (\mathbf{a}_n)^\pm_\alpha(a_{j_n(1)}, \ldots, a_{j_n(k_n)})).$$

If $\mathbf{u}$ is $\mathbf{p}\mathbf{a}_1 \ldots \mathbf{a}_n$ and if $\mathbf{x}_{j_i(1)}, \ldots, \mathbf{x}_{j_i(k_i)}$ are the variables occurring in $\mathbf{a}_i$ in alphabetical order, set

$$\mathbf{u}^\pm_\alpha(a_1, \ldots, a_n) \leftrightarrow \mathbf{p}^\pm_\alpha((\mathbf{a}_1)^\pm_\alpha(a_{j_1(1)}, \ldots, a_{j_1(k_1)}), \ldots, (\mathbf{a}_n)^\pm_\alpha(a_{j_n(1)}, \ldots, a_{j_n(k_n)}))$$

If $\mathbf{u}$ is $\mathbf{B} \vee \mathbf{C}$ and if $\mathbf{x}_{i_1}, \ldots, \mathbf{x}_{i_k}$ (resp. $\mathbf{x}_{j_1}, \ldots, \mathbf{x}_{j_l}$) are the variables occurring in $\mathbf{B}$ (resp. in $\mathbf{C}$) in alphabetical order, set $\mathbf{u}^\pm_\alpha(a_1, \ldots, a_n) \leftrightarrow \mathbf{B}^\pm_\alpha(a_{i_1}, \ldots, a_{i_k}) \vee \mathbf{C}^\pm_\alpha(a_{j_1}, \ldots, a_{j_l})$. If $\mathbf{u}$ is $\neg\mathbf{B}$, set $\mathbf{u}^\pm_\alpha(a_1, \ldots, a_n) \leftrightarrow \neg\mathbf{B}^\mp_\alpha(a_1, \ldots, a_n)$. Finally, if $\mathbf{u}$ is $\exists\mathbf{x}\mathbf{B}$, set $\mathbf{u}^\pm_\alpha(a_1, \ldots, a_n) \leftrightarrow \mathbf{B}^\pm_\alpha(a_1, \ldots, a_n)$ if $\mathbf{x}$ is not free in $\mathbf{B}$ and $\mathbf{u}^\pm_\alpha(a_1, \ldots, a_n) \leftrightarrow \exists a\mathbf{B}^\pm_\alpha(a_1, \ldots, a_i, a, a_{i+1}, \ldots, a_n)$ otherwise, where i is such that $\mathbf{x}$ precedes $\mathbf{x}_{i+1}$ and is preceded by $\mathbf{x}_i$ in the alphabetical order ($0 \le i \le n$). Let $\mathbf{u}_\alpha$ be $\mathbf{u}^+_\alpha$.

Two formulae $\mathbf{A}$ and $\mathbf{B}$ of a first-order language L will be called *numerically equivalent* if

(i) $\mathbf{A}$ and $\mathbf{B}$ have exactly the same free variables;

(ii) $\mathbf{A} \leftrightarrow \mathbf{B}$ is derivable without nonlogical axioms;

(iii) for every numerical realization α of L the predicates $\mathbf{A}_\alpha$ and $\mathbf{B}_\alpha$ are equal.

We shall use the notation $\mathbf{A} \sim \mathbf{B}$ (in this section only) to mean that $\mathbf{A}$ and $\mathbf{B}$ are numerically equivalent. Observe that numerical equivalence is an equivalence relation.

If $\mathbf{u}$ is a designator of the form $\mathbf{s}\mathbf{u}_1 \ldots \mathbf{u}_n$, we see at once from the inductive definition of $\mathbf{u}^\pm_\alpha$ that $\mathbf{u}^\pm_\alpha$ depends only on each $(\mathbf{u}_i)^\pm_\alpha$ and on which variables are free in which $\mathbf{u}_i$. Also, condition (iii) implies that $\mathbf{A}^-_\alpha$ and $\mathbf{B}^-_\alpha$ are equal for every α, as we see by considering the "opposite" realization. These two remarks, together with the equivalence theorem, immediatly imply the following result.

Proposition 1. If $\mathbf{B}$ is obtained from $\mathbf{A}$ by replacing an occurence of $\mathbf{C}$ by $\mathbf{D}$ and if $\mathbf{C} \sim \mathbf{D}$, then $\mathbf{A} \sim \mathbf{B}$.

Proposition 2. The formulae $\mathbf{A}[\mathbf{x}|\mathbf{a}]$ and $\exists\mathbf{x}(\mathbf{x} = \mathbf{a} \wedge \mathbf{A})$ are numerically equivalent, provided that $\mathbf{a}$ is substitutible for $\mathbf{x}$ in $\mathbf{A}$, that $\mathbf{x}$ is free in $\mathbf{A}$, and that $\mathbf{x}$ does not occur in $\mathbf{a}$

Proof. Condition (i) is obvious and condition (ii) is true by the replacement theorem. To prove (iii), consider the more general situation of a designator $\mathbf{u}$ in which $\mathbf{x}$ is free and $\mathbf{a}$ is substitutible for $\mathbf{x}$. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be the variables free in $\mathbf{u}[\mathbf{x}|\mathbf{a}]$, $\mathbf{x}_{i_1}, \ldots, \mathbf{x}_{i_k}$ those among them free in $\mathbf{u}$, and $\mathbf{x}_{j_1}, \ldots, \mathbf{x}_{j_l}$ those among them occurring in $\mathbf{a}$ (all arranged in alphabetical order). Suppose that $\mathbf{x}$ comes between $\mathbf{x}_{i_r}$ and $\mathbf{x}_{i_{r+1}}$ in the alphabetical order. It is then easily proved by induction on the length of $\mathbf{u}$ that

$$\mathbf{u}[\mathbf{x}|\mathbf{a}]^\pm_\alpha(a_1, \ldots, a_n) \text{ is } \mathbf{u}^\pm_\alpha(a_{i_1}, \ldots, a_{i_r}, \mathbf{a}_\alpha(a_{j_1}, \ldots, a_{j_l}), a_{i_{r+1}}, \ldots, a_{i_k}),$$

i.e., the left-hand and right-hand sides are the same number if $\mathbf{u}$ is a term and are equivalent if $\mathbf{u}$ is a formula. On the other hand, since $=^+_\alpha$ is $=$, we have

$$(\exists\mathbf{x}(\mathbf{x} = \mathbf{a} \wedge \mathbf{A}))_\alpha(a_1, \ldots, a_n) \leftrightarrow \exists a(a = \mathbf{a}_\alpha(a_{j_1}, \ldots, a_{j_l}) \wedge \mathbf{A}_\alpha(a_{i_1}, \ldots, a_{i_r}, a, a_{i_{r+1}}, \ldots, a_{i_k})),$$

whence the result. □

1.2 RE-formulae. Let L be a first-order language. The *strict RE-formulae* of L are defined inductively as follows:

(i) $\mathbf{y} = \mathbf{f}\mathbf{x}_1 \dots \mathbf{x}_n$ is a strict RE-formula;

(ii) $\mathbf{p}\mathbf{x}_1 \dots \mathbf{x}_n$ and $\neg\mathbf{p}\mathbf{x}_1 \dots \mathbf{x}_n$ are strict RE-formulae;

(iii) if $\mathbf{A}$ and $\mathbf{B}$ are strict RE-formulae, then $\mathbf{A} \vee \mathbf{B}$ and $\mathbf{A} \wedge \mathbf{B}$ are strict RE-formulae;

(iv) if L has the binary predicate symbol $<$ and if $\mathbf{A}$ is a strict RE-formula, then $\forall\mathbf{x}(\mathbf{x} < \mathbf{y} \to \mathbf{A})$ is a strict RE-formula;

(v) if $\mathbf{A}$ is a strict RE-formula, then $\exists\mathbf{x}\mathbf{A}$ is a strict RE-formula.

A formula $\mathbf{A}$ of L is called an *RE-formula* if it is numerically equivalent to a strict RE-formula of L. It is a *PR-formula* if both $\mathbf{A}$ and $\neg\mathbf{A}$ are RE-formulae.

We shall establish the following closure properties of the classes of RE-formulae and of PR-formulae.

(i′) Open formulae are PR-formulae.

(ii′) Any instance of an RE-formula is an RE-formula.

(iii′) If $\mathbf{A}$ and $\mathbf{B}$ are RE-formulae, then $\mathbf{A} \vee \mathbf{B}$, $\mathbf{A} \wedge \mathbf{B}$, $\forall\mathbf{x}(\mathbf{x} < \mathbf{a} \to \mathbf{A})$, and $\exists\mathbf{x}\mathbf{A}$ are RE-formulae.

(iv′) If $\mathbf{A}$ and $\mathbf{B}$ are PR-formulae, then $\mathbf{A} \vee \mathbf{B}$, $\neg\mathbf{A}$, $\exists\mathbf{x}(\mathbf{x} < \mathbf{a} \wedge \mathbf{A})$, and $\forall\mathbf{x}(\mathbf{x} < \mathbf{a} \to \mathbf{A})$ are PR-formulae.

We first show that $\mathbf{x} = \mathbf{a}$ is an RE-formula by induction on the length of $\mathbf{a}$. If $\mathbf{a}$ is a variable, $\mathbf{x} = \mathbf{a}$ is a strict RE-formula. Suppose that $\mathbf{a}$ is $\mathbf{f}\mathbf{a}_1 \dots \mathbf{a}_n$ and choose new variables $\mathbf{x}_1, \dots, \mathbf{x}_n$. Then by Proposition 2 of §1.1, we have

$$\mathbf{x} = \mathbf{a} \sim \exists\mathbf{x}_1(\mathbf{x}_1 = \mathbf{a}_1 \wedge \cdots \exists\mathbf{x}_n(\mathbf{x}_n = \mathbf{a}_n \wedge \mathbf{x} = \mathbf{f}\mathbf{x}_1 \dots \mathbf{x}_n)\cdots),$$

which is an RE-formula by the induction hypothesis and Proposition 1 of §1.1.

We now prove (i′). It suffices to prove that any open formula $\mathbf{A}$ is an RE-formula, and we proceed by induction on the length of $\mathbf{A}$. If $\mathbf{A}$ is $\mathbf{p}\mathbf{a}_1 \dots \mathbf{a}_n$ and if $\mathbf{x}_1, \dots, \mathbf{x}_n$ do not occur in $\mathbf{A}$, then

$$\mathbf{A} \sim \exists\mathbf{x}_1(\mathbf{x}_1 = \mathbf{a}_1 \wedge \cdots \exists\mathbf{x}_n(\mathbf{x}_n = \mathbf{a}_n \wedge \mathbf{p}\mathbf{x}_1 \dots \mathbf{x}_n)\cdots)$$

by Proposition 2 of §1.1, and this is an RE-formula by the preliminary result and Proposition 1 of §1.1. The same proof shows that $\neg\mathbf{p}\mathbf{a}_1 \dots \mathbf{a}_n$ is an RE-formula. If $\mathbf{A}$ is $\mathbf{B} \vee \mathbf{C}$ with $\mathbf{B}$ and $\mathbf{C}$ open, $\mathbf{A}$ is an RE-formula by induction hypothesis and Proposition 1 of §1.1. Finally, suppose that $\mathbf{A}$ is $\neg\mathbf{B}$ with $\mathbf{B}$ open. If $\mathbf{B}$ is atomic, we have already seen that $\mathbf{A}$ is an RE-formula. If $\mathbf{B}$ is of the form $\neg\mathbf{C}$, then clearly $\mathbf{A} \sim \mathbf{C}$ and $\mathbf{C}$ is an RE-formula by induction hypothesis, so $\mathbf{A}$ is an RE-formula. If $\mathbf{B}$ is $\mathbf{C} \vee \mathbf{D}$, then $\mathbf{A} \sim \neg\mathbf{C} \wedge \neg\mathbf{D}$ and $\neg\mathbf{C}$ and $\neg\mathbf{D}$ are RE-formulae by induction hypothesis, so $\mathbf{A}$ is an RE-formula by Proposition 1 of §1.1.

To prove (ii′), it suffices to prove that if $\mathbf{A}$ is an RE-formula and $\mathbf{x}$ does not occur in $\mathbf{a}$, then $\mathbf{A}[\mathbf{x}|\mathbf{a}]$ is an RE-formula, since arbitrary instances can be obtained by taking several instances of this form. If $\mathbf{x}$ is not free in $\mathbf{A}$, there is nothing to prove. Otherwise, (i′) and Propositions 1 and 2 of §1.1 immediatly show that $\mathbf{A}[\mathbf{x}|\mathbf{a}]$ is an RE-formula.

Finally, the assertions (iii′) and (iv′) are easy consequences of (i′) and Propositions 1 and 2 of §1.1.

1.3 Recursive extensions. Let L be a first-order language and let T be a numerical first-order theory whose language is an extension of L. A function symbol $\mathbf{f}$ of T is called *recursive on L in T* if $\vdash_T y = \mathbf{f}x_1 \dots x_n \leftrightarrow \mathbf{A}$ for some RE-formula $\mathbf{A}$ of L. A predicate symbol $\mathbf{p}$ of T is called *recursive on L in T* if $\vdash_T (\mathbf{p}x_1 \dots x_n \wedge y = \dot{0}) \vee (\neg\mathbf{p}x_1 \dots x_n \wedge y = \dot{1}) \leftrightarrow \mathbf{A}$ for some RE-formula $\mathbf{A}$ of L. A predicate symbol $\mathbf{p}$ of T is called *recursively enumerable on L in T* if $\vdash_T \mathbf{p}x_1 \dots x_n \leftrightarrow \mathbf{A}$ for some RE-formula $\mathbf{A}$ of L. Observe that if $\vdash_T \mathbf{p}x_1 \dots x_n \leftrightarrow \mathbf{A}$ for some PR-formula $\mathbf{A}$ of L, then $\mathbf{p}$ is recursive on L in T. In these definitions we sometimes identify a list of nonlogical symbols of T with the first-order language having these symbols as its only nonlogical symbols, and we often drop the reference to T when the context makes it clear which T is intended.

Let L' be another first-order language of which $L(T)$ is an extension. We say that L' *is recursive on L in T* if every nonlogical symbol of L' is recursive on L in T. Note that L is always recursive on extensions of itself in T since $y = \mathbf{f}x_1 \dots x_n$ and $(\mathbf{p}x_1 \dots x_n \wedge y = \dot{0}) \vee (\neg\mathbf{p}x_1 \dots x_n \wedge y = \dot{1})$ are RE-formulae. We say that T is recursive on L if $L(T)$ is recursive on L in T. A *recursive extension* of T is an extension of T that is recursive on $L(T)$.

PROPOSITION. Let L and L' be numerical first-order languages having the binary predicate symbol $<$ and let T be a first-order theory in which L' is recursive on L. Suppose that $\vdash_T \dot{0} \neq \dot{1}$. Then for every RE-formula $\mathbf{A}$ of L', we can find an RE-formula $\mathbf{A}'$ of L such that $\vdash_T \mathbf{A} \leftrightarrow \mathbf{A}'$.

Proof. We may suppose that $\mathbf{A}$ is a strict RE-formula, and we proceed by induction on the length of $\mathbf{A}$. Suppose that $\mathbf{A}$ is $\mathbf{y} = \mathbf{f}\mathbf{x}_1 \ldots \mathbf{x}_n$. Since $\mathbf{f}$ is recursive on L, there is an RE-formula $\mathbf{A}'$ of L such that $\vdash_T \mathbf{A} \leftrightarrow \mathbf{A}'$, and the proposition is proved in this case. Suppose that $\mathbf{A}$ is $\mathbf{p}\mathbf{x}_1 \ldots \mathbf{x}_n$. Since $\mathbf{p}$ is recursive on L, there is an RE-formula $\mathbf{B}$ of L such that $\vdash_T (\mathbf{p}\mathbf{x}_1 \ldots \mathbf{x}_n \wedge \mathbf{y} = \dot{0}) \vee (\neg\mathbf{p}\mathbf{x}_1 \ldots \mathbf{x}_n \wedge \mathbf{y} = \dot{1}) \leftrightarrow \mathbf{B}$ for some $\mathbf{y}$ distinct from $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Since $\vdash_T \dot{0} \neq \dot{1}$, we have $\vdash_T \mathbf{y} = \dot{0} \rightarrow \mathbf{p}\mathbf{x}_1 \ldots \mathbf{x}_n \leftrightarrow \mathbf{B}$ by the tautology theorem and the equality theorem, whence $\vdash_T \mathbf{p}\mathbf{x}_1 \ldots \mathbf{x}_n \leftrightarrow \mathbf{B}[\mathbf{y}|\dot{0}]$ by the equality theorem, and the right-hand side is an RE-formula of L. We derive similarly $\vdash_T \neg\mathbf{p}\mathbf{x}_1 \ldots \mathbf{x}_n \leftrightarrow \mathbf{B}[\mathbf{y}|\dot{1}]$, and the right-hand side is again an RE-formula of L. If $\mathbf{A}$ is $\mathbf{B} \vee \mathbf{C}$ (resp. $\mathbf{B} \wedge \mathbf{C}$), then the desired result follows by the induction hypothesis and the tautolgy theorem. Finally, suppose that $\mathbf{A}$ is $\forall\mathbf{x}(\mathbf{x} < \mathbf{y} \rightarrow \mathbf{B})$ (resp. $\exists\mathbf{x}\mathbf{B}$). By induction hypothesis, there is an RE-formula $\mathbf{B}'$ of L such that $\vdash_T \mathbf{B} \leftrightarrow \mathbf{B}'$. By the tautology theorem and the distribution rule, $\vdash_T \mathbf{A} \leftrightarrow \forall\mathbf{x}(\mathbf{x} < \mathbf{y} \rightarrow \mathbf{B}')$ (resp. $\vdash_T \mathbf{A} \leftrightarrow \exists\mathbf{x}\mathbf{B}'$), and the right-hand side is an RE-formula of L. □

The proposition can be rephrased as follows. For any numerical first-order theory T such that $\vdash_T \dot{0} \neq \dot{1}$, the relation "$L'$ is recursive on L" is transitive in the collection of numerical first-order languages having $L(T)$ as an extension and containing the symbol $<$. A special case of this situation that is worth mentioning is obtained when T' is an extension by definitions of T with exactly one new nonlogical symbol $\mathbf{s}$: if it has been previously proved that T is recursive on L, then it suffices that $\mathbf{s}$ be recursive on $L(T)$ in T' for T' to be recursive on L, and the former certainly happens e.g. if the right-hand side of the defining axiom of $\mathbf{s}$ is a PR-formula or, if $\mathbf{s}$ is a function symbol, an RE-formula.

Remark. A recursive function or predicate symbol is intended to be a formal analogue to a recursive function or predicate. However, even if T is a substantial extension of N, it is not always true that a recursive function or predicate can be represented by a symbol recursive on $L(\mathrm{N})$ in an extension by definitions of T. A function or predicate is sometimes called *provably recursive in T* if it has the latter property. If we use the construction given in the proposition of ch. III §2.2 to obtain a defined function symbol representing a given recursive function, this function symbol need not be recursive on $L(\mathrm{N})$ since its defining axiom might not be an RE-formula. We shall prove in chapter V that the recursive functions and predicates associated with a first-order theory arithmetized from a recursive numerotation by a recursive coding function are provably recursive in Peano arithmetic (see §2) if the numerotation and the coding function are.

1.4 The theorem on RE-formulae. Let L be a first-order language and α a numerical realization of L. If T is a numerical first-order theory whose language is an extension of L, we say that α is *faithful in T* if for every function symbol $\mathbf{f}$ of L, $\mathbf{f}$ represents $\mathbf{f}_\alpha$ in T, and for every predicate symbol $\mathbf{p}$ of L, $\mathbf{p}$ positively (negatively) represents $\mathbf{p}_\alpha^+$ ($\mathbf{p}_\alpha^-$) in T. We say that α is *well-founded in T* either if L does not have the binary predicate symbol $<$ or if for all n there exist $a_1, \ldots, a_k$ with $a_i <_\alpha^- n$ such that $\vdash_T x < \dot{n} \rightarrow x = \dot{a}_1 \vee \cdots \vee x = \dot{a}_k$.

THEOREM ON RE-FORMULAE. Let L be a first-order language, T a numerical first-order theory whose language is an extension of L, and α a numerical realization of L that is faithful and well-founded in T. If $\mathbf{A}$ is an RE-formula of L and if $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are the variables free in $\mathbf{A}$ in alphabetical order, then $\mathbf{A}$ with $\mathbf{x}_1, \ldots, \mathbf{x}_n$ positively represents $\mathbf{A}_\alpha$ in T.

Proof. We need only prove the theorem when $\mathbf{A}$ is a strict RE-formula. We assume that $\mathbf{A}_\alpha(a_1, \ldots, a_n)$ and we proceed by induction on the length of $\mathbf{A}$. Let $\mathbf{A}'$ be $\mathbf{A}[\mathbf{x}_1, \ldots, \mathbf{x}_n|\dot{a}_1, \ldots, \dot{a}_n]$, the formula to be derived. If $\mathbf{A}$ is $\mathbf{y} = \mathbf{f}\mathbf{x}_1 \ldots \mathbf{x}_n$, then $\mathbf{A}'$ has the form $\dot{a} = \mathbf{f}\dot{b}_1 \ldots \dot{b}_n$ where $a =_\alpha^+ \mathbf{f}_\alpha(b_1, \ldots, b_n)$, so $\vdash_T \mathbf{A}'$ by representability and the equality theorem. If $\mathbf{A}$ is $\mathbf{p}\mathbf{x}_1 \ldots \mathbf{x}_n$ (resp. $\neg\mathbf{p}\mathbf{x}_1 \ldots \mathbf{x}_n$) then $\mathbf{A}'$ has the form $\mathbf{p}\dot{b}_1 \ldots \dot{b}_n$ (resp. $\neg\mathbf{p}\dot{b}_1 \ldots \dot{b}_n$) where $\mathbf{p}_\alpha^+(b_1, \ldots, b_n)$ (resp. $\neg\mathbf{p}_\alpha^-(b_1, \ldots, b_n)$), and hence $\vdash_T \mathbf{A}'$ by positive (resp. negative) representability. If $\mathbf{A}$ is $\mathbf{B} \vee \mathbf{C}$ or $\mathbf{B} \wedge \mathbf{C}$, the result follows from the induction hypothesis by the tautology theorem. If $\mathbf{A}$ is $\forall\mathbf{x}(\mathbf{x} < \mathbf{y} \rightarrow \mathbf{B})$, then $\mathbf{A}'$ has the form $\forall\mathbf{x}(\mathbf{x} < \dot{n} \rightarrow \mathbf{B}')$ and by induction hypothesis $\vdash_T \mathbf{B}'[\mathbf{x}|\dot{m}]$ for all $m <_\alpha^- n$. By the equality theorem, $\vdash_T \mathbf{x} = \dot{m} \rightarrow \mathbf{B}'$ for all $m <_\alpha^- n$, whence by well-foundedness and the tautology theorem, $\vdash_T \mathbf{x} < \dot{n} \rightarrow \mathbf{B}'$. By the generalization rule, $\vdash_T \mathbf{A}'$. Finally, suppose that $\mathbf{A}$ is $\exists\mathbf{x}\mathbf{B}$ and that $\mathbf{A}'$ is $\exists\mathbf{x}\mathbf{B}'$. Then $\vdash_T \mathbf{B}'[\mathbf{x}|\dot{n}]$ for some n by induction hypothesis. Hence $\vdash_T \mathbf{A}'$ by the substitution axioms. □

Let α be faithful and well-founded in T. If $\mathbf{A}$ is PR-formula of L with free variables $\mathbf{x}_1, \ldots, \mathbf{x}_n$ in alphabetical order, the theorem on RE-formulae shows that $\mathbf{A}$ with $\mathbf{x}_1, \ldots, \mathbf{x}_n$ positively represents $\mathbf{A}_\alpha$ in

T and negatively represents $\mathbf{A}_\alpha^-$ in T. Thus in case α decides the predicate symbols occurring in $\mathbf{A}$, we obtain that $\mathbf{A}$ with $\mathbf{x}_1, \ldots, \mathbf{x}_n$ represents $\mathbf{A}_\alpha$ in T.

We define a numerical realization ν of $L(\mathrm{N})$ as follows: $\dot{0}_\nu$ is 0, $\mathrm{S}_\nu(a) = a + 1$, $+_\nu$ is $+$, $\cdot_\nu$ is $\cdot$, $=_\nu^\pm$ are both $=$, and $<_\nu^\pm$ are both $<$. By (i)–(v) and (x) of ch. III §5.2, ν is faithful in N. By (vii) of ch. III §5.2, ν is well-founded in N.

Corollary. Let $\mathbf{A}$ be an RE-formula (resp. a PR-formula) of $L(\mathrm{N})$ and $\mathbf{x}_1, \ldots, \mathbf{x}_n$ the variables free in $\mathbf{A}$ in alphabetical order. Then $\mathbf{A}$ with $\mathbf{x}_1, \ldots, \mathbf{x}_n$ positively represents (resp. represents) $\mathbf{A}_\nu$ in N.

Here is an important example of application of this theorem. Let L be a first-order language and T a numerical first-order theory whose language is an extension of L. Suppose given a faithful and well-founded realization α of L. Then the theorem on RE-formula gives a method to derive representability conditions for symbols recursive on L in T. Suppose for instance that $\vdash_T y = \mathbf{f}x_1 \ldots x_n \leftrightarrow \mathbf{A}$ for some RE-formula $\mathbf{A}$ of L in which precisely $x_1, \ldots, x_n, y$ are free. If there is an n-ary function f such that for all numbers $a_1, \ldots, a_n$, $\mathbf{A}_\alpha(f(a_1, \ldots, a_n), a_1, \ldots, a_n)$, then $\mathbf{f}$ represents f in T. For by the theorem on RE-formulae, $\vdash_T \mathbf{A}[x_1, \ldots, x_n, y | \dot{a}_1, \ldots, \dot{a}_n, \dot{f}(a_1, \ldots, a_n)]$, whence $\vdash_T \dot{f}(a_1, \ldots, a_n) = \mathbf{f}\dot{a}_1 \ldots \dot{a}_n$. A similar conclusion holds for recursive predicate symbols: if $\vdash_T (\mathbf{p}x_1 \ldots x_n \wedge y = \dot{0}) \vee (\neg \mathbf{p}x_1 \ldots x_n \wedge y = \dot{1}) \leftrightarrow \mathbf{A}$ for some RE-formula $\mathbf{A}$ of L in which exactly $x_1, \ldots, x_n, y$ are free, and if we can find a predicate p such that for all numbers $a_1, \ldots, a_n$, $\mathbf{A}_\alpha(\chi_p(a_1, \ldots, a_n), a_1, \ldots, a_n)$, we obtain $\vdash_T (\mathbf{p}\dot{a}_1 \ldots \dot{a}_n \wedge \dot{\chi}_p(a_1, \ldots, a_n) = \dot{0}) \vee (\neg \mathbf{p}\dot{a}_1 \ldots \dot{a}_n \wedge \dot{\chi}_p(a_1, \ldots, a_n) = \dot{1})$; then if $\vdash_T \dot{0} \neq \dot{1}$, we conclude that $\mathbf{p}$ represents p in T.

We see that the theorem on RE-formulae is a powerful tool to obtain constructive proofs of representability of functions and predicates in recursive extensions.

§2 The first-order theory PA

2.1 Peano arithmetic. We introduce a first-order theory known as *Peano arithmetic* and denoted by PA. The language of PA is $L(\mathrm{N})$ and the axioms of PA are the axioms N1–N8 of N and all the formulae of the form

$$\mathbf{A}[\mathbf{x}|\dot{0}] \rightarrow \forall \mathbf{x}(\mathbf{A} \rightarrow \mathbf{A}[\mathbf{x}|\mathrm{S}\mathbf{x}]) \rightarrow \mathbf{A}, \tag{1}$$

called *induction axioms*.

An extension P of PA is called a *good extension* if the formula (1) is a theorem of P for any formula $\mathbf{A}$ of P. Such a formula is then also called an induction axiom of P. It is certainly the case that if P is obtained from PA by the adjunction of new constants and new axioms, then P is a good extension of PA (by the substitution rule). Note also that if P is a good extension of PA and if P' is an extension by definitions of P, then P' is a good extension as well. For a translation of (1) into P is simply obtained by replacing $\mathbf{A}$ by a translation $\mathbf{A}^*$, and hence is an induction axiom of P.

2.2 PA is an extension of N. We note that PA is an extension of N by deriving N9 in PA. We have $\vdash_{\mathrm{PA}} \dot{0} < \dot{0} \vee \dot{0} = \dot{0}$ by the identity axioms and, by N8, $\vdash_{\mathrm{PA}} \dot{0} < y \vee \dot{0} = y \rightarrow \dot{0} < \mathrm{S}y \vee \dot{0} = \mathrm{S}y$, whence

$$\vdash_{\mathrm{PA}} \dot{0} < y \vee \dot{0} = y \tag{2}$$

by the induction axioms. Let $\mathbf{A}$ be the formula $x < y \rightarrow \mathrm{S}x < \mathrm{S}y$. By N7, $\vdash_{\mathrm{PA}} \mathbf{A}[y|\dot{0}]$, and by two instances of N8 and the equivalence theorem, $\vdash_{\mathrm{PA}} \mathbf{A}[y|\mathrm{S}y] \leftrightarrow x < y \vee x = y \rightarrow \mathrm{S}x < \mathrm{S}y \vee \mathrm{S}x = \mathrm{S}y$. By the equality axioms, $\vdash_{\mathrm{PA}} x = y \rightarrow \mathrm{S}x = \mathrm{S}y$, and so by the tautology theorem, $\vdash_{PA} \mathbf{A} \rightarrow \mathbf{A}[y|\mathrm{S}y]$. Hence $\vdash_{\mathrm{PA}} \mathbf{A}$ by the induction axioms. Let $\mathbf{B}$ be N9. By (2), $\vdash_{\mathrm{PA}} \mathbf{B}[x|\dot{0}]$. The formula $\mathbf{B} \rightarrow \mathbf{B}[x|\mathrm{S}x]$ is a tautological consequence of $x = y \vee x < y \rightarrow y < \mathrm{S}x$ and $\mathbf{A}$. But $\vdash_{\mathrm{PA}} x = y \vee y < x \rightarrow y < \mathrm{S}x$ by an instance of N8, the symmetry theorem, and the tautology theorem. Hence $\vdash_{\mathrm{PA}} \mathbf{B} \rightarrow \mathbf{B}[x|\mathrm{S}x]$, and we conclude using the induction axioms.

2.3 A few theorems of PA. We introduce some definitions: $x \leq y \leftrightarrow x < y \vee x = y$, $\mathrm{Div}\, xy \leftrightarrow \exists z(x = y \cdot z)$, and $z = -xy \leftrightarrow (y \leq x \wedge x = y + z) \vee (x < y \wedge z = \dot{0})$. We must check that

$$\vdash_{\mathrm{PA}} \exists z((y \leq x \wedge x = y + z) \vee (x < y \wedge z = \dot{0})) \text{ and}$$
$$\vdash_{\mathrm{PA}} (y \leq x \wedge x = y + z) \vee (x < y \wedge z = \dot{0}) \rightarrow (y \leq x \wedge x = y + z') \vee (x < y \wedge z = \dot{0}) \rightarrow z = z'.$$

These follow from (i), (vi), and (x) below. We abbreviate $-\mathbf{ab}$ by $(\mathbf{a} - \mathbf{b})$.

(i) $\vdash_{\mathrm{PA}} x + y = y + x$;

(ii) $\vdash_{\mathrm{PA}} x + (y + z) = (x + y) + z$;

(iii) $\vdash_{\mathrm{PA}} x \cdot y = y \cdot x$;

(iv) $\vdash_{\mathrm{PA}} x \cdot (y \cdot z) = (x \cdot y) \cdot z$;

(v) $\vdash_{\mathrm{PA}} x \cdot (y + z) = (x \cdot y) + (x \cdot z)$;

(vi) $\vdash_{\mathrm{PA}} x + z = y + z \rightarrow x = y$;

(vii) $\vdash_{\mathrm{PA}} z \neq \dot{0} \rightarrow x \cdot z = y \cdot z \rightarrow x = y$;

(viii) $\vdash_{\mathrm{PA}} x \neq 0 \leftrightarrow \exists y(x = \mathrm{S}y)$;

(ix) $\vdash_{\mathrm{PA}} x < y \leftrightarrow \exists z(y = x + \mathrm{S}z)$;

(x) $\vdash_{\mathrm{PA}} x \leq y \leftrightarrow \exists z(y = x + z)$;

(xi) $\vdash_{\mathrm{PA}} x \neq \dot{0} \leftrightarrow \dot{0} < x$;

(xii) $\vdash_{\mathrm{PA}} \dot{0} \leq x$;

(xiii) $\vdash_{\mathrm{PA}} x < y \rightarrow y < z \rightarrow x < z$;

(xiv) $\vdash_{\mathrm{PA}} x = y \rightarrow \neg(x < y)$

(xv) $\vdash_{\mathrm{PA}} x < y \rightarrow \neg(x = y \vee y < x)$;

(xvi) $\vdash_{\mathrm{PA}} x < y \rightarrow z \leq w \rightarrow x + z < y + w$;

(xvii) $\vdash_{\mathrm{PA}} w \neq \dot{0} \rightarrow x < y \rightarrow z \leq w \rightarrow x \cdot z < y \cdot w$;

(xviii) $\vdash_{\mathrm{PA}} x < x + \mathrm{S}y$;

(xix) $\vdash_{\mathrm{PA}} x \leq x + y$;

(xx) $\vdash_{\mathrm{PA}} y \neq \dot{0} \rightarrow x \leq x \cdot y$;

(xxi) $\vdash_{\mathrm{PA}} (x + y) - y = x$;

(xxii) $\vdash_{\mathrm{PA}} y \leq x \rightarrow (x - y) + y = x$;

(xxiii) $\vdash_{\mathrm{PA}} x - y \leq x$;

(xxiv) $\vdash_{\mathrm{PA}} x \neq \dot{0} \rightarrow y \neq \dot{0} \rightarrow x - y < x$;

(xxv) $\vdash_{\mathrm{PA}} \mathrm{Div}\, xy \leftrightarrow \exists z(z < \mathrm{S}x \wedge x = y \cdot z)$;

(xxvi) $\vdash_{\mathrm{PA}} x \neq \dot{0} \rightarrow \mathrm{Div}\, xy \rightarrow y \leq x$;

(xxvii) $\vdash_{\mathrm{PA}} \mathrm{Div}\, \dot{0}x$;

(xxviii) $\vdash_{\mathrm{PA}} \mathrm{Div}\, \dot{1}x \rightarrow x = \dot{1}$;

(xxix) $\vdash_{\mathrm{PA}} \mathrm{Div}\, x\dot{0} \leftrightarrow x = \dot{0}$;

(xxx) $\vdash_{\mathrm{PA}} \mathrm{Div}\, xy \rightarrow \mathrm{Div}\, yz \rightarrow \mathrm{Div}\, xz$;

(xxxi) $\vdash_{\mathrm{PA}} \mathrm{Div}\, xy \rightarrow \mathrm{Div}(x \cdot z)(y \cdot z)$;

(xxxii) $\vdash_{\mathrm{PA}} z \neq \dot{0} \rightarrow \mathrm{Div}(x \cdot z)(y \cdot z) \rightarrow \mathrm{Div}\, xy$;

(xxxiii) $\vdash_{\mathrm{PA}} \mathrm{Div}\, xz \rightarrow \mathrm{Div}\, yz \rightarrow \mathrm{Div}(x + y)z$;

(xxxiv) $\vdash_{\mathrm{PA}} \mathrm{Div}\, xz \rightarrow \mathrm{Div}(x + y)z \rightarrow \mathrm{Div}\, yz$;

By the identity axioms, $\vdash_{\mathrm{PA}} \dot{0} + \dot{0} = \dot{0} + \dot{0}$, and by N1, N4, and the equality theorem, $\vdash_{\mathrm{PA}} x + \dot{0} = \dot{0} + x \rightarrow \mathrm{S}x + \dot{0} = \dot{0} + \mathrm{S}x$, so

$$\vdash_{\mathrm{PA}} x + \dot{0} = \dot{0} + x \tag{3}$$

by the induction axioms. By N1 and the equality theorem, $\vdash_{\mathrm{PA}} \mathrm{S}y + \dot{0} = \mathrm{S}(y + \dot{0})$, and by N4 $\vdash_{\mathrm{PA}} \mathrm{S}y + \mathrm{S}x = \mathrm{S}(\mathrm{S}y + x)$ whence $\vdash_{\mathrm{PA}} \mathrm{S}y + x = \mathrm{S}(y + x) \rightarrow \mathrm{S}y + \mathrm{S}x = \mathrm{SS}(y + x)$. But by N4 and the equality theorem, $\vdash_{\mathrm{PA}} \mathrm{SS}(y + x) = \mathrm{S}(y + \mathrm{S}x)$, so by the equality theorem, $\vdash_{\mathrm{PA}} \mathrm{S}y + x = \mathrm{S}(y + x) \rightarrow \mathrm{S}y + \mathrm{S}x = \mathrm{S}(y + \mathrm{S}x)$. By the induction axioms $\vdash_{\mathrm{PA}} \mathrm{S}y + x = \mathrm{S}(y + x)$, whence by N4 and the equality theorem, $\vdash_{\mathrm{PA}} \mathrm{S}y + x = y + \mathrm{S}x$. Therefore by N4 and the equality theorem,

$$\vdash_{\mathrm{PA}} x + y = y + x \rightarrow x + \mathrm{S}y = \mathrm{S}y + x. \tag{4}$$

From (3) and (4) by the induction axioms, we obtain (i).

The derivations of (ii)–(xxxiv) are of a similar nature and are omitted. Complete derivations of many of them can be found in [5].

Clearly $\le$ and $-$ are recursive on $L(\mathrm{N})$. By (xxv), Div is also recursive on $L(\mathrm{N})$. The above theorems and N1–N9 allow us to derive all the elementary identities of arithmetic, such as $(x + y) \cdot (x + y) = ((x \cdot x) + (\dot{2} \cdot (x \cdot y))) + (y \cdot y)$, using only the logical rules. All such results will often be used tacitly afterwards as the complexity increases.

§3 Definitions in PA

3.1 Complete induction. In this section we shall prove general methods for building extensions by definitions of PA and introduce such extensions which will be useful later on. In this paragraph, P is a good extension of PA.

Principle of Complete Induction. Let $\mathbf{A}$ be a formula of P and $\mathbf{x}$, $\mathbf{y}$ distinct variables such that $\mathbf{y}$ does not occur in $\mathbf{A}$. Then $\vdash_P \forall\mathbf{x}(\forall\mathbf{y}(\mathbf{y} < \mathbf{x} \to \mathbf{A}[\mathbf{x}|\mathbf{y}]) \to \mathbf{A}) \to \forall\mathbf{x}\mathbf{A}$.

Proof. Let $\mathbf{B}$ be $\forall\mathbf{y}(\mathbf{y} < \mathbf{x} \to \mathbf{A}[\mathbf{x}|\mathbf{y}])$. By N7 and the tautology theorem, $\vdash_P \mathbf{y} < \dot{0} \to \mathbf{A}[\mathbf{x}|\mathbf{y}]$, whence

$$\vdash_P \mathbf{B}[\mathbf{x}|\dot{0}] \tag{1}$$

by the generalization rule. By N8 and the tautology theorem, $\vdash_P (\mathbf{y} < \mathrm{S}\mathbf{x} \to \mathbf{A}[\mathbf{x}|\mathbf{y}]) \leftrightarrow (\mathbf{y} < \mathbf{x} \to \mathbf{A}[\mathbf{x}|\mathbf{y}]) \land (\mathbf{y} = \mathbf{x} \to \mathbf{A}[\mathbf{x}|\mathbf{y}])$. From this by the equivalence theorem, $\forall$–$\land$ distributivity, and the replacement theorem, $\vdash_P \mathbf{B}[\mathbf{x}|\mathrm{S}\mathbf{x}] \leftrightarrow \mathbf{B} \land \mathbf{A}$. Now by the substitution theorem and the tautology theorem,

$$\vdash_P \forall\mathbf{x}(\mathbf{B} \to \mathbf{A}) \to \mathbf{B} \to \mathbf{B} \land \mathbf{A}, \tag{2}$$

so $\vdash_P \forall\mathbf{x}(\mathbf{B} \to \mathbf{A}) \to \mathbf{B} \to \mathbf{B}[\mathbf{x}|\mathrm{S}\mathbf{x}]$, and hence

$$\vdash_P \forall\mathbf{x}(\mathbf{B} \to \mathbf{A}) \to \forall\mathbf{x}(\mathbf{B} \to \mathbf{B}[\mathbf{x}|\mathrm{S}\mathbf{x}]) \tag{3}$$

by the $\forall$-introduction rule. From (1), (3), and the induction axioms, $\vdash_P \forall\mathbf{x}(\mathbf{B} \to \mathbf{A}) \to \mathbf{B}$, and by (2), the tautology theorem, and the $\forall$-introduction rule, $\vdash_P \forall\mathbf{x}(\mathbf{B} \to \mathbf{A}) \to \forall\mathbf{x}\mathbf{A}$. □

The following corollary will also be called the principle of complete induction.

Corollary. Let $\mathbf{A}$ be a formula of P and $\mathbf{x}_1, \dots, \mathbf{x}_n, \mathbf{y}_1, \dots, \mathbf{y}_n$ distinct variables such that $\mathbf{y}_1, \dots, \mathbf{y}_n$ do not occur in $\mathbf{A}$. If $\mathbf{f}$ is an n-ary function symbol of P, then

$$\vdash_P \forall\mathbf{x}_1 \dots \forall\mathbf{x}_n(\forall\mathbf{y}_1 \dots \forall\mathbf{y}_n(\mathbf{f}\mathbf{y}_1 \dots \mathbf{y}_n < \mathbf{f}\mathbf{x}_1 \dots \mathbf{x}_n \to \mathbf{A}[\mathbf{x}_1, \dots, \mathbf{x}_n|\mathbf{y}_1, \dots, \mathbf{y}_n]) \to \mathbf{A}) \to \forall\mathbf{x}_1 \dots \forall\mathbf{x}_n\mathbf{A}.$$

Proof. Let $\mathbf{z}$ and $\mathbf{w}$ be distinct from $\mathbf{x}_1, \dots, \mathbf{x}_n, \mathbf{y}_1, \dots, \mathbf{y}_n$ and not occurring in $\mathbf{A}$ and let $\mathbf{B}$ be the formula $\forall\mathbf{x}_1 \dots \forall\mathbf{x}_n(\mathbf{z} = \mathbf{f}\mathbf{x}_1 \dots \mathbf{x}_n \to \mathbf{A})$. Using prenex operations, the equality theorem, and the equivalence theorem, we find

$$\begin{aligned}\vdash_P \forall\mathbf{x}_1 \dots \forall\mathbf{x}_n(\forall\mathbf{y}_1 \dots \forall\mathbf{y}_n(\mathbf{f}\mathbf{y}_1 \dots \mathbf{y}_n < \mathbf{f}\mathbf{x}_1 \dots \mathbf{x}_n \to \mathbf{A}[\mathbf{x}_1, \dots, \mathbf{x}_n|\mathbf{y}_1, \dots, \mathbf{y}_n]) \to \mathbf{A})&\\ \leftrightarrow \forall\mathbf{z}(\forall\mathbf{w}(\mathbf{w} < \mathbf{z} \to \mathbf{B}[\mathbf{z}|\mathbf{w}]) \to \mathbf{B}).&\end{aligned}$$

We also clearly have $\vdash_P \forall\mathbf{z}\mathbf{B} \to \forall\mathbf{x}_1 \dots \forall\mathbf{x}_n\mathbf{A}$. By the principle of complete induction, $\vdash_P \forall\mathbf{z}(\forall\mathbf{w}(\mathbf{w} < \mathbf{z} \to \mathbf{B}[\mathbf{z}|\mathbf{w}]) \to \mathbf{B}) \to \forall\mathbf{z}\mathbf{B}$. The desired theorem is a tauological consequence of those three formulae. □

Least Number Principle. Let $\mathbf{A}$ be a formula of P and $\mathbf{x}$, $\mathbf{y}$ distinct variables such that $\mathbf{y}$ does not occur in $\mathbf{A}$. Then $\vdash_P \exists\mathbf{x}\mathbf{A} \to \exists\mathbf{x}(\mathbf{A} \land \forall\mathbf{y}(\mathbf{y} < \mathbf{x} \to \neg\mathbf{A}[\mathbf{x}|\mathbf{y}]))$. If $\vdash_P \exists\mathbf{x}\mathbf{A}$, then existence and uniqueness conditions for $\mathbf{x}$ in $\mathbf{A} \land \forall\mathbf{y}(\mathbf{y} < \mathbf{x} \to \neg\mathbf{A}[\mathbf{x}|\mathbf{y}])$ are theorems of P.

Proof. The first assertion follows from the principle of complete induction (with $\neg\mathbf{A}$ instead of $\mathbf{A}$) by the tautology theorem and the equivalence theorem. Assuming $\vdash_P \exists\mathbf{x}\mathbf{A}$, the existence condition is obtained from the first assertion by the detachment rule. A uniqueness condition has the form

$$\mathbf{A} \land \forall\mathbf{y}(\mathbf{y} < \mathbf{x} \to \neg\mathbf{A}[\mathbf{x}|\mathbf{y}]) \to \mathbf{A}[\mathbf{x}|\mathbf{x}'] \land \forall\mathbf{y}(\mathbf{y} < \mathbf{x}' \to \neg\mathbf{A}[\mathbf{x}|\mathbf{y}]) \to \mathbf{x} = \mathbf{x}' \tag{4}$$

for some suitable $\mathbf{x}'$. By the substitution theorem and the tautology theorem, $\vdash_P \mathbf{A} \land \forall\mathbf{y}(\mathbf{y} < \mathbf{x}' \to \neg\mathbf{A}[\mathbf{x}|\mathbf{x}']) \to \neg(\mathbf{x} < \mathbf{x}')$ and $\vdash_P \mathbf{A}[\mathbf{x}|\mathbf{x}'] \land \forall\mathbf{y}(\mathbf{y} < \mathbf{x} \to \neg\mathbf{A}) \to \neg(\mathbf{x}' < \mathbf{x})$. Now (4) is a tautological consequence of these formulae and an instance of N9. □

Thus if $\mathbf{x}, \mathbf{x}_1, \ldots, \mathbf{x}_n$ are distinct and include the variables free in $\mathbf{A}$ and if $\mathbf{y}$ does not occur in $\mathbf{A}$, the first-order theory obtained from P by the adjunction of a new n-ary function symbol $\mathbf{f}$ and the axiom $\mathbf{x} = \mathbf{f}\mathbf{x}_1 \ldots \mathbf{x}_n \leftrightarrow \mathbf{A} \wedge \forall \mathbf{y}(\mathbf{y} < \mathbf{x} \rightarrow \neg\mathbf{A}[\mathbf{x}|\mathbf{y}])$ is an extension by definitions of P. We shall often abbreviate the defining axiom of $\mathbf{f}$ by $\mathbf{f}\mathbf{x}_1 \ldots \mathbf{x}_n = \mu\mathbf{x}\mathbf{A}$. The loss of the variable $\mathbf{y}$ in this abbreviation does not matter since different choices of $\mathbf{y}$, as long as they do not occur in $\mathbf{A}$, yield equivalent extensions by definitions by the variant theorem. Sometimes we even use $\mu\mathbf{x}\mathbf{A}$ as a term, the necessary extension by definitions being taken implicitly.

It is clear that if $\vdash_P \mathbf{A} \leftrightarrow \mathbf{B}$ for some PR-formula $\mathbf{B}$ of P, and if $\mathbf{f}$ is defined as above, then $\mathbf{f}$ is recursive on any language containing $<$ and the nonlogical symbols occurring in $\mathbf{B}$.

We remark now once and for all that if $\mathbf{A}$ has the form $\mathbf{x} = \mathbf{a} \vee \mathbf{B}$ where $\mathbf{x}$ does not occur in $\mathbf{a}$, then since $\vdash_P \exists\mathbf{x}(\mathbf{x} = \mathbf{a})$ and $\vdash_P \mathbf{x} = \mathbf{a} \rightarrow \mathbf{x} = \mathbf{a} \vee \mathbf{B}$, we have $\vdash_P \exists\mathbf{x}\mathbf{A}$ by the distribution rule and the detachment rule. Thus in this particular case, the hypothesis of the least number principle is verified.

3.2 The theorem on sequences. We define $\mathrm{RP}\,xy \leftrightarrow \forall z(\mathrm{Div}(x \cdot z)y \rightarrow \mathrm{Div}\,zy)$, $\mathrm{OP}\,xy = ((x + y) \cdot (x + y)) + \mathrm{S}x$, and

$$\mathrm{B}xy = \mu z(z = x - \dot{1} \vee \exists x' \exists y'(x' < x \wedge y' < x \wedge x = \mathrm{OP}\,x'y' \wedge \mathrm{Div}\,x'\mathrm{S}(\mathrm{S}\,\mathrm{OP}\,zy \cdot y'))).$$

Note that $\vdash_{\mathrm{PA}} x < \mathrm{OP}\,xy \wedge y < \mathrm{OP}\,xy$, $\vdash_{\mathrm{PA}} \mathrm{OP}\,xy \neq \dot{0}$, and $\vdash_{\mathrm{PA}} \mathrm{B}xy \leq x - \dot{1}$. It is clear that OP and B are recursive on $L(\mathrm{N})$. By the theorem on RE-formulae, we see that B represents in PA the coding function β of ch. III §1.2.

LEMMA 1. $\vdash_{\mathrm{PA}} y \neq \dot{0} \rightarrow \mathrm{RP}\,xy \rightarrow \mathrm{RP}\,yx$.

Proof. Let P be the first-order theory obtained from PA by the adjunction of new constants $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ and the axioms $\mathbf{e}_2 \neq \dot{0}$, $\mathrm{RP}\,\mathbf{e}_1\mathbf{e}_2$, and $\mathrm{Div}(\mathbf{e}_2 \cdot \mathbf{e}_3)\mathbf{e}_1$. By the deduction theorem and the definition of RP, it will suffice to prove $\vdash_P \mathrm{Div}\,\mathbf{e}_3\mathbf{e}_1$. By the substitution axioms, the definition of Div, and the symmetry theorem, $\vdash_P \mathbf{e}_2 \cdot \mathbf{e}_3 = \mathbf{e}_1 \cdot z \rightarrow \mathrm{Div}(\mathbf{e}_1 \cdot z)\mathbf{e}_2$, whence

$$\vdash_P \mathbf{e}_2 \cdot \mathbf{e}_3 = \mathbf{e}_1 \cdot z \rightarrow \mathrm{Div}\,z\mathbf{e}_2 \tag{5}$$

by definition of RP and the new axioms. By the equality theorem, $\vdash_P \mathbf{e}_2 \cdot \mathbf{e}_3 = \mathbf{e}_1 \cdot z \wedge z = \mathbf{e}_2 \cdot z' \rightarrow \mathbf{e}_2 \cdot \mathbf{e}_3 = \mathbf{e}_1 \cdot (\mathbf{e}_2 \cdot z')$, whence $\vdash_P \mathbf{e}_2 \cdot \mathbf{e}_3 = \mathbf{e}_1 \cdot z \wedge z = \mathbf{e}_2 \cdot z' \rightarrow \mathbf{e}_2 \cdot \mathbf{e}_3 = (\mathbf{e}_2 \cdot \mathbf{e}_1) \cdot z'$ by properties of $\cdot$. From this by the distribution rule and prenex operations, $\vdash_P \mathbf{e}_2 \cdot \mathbf{e}_3 = \mathbf{e}_1 \cdot z \wedge \mathrm{Div}\,z\mathbf{e}_2 \rightarrow \mathrm{Div}(\mathbf{e}_2 \cdot \mathbf{e}_3)(\mathbf{e}_2 \cdot \mathbf{e}_1)$. Using (5) and the $\exists$-introduction rule, $\vdash_P \mathrm{Div}(\mathbf{e}_2 \cdot \mathbf{e}_3)\mathbf{e}_1 \rightarrow \mathrm{Div}(\mathbf{e}_2 \cdot \mathbf{e}_3)(\mathbf{e}_2 \cdot \mathbf{e}_1)$, and finally $\vdash_P \mathrm{Div}(\mathbf{e}_2 \cdot \mathbf{e}_3)\mathbf{e}_1 \rightarrow \mathrm{Div}\,\mathbf{e}_3\mathbf{e}_1$ because $\vdash_P \mathbf{e}_2 \neq \dot{0}$. By the axioms of P, $\vdash_P \mathrm{Div}\,\mathbf{e}_3\mathbf{e}_1$. □

LEMMA 2. $\vdash_{\mathrm{PA}} \mathrm{OP}\,xy = \mathrm{OP}\,x'y' \leftrightarrow x = x' \wedge y = y'$.

Proof. The implication from right to left is tautologically equivalent to an equality axiom. To derive the other implication, we shall first prove

$$\vdash_{\mathrm{PA}} x + y < x' + y' \rightarrow \mathrm{OP}\,xy < \mathrm{OP}\,x'y' \text{ and} \tag{6}$$

$$\vdash_{\mathrm{PA}} x' + y' < x + y \rightarrow \mathrm{OP}\,x'y' < \mathrm{OP}\,xy. \tag{7}$$

We have $\vdash_{\mathrm{PA}} \mathrm{S}(x+y) \cdot \mathrm{S}(x+y) = ((x+y) \cdot (x+y)) + (\mathrm{S}(x+y) + (x+y))$, and since $\vdash_{\mathrm{PA}} \mathrm{S}x \leq \mathrm{S}(x+y) + (x+y)$, we find

$$\vdash_{\mathrm{PA}} \mathrm{OP}\,xy \leq \mathrm{S}(x + y) \cdot \mathrm{S}(x + y). \tag{8}$$

On the other hand we can easily derive

$$\vdash_{\mathrm{PA}} (x' + y') \cdot (x' + y') < \mathrm{OP}\,x'y'. \tag{9}$$

Now $\vdash_{\mathrm{PA}} x + y < x' + y' \rightarrow \mathrm{S}(x+y) \leq x' + y'$. Combining this with (8) and (9) we obtain (6). The derivation of (7) is similar. From (6) and (7), we obtain

$$\vdash_{\mathrm{PA}} \mathrm{OP}\,xy = \mathrm{OP}\,x'y' \rightarrow x + y = x' + y' \tag{10}$$

whence $\vdash_{\mathrm{PA}} \mathrm{OP}\,xy = \mathrm{OP}\,x'y' \rightarrow (x + y) \cdot (x + y) = (x' + y') \cdot (x' + y')$. From this by definition of OP and properties of $+$, $\vdash_{\mathrm{PA}} \mathrm{OP}\,xy = \mathrm{OP}\,x'y' \rightarrow \mathrm{S}x = \mathrm{S}x'$, and hence $\vdash_{\mathrm{PA}} \mathrm{OP}\,xy = \mathrm{OP}\,x'y' \rightarrow x = x'$ by N2. Together with (10) and properties of $+$, this implies $\vdash_{\mathrm{PA}} \mathrm{OP}\,xy = \mathrm{OP}\,x'y' \rightarrow y = y'$. □

LEMMA 3. $\vdash_{\mathrm{PA}} \mathrm{Div}\, xy \to \mathrm{RP}\, \mathrm{S}((z+y)\cdot x)\mathrm{S}(z\cdot x)$.

Proof. Let P be obtained from PA by the adjunction of new constants $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$, and $\mathbf{e}_4$, and the axioms $\mathrm{Div}\, \mathbf{e}_1\mathbf{e}_2$ and $\mathrm{Div}(\mathrm{S}((\mathbf{e}_3+\mathbf{e}_2)\cdot\mathbf{e}_1)\cdot\mathbf{e}_4)\mathrm{S}(\mathbf{e}_3\cdot\mathbf{e}_1)$ (henceforth referred to as first and second axiom). By the deduction theorem and the definition of RP, the lemma is reduced to

$$\vdash_P \mathrm{Div}\, \mathbf{e}_4\mathrm{S}(\mathbf{e}_3\cdot\mathbf{e}_1). \tag{11}$$

We shall first prove

$$\vdash_{\mathrm{PA}} \mathrm{RP}\, x\mathrm{S}(z\cdot x). \tag{12}$$

We know that $\vdash_{\mathrm{PA}} \mathrm{Div}((w\cdot z)\cdot x)x$ and $\vdash_{\mathrm{PA}} \mathrm{Div}((w\cdot z)\cdot x)x \to \mathrm{Div}(w+((w\cdot z)\cdot x))x \to \mathrm{Div}\, wx$, whence $\vdash_{\mathrm{PA}} \mathrm{Div}(w+((w\cdot z)\cdot x))x \to \mathrm{Div}\, wx$. Since $\vdash_{\mathrm{PA}} \mathrm{S}(z\cdot x)\cdot w = w+((w\cdot z)\cdot x)$, we obtain (12).

We have $\vdash_P \mathrm{S}((\mathbf{e}_3+\mathbf{e}_2)\cdot\mathbf{e}_1)\cdot\mathbf{e}_4 = (\mathrm{S}(\mathbf{e}_3\cdot\mathbf{e}_1)\cdot\mathbf{e}_4)+((\mathbf{e}_2\cdot\mathbf{e}_1)\cdot\mathbf{e}_4)$. Using $\vdash_P \mathrm{Div}(\mathrm{S}(\mathbf{e}_3\cdot\mathbf{e}_1)\cdot\mathbf{e}_4)\mathrm{S}(\mathbf{e}_3\cdot\mathbf{e}_1)$ and the second axiom, we infer $\vdash_P \mathrm{Div}((\mathbf{e}_2\cdot\mathbf{e}_1)\cdot\mathbf{e}_4)\mathrm{S}(\mathbf{e}_3\cdot\mathbf{e}_1)$. By an instance of (12) and properties of $\cdot$, we obtain

$$\vdash_P \mathrm{Div}(\mathbf{e}_4\cdot\mathbf{e}_2)\mathrm{S}(\mathbf{e}_3\cdot\mathbf{e}_1) \tag{13}$$

By the first axiom and properties of Div, $\vdash_P \mathrm{Div}(\mathbf{e}_4\cdot\mathbf{e}_1)(\mathbf{e}_4\cdot\mathbf{e}_2)$. By (13) and the transitivity of Div, $\vdash_P \mathrm{Div}(\mathbf{e}_4\cdot\mathbf{e}_1)\mathrm{S}(\mathbf{e}_3\cdot\mathbf{e}_1)$, and hence (11) by an instance of (12) and the substitution theorem. □

LEMMA 4. Let P be a good extension of PA, $\mathbf{a}$ a term of P. If $\mathbf{x}$ and $\mathbf{z}$ do not occur in $\mathbf{a}$, then $\vdash_P \exists\mathbf{x}\forall\mathbf{y}(\mathbf{y}<\mathbf{z}\to\mathbf{a}<\mathbf{x})$.

Proof. Let $\mathbf{A}$ be the formula to be derived. We have $\vdash_P \mathbf{A}[\mathbf{z}|\dot{0}]$ and $\vdash_P \forall\mathbf{y}(\mathbf{y}<\mathbf{z}\to\mathbf{a}<\mathbf{x})\to\forall\mathbf{y}(\mathbf{y}<\mathrm{S}\mathbf{z}\to\mathbf{a}<\mathrm{S}\mathbf{x}+\mathbf{a}[\mathbf{y}|\mathbf{z}])$, whence $\vdash_P \mathbf{A}\to\mathbf{A}[\mathbf{z}|\mathrm{S}\mathbf{z}]$ by the substitution axioms and the $\exists$-introduction rule. By the induction axioms, $\vdash_P \mathbf{A}$. □

LEMMA 5. Let P be a good extension of PA, $\mathbf{A}$ and $\mathbf{B}$ formulae of P, and $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ variables such that $\mathbf{y}$, $\mathbf{z}$ are not free in $\mathbf{A}$ and $\mathbf{x}$, $\mathbf{z}$ are not free in $\mathbf{B}$. Suppose that

(i) $\vdash_P \exists\mathbf{z}(\mathbf{A}\to\mathbf{x}<\mathbf{z})$;
(ii) $\vdash_P \neg\mathbf{B}[\mathbf{y}|\dot{1}]$; and
(iii) $\vdash_P \mathbf{A}\to\mathbf{B}\to\mathrm{RP}\,\mathbf{xy}$.

Then $\vdash_P \exists\mathbf{z}(\forall\mathbf{x}(\mathbf{A}\to\mathrm{Div}\,\mathbf{zx})\wedge\forall\mathbf{y}(\mathbf{B}\to\neg\mathrm{Div}\,\mathbf{zy}))$.

Proof. Let $\mathbf{w}$ be distinct from $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ and not occurring in $\mathbf{A}$. We first derive in P the formula $\exists\mathbf{z}\mathbf{C}$, where $\mathbf{C}$ is

$$\forall\mathbf{x}(\mathbf{x}<\mathbf{w}\to\mathbf{A}\to\mathrm{Div}\,\mathbf{zx})\wedge\forall\mathbf{y}(\mathbf{B}\to\neg\mathrm{Div}\,\mathbf{zy}),$$

using an induction axiom. We have $\vdash_P \mathbf{B}\to\mathbf{y}\neq\dot{1}$ by (ii), whence $\vdash_P \mathbf{B}\to\neg\mathrm{Div}\,\dot{1}\mathbf{y}$ by properties of Div. By N7 and the tautology theorem, $\vdash_P \mathbf{x}<\dot{0}\to\mathbf{A}\to\mathrm{Div}\,\dot{1}\mathbf{x}$. From the last two formulae, we find

$$\vdash_P \exists\mathbf{z}\mathbf{C}[\mathbf{w}|\dot{0}]. \tag{14}$$

We claim that

$$\vdash_P \mathbf{C}\to\mathbf{A}[\mathbf{x}|\mathbf{w}]\to\mathbf{x}<\mathbf{w}\to\mathbf{A}\to\mathrm{Div}(\mathbf{z}\cdot\mathbf{w})\mathbf{x}, \tag{15}$$

$$\vdash_P \mathbf{C}\to\mathbf{A}[\mathbf{x}|\mathbf{w}]\to\mathbf{x}=\mathbf{w}\to\mathbf{A}\to\mathrm{Div}(\mathbf{z}\cdot\mathbf{w})\mathbf{x}, \tag{16}$$

$$\vdash_P \mathbf{C}\to\mathbf{A}[\mathbf{x}|\mathbf{w}]\to\mathbf{B}\to\neg\mathrm{Div}(\mathbf{z}\cdot\mathbf{w})\mathbf{y}, \tag{17}$$

$$\vdash_P \mathbf{C}\to\neg\mathbf{A}[\mathbf{x}|\mathbf{w}]\to\mathbf{x}<\mathbf{w}\to\mathbf{A}\to\mathrm{Div}\,\mathbf{zx}, \tag{18}$$

$$\vdash_P \mathbf{C}\to\neg\mathbf{A}[\mathbf{x}|\mathbf{w}]\to\mathbf{x}=\mathbf{w}\to\mathbf{A}\to\mathrm{Div}\,\mathbf{zx}, \tag{19}$$

$$\vdash_P \mathbf{C}\to\neg\mathbf{A}[\mathbf{x}|\mathbf{w}]\to\mathbf{B}\to\neg\mathrm{Div}\,\mathbf{zy}. \tag{20}$$

To obtain (15), apply the substitution theorem to the left-hand side of $\mathbf{C}$, and use $\vdash_P \mathrm{Div}\,\mathbf{zx}\to\mathrm{Div}(\mathbf{z}\cdot\mathbf{w})\mathbf{x}$. To obtain (16), use $\vdash_P \mathrm{Div}(\mathbf{z}\cdot\mathbf{x})\mathbf{x}$ and the equality theorem. By hypothesis (iii) and the defining axiom of RP, $\vdash_P \mathbf{A}[\mathbf{x}|\mathbf{w}]\to\mathbf{B}\to\neg\mathrm{Div}\,\mathbf{zy}\to\neg\mathrm{Div}(\mathbf{z}\cdot\mathbf{w})\mathbf{y}$; (17) is a tautological consequence of the latter and $\vdash_P \mathbf{C}\to\mathbf{B}\to\neg\mathrm{Div}\,\mathbf{zy}$. Finally, (18) and (20) are straightforward and (19) is obtained from the tautology $\neg\mathbf{A}[\mathbf{x}|\mathbf{w}]\to$

$\mathbf{A}[\mathbf{x}|\mathbf{w}] \to \mathrm{Div}\,\mathbf{zw}$ using the equality theorem. The following formulae are tautological consequences of an instance of N8 and (15)–(16) (resp. (18)–(19)):

$$\vdash_P \mathbf{C} \to \mathbf{A}[\mathbf{x}|\mathbf{w}] \to \mathbf{x} < \mathrm{S}\mathbf{w} \to \mathbf{A} \to \mathrm{Div}(\mathbf{z}\cdot\mathbf{w})\mathbf{x}, \tag{21}$$

$$\vdash_P \mathbf{C} \to \neg\mathbf{A}[\mathbf{x}|\mathbf{w}] \to \mathbf{x} < \mathrm{S}\mathbf{w} \to \mathbf{A} \to \mathrm{Div}\,\mathbf{zx}. \tag{22}$$

From (17) and (21) by the $\forall$-introduction rule, we obtain $\vdash_P \mathbf{C} \to \mathbf{A}[\mathbf{x}|\mathbf{w}] \to \mathbf{C}[\mathbf{z},\mathbf{w}|\mathbf{z}\cdot\mathbf{w},\mathrm{S}\mathbf{w}]$ and similarly from (20) and (22), $\vdash_P \mathbf{C} \to \neg\mathbf{A}[\mathbf{x}|\mathbf{w}] \to \mathbf{C}[\mathbf{z},\mathbf{w}|\mathbf{z}\cdot\mathbf{w},\mathrm{S}\mathbf{w}]$. Using the substitution axioms, these become $\vdash_P \mathbf{C} \to \mathbf{A}[\mathbf{x}|\mathbf{w}] \to \exists\mathbf{z}\mathbf{C}[\mathbf{w}|\mathrm{S}\mathbf{w}]$ and $\vdash_P \mathbf{C} \to \neg\mathbf{A}[\mathbf{x}|\mathbf{w}] \to \exists\mathbf{z}\mathbf{C}[\mathbf{w}|\mathrm{S}\mathbf{w}]$. Combining these two formulae with the tautology theorem and using the $\exists$-introduction rule, we obtain

$$\vdash_P \exists\mathbf{z}\mathbf{C} \to \exists\mathbf{z}\mathbf{C}[\mathbf{w}|\mathrm{S}\mathbf{w}]. \tag{23}$$

From (14) and (23) by the induction axioms, we find $\vdash_P \exists\mathbf{z}\mathbf{C}$. We let $\mathbf{D}$ be the formula $(\mathbf{x} < \mathbf{w} \to \mathbf{A} \to \mathrm{Div}\,\mathbf{zx}) \wedge (\mathbf{B} \to \neg\,\mathrm{Div}\,\mathbf{zy})$, and $\mathbf{D}'$ the formula $(\mathbf{A} \to \mathrm{Div}\,\mathbf{zx}) \wedge (\mathbf{B} \to \neg\,\mathrm{Div}\,\mathbf{zy})$. Then $(\mathbf{A} \to \mathbf{x} < \mathbf{w}) \wedge \mathbf{D} \to \mathbf{D}'$ is a tautology. Using the distribution rule thrice and the $\exists$-introduction rule, we obtain $\vdash_P \exists\mathbf{w}\exists\mathbf{z}\forall\mathbf{x}\forall\mathbf{y}((\mathbf{A} \to \mathbf{x} < \mathbf{w}) \wedge \mathbf{D}) \to \exists\mathbf{z}\forall\mathbf{x}\forall\mathbf{y}\mathbf{D}'$, whence $\exists\mathbf{w}(\mathbf{A} \to \mathbf{x} < \mathbf{w}) \wedge \exists\mathbf{z}\mathbf{C} \to \exists\mathbf{z}(\forall\mathbf{x}(\mathbf{A} \to \mathrm{Div}\,\mathbf{zx}) \wedge \forall\mathbf{y}(\mathbf{B} \to \neg\,\mathrm{Div}\,\mathbf{zy}))$ by prenex operations. Using the hypothesis (i) and our previous result, we obtain $\vdash_P \exists\mathbf{z}(\forall\mathbf{x}(\mathbf{A} \to \mathrm{Div}\,\mathbf{zx}) \wedge \forall\mathbf{y}(\mathbf{B} \to \neg\,\mathrm{Div}\,\mathbf{zy}))$ as desired. □

We are now able to prove the main result of this section.

Theorem on sequences. Let P be a good extension of PA, $\mathbf{a}$ a term of P, and $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ variables such that $\mathbf{x}$ and $\mathbf{z}$ do not occur in $\mathbf{a}$. Then $\vdash_P \exists\mathbf{x}\forall\mathbf{y}(\mathbf{y} < \mathbf{z} \to \mathrm{B}\mathbf{xy} = \mathbf{a})$.

Proof. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be the variables other than $\mathbf{y}$ occurring in $\mathbf{a}$, and let $\mathbf{x}'$, $\mathbf{w}$, $\mathbf{w}'$ be variables distinct from $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ and not occuring in $\mathbf{a}$. By lemma 4, $\vdash_P \exists\mathbf{x}\forall\mathbf{y}(\mathbf{y} < \mathbf{z} \to \mathrm{S\,OP}\,\mathbf{ay} < \mathbf{x})$. Define $\mathbf{fzx}_1\ldots\mathbf{x}_n = \mu\mathbf{x}\forall\mathbf{y}(\mathbf{y} < \mathbf{z} \to \mathrm{S\,OP}\,\mathbf{ay} < \mathbf{x})$. The hypotheses of lemma 5 are clearly satisfied when $\mathbf{A}$ and $\mathbf{B}$ are $\mathbf{x} < \mathbf{fzx}_1\ldots\mathbf{x}_n \wedge \mathbf{x} \neq \dot{0}$ and $\mathbf{y} = \dot{0}$. Hence we can define $\mathbf{g}$ by $\mathbf{gzx}_1\ldots\mathbf{x}_n = \mu\mathbf{w}(\forall\mathbf{x}(\mathbf{x} < \mathbf{fzx}_1\ldots\mathbf{x}_n \wedge \mathbf{x} \neq \dot{0} \to \mathrm{Div}\,\mathbf{wx}) \wedge \forall\mathbf{y}(\mathbf{y} = \dot{0} \to \neg\,\mathrm{Div}\,\mathbf{wy}))$. We denote by $\mathbf{b}$ the term $\mathbf{gzx}_1\ldots\mathbf{x}_n$. Note that since $\vdash_P \mathrm{Div}\,\dot{0}\dot{0}$,

$$\vdash_P \mathbf{b} \neq \dot{0}. \tag{24}$$

Using $\vdash_P z < y \to z + (y - z) = y$, $\vdash_P z < y \to y - z \neq \dot{0}$, and lemma 3, we obtain $\vdash_P z < y \to x \neq \dot{0} \to \mathrm{Div}\,x(y - z) \to \mathrm{RP\,S}(y\cdot x)\mathrm{S}(z\cdot x)$, an instance of which is $\vdash_P z < y \to \mathbf{g}wx_1\ldots x_n \neq \dot{0} \to \mathrm{Div}\,\mathbf{g}wx_1\ldots x_n(y - z) \to \mathrm{RP\,S}(y\cdot\mathbf{g}wx_1\ldots x_n)\mathrm{S}(z\cdot\mathbf{g}wx_1\ldots x_n)$. But the definition of $\mathbf{g}$ is such that $\vdash_P z < y \to y < \mathbf{f}wx_1\ldots x_n \to \mathrm{Div}\,\mathbf{g}wx_1\ldots x_n(y - z)$, and so, using (24),

$$\vdash_P z < y \to y < \mathbf{f}wx_1\ldots x_n \to \mathrm{RP\,S}(y\cdot\mathbf{g}wx_1\ldots x_n)\mathrm{S}(z\cdot\mathbf{g}wx_1\ldots x_n). \tag{25}$$

Inverting the rôles of y and z in the above formula and using lemma 1, we obtain

$$\vdash_P y < z \to z < \mathbf{f}wx_1\ldots x_n \to \mathrm{RP\,S}(y\cdot\mathbf{g}wx_1\ldots x_n)\mathrm{S}(z\cdot\mathbf{g}wx_1\ldots x_n). \tag{26}$$

From (25) and (26),

$$\vdash_P y \neq z \to y < \mathbf{f}wx_1\ldots x_n \to z < \mathbf{f}wx_1\ldots x_n \to \mathrm{RP\,S}(y\cdot\mathbf{g}wx_1\ldots x_n)\mathrm{S}(z\cdot\mathbf{g}wx_1\ldots x_n). \tag{27}$$

We let $\mathbf{A}$ be $\exists\mathbf{y}(\mathbf{y} < \mathbf{z} \wedge \mathbf{w} = \mathrm{S}(\mathrm{S\,OP}\,\mathbf{ay}\cdot\mathbf{b}))$ and $\mathbf{B}$ be $\exists\mathbf{x}(\mathbf{x} < \mathbf{fzx}_1\ldots\mathbf{x}_n \wedge \dot{1} < \mathbf{x} \wedge \forall\mathbf{y}(\mathbf{y} < \mathbf{z} \to \mathbf{x} \neq \mathrm{S\,OP}\,\mathbf{ay}) \wedge \mathbf{w}' = \mathrm{S}(\mathbf{x}\cdot\mathbf{b})) \vee \mathbf{w}' = \dot{0}$. We claim that

$$\vdash_P \exists\mathbf{x}(\mathbf{A} \to \mathbf{w} < \mathbf{x}), \tag{28}$$

$$\vdash_P \neg\mathbf{B}[\mathbf{w}'|\dot{1}], \text{ and} \tag{29}$$

$$\vdash_P \mathbf{A} \to \mathbf{B} \to \mathrm{RP}\,\mathbf{ww}'. \tag{30}$$

Now by the definition of $\mathbf{f}$ and (24), $\vdash_P \mathbf{A} \to \mathbf{w} < \mathrm{S}(\mathbf{fzx}_1\ldots\mathbf{x}_n\cdot\mathbf{b})$, and thus (28) holds. Also by (24), $\vdash_P \dot{1} < \mathbf{x} \to \mathrm{S}(\mathbf{x}\cdot\mathbf{b}) \neq \dot{1}$ and hence (29) holds. By the definition of $\mathbf{f}$, $\vdash_P \mathbf{y} < \mathbf{z} \to \mathrm{S\,OP}\,\mathbf{ay} < \mathbf{fzx}_1\ldots\mathbf{x}_n$ and by the substitution theorem, $\vdash_P \mathbf{y} < \mathbf{z} \to \forall\mathbf{y}(\mathbf{y} < \mathbf{z} \to \mathbf{x} \neq \mathrm{S\,OP}\,\mathbf{ay}) \to \mathbf{x} \neq \mathrm{S\,OP}\,\mathbf{ay}$. From these and (27) we obtain $\vdash_P \mathbf{y} < \mathbf{z} \to \mathbf{x} < \mathbf{fzx}_1\ldots\mathbf{x}_n \to \forall\mathbf{y}(\mathbf{y} < \mathbf{z} \to \mathbf{x} \neq \mathrm{S\,OP}\,\mathbf{ay}) \to \mathbf{w} = \mathrm{S}(\mathrm{S\,OP}\,\mathbf{ay}\cdot\mathbf{b}) \to \mathbf{w}' = \mathrm{S}(\mathbf{x}\cdot\mathbf{b}) \to \mathrm{RP}\,\mathbf{ww}'$. By the tautology theorem, the ∃-introduction rule, and prenex operations, we obtain (30). Thus by lemma 5, we may define $\mathbf{h}$ by $\mathbf{hzx}_1 \dots \mathbf{x}_n = \mu\mathbf{x}'(\forall\mathbf{w}(\mathbf{A} \to \mathrm{Div}\,\mathbf{x}'\mathbf{w}) \wedge \forall\mathbf{w}'(\mathbf{B} \to \neg\,\mathrm{Div}\,\mathbf{x}'\mathbf{w}'))$. We let $\mathbf{c}$ be $\mathbf{hzx}_1 \dots \mathbf{x}_n$.

We now show that $\vdash_P \mathbf{y} < \mathbf{z} \to \mathrm{B}\,\mathrm{OP}\,\mathbf{cby} = \mathbf{a}$. By the generalization rule and the substitution axioms, this will complete the proof of the theorem. By the definition of B, it will suffice to prove

$$\vdash_P \mathbf{y} < \mathbf{z} \to \mathbf{c} < \mathrm{OP}\,\mathbf{cb} \wedge \mathbf{b} < \mathrm{OP}\,\mathbf{cb} \wedge \mathrm{Div}\,\mathbf{c}\mathrm{S}(\mathrm{S}\,\mathrm{OP}\,\mathbf{ay}\cdot\mathbf{b}), \tag{31}$$

$$\vdash_P \mathbf{y} < \mathbf{z} \to \mathbf{x} < \mathbf{a} \to \mathbf{x} \neq \mathrm{OP}\,\mathbf{cb} - \dot{1}, \text{ and} \tag{32}$$

$$\vdash_P \mathbf{y} < \mathbf{z} \to \mathbf{x} < \mathbf{a} \to \mathbf{w} < \mathrm{OP}\,\mathbf{cb} \to \mathbf{w}' < \mathrm{OP}\,\mathbf{cb} \to \mathrm{OP}\,\mathbf{cb} \neq \mathrm{OP}\,\mathbf{ww}' \vee \neg\,\mathrm{Div}\,\mathbf{w}\mathrm{S}(\mathrm{S}\,\mathrm{OP}\,\mathbf{xy}\cdot\mathbf{w}')). \tag{33}$$

Now (31) follows at once from the definition of $\mathbf{h}$ and the fact that $\vdash_P x < \mathrm{OP}\,xy \wedge y < \mathrm{OP}\,xy$. To prove (32), since $\vdash_P \mathrm{OP}\,xy \neq \dot{0}$ and $\vdash_P y < \mathrm{OP}\,xy$, it will suffice to prove

$$\vdash_P \mathbf{y} < \mathbf{z} \to \mathbf{a} < \mathbf{b} \tag{34}$$

But by definition of $\mathbf{g}$, $\vdash_P \mathbf{y} < \mathbf{z} \to \mathrm{Div}\,\mathbf{b}\mathrm{S}\,\mathrm{OP}\,\mathbf{ay}$, and so by (24), $\vdash_P \mathbf{y} < \mathbf{z} \to \mathrm{S}\,\mathrm{OP}\,\mathbf{ay} < \mathbf{b}$, whence (34). By lemma 2, $\vdash_P \mathbf{w} \neq \mathbf{c} \vee \mathbf{w}' \neq \mathbf{b} \to \mathrm{OP}\,\mathbf{cb} \neq \mathrm{OP}\,\mathbf{ww}'$, and so to prove (33) we need only prove $\vdash_P \mathbf{y} < \mathbf{z} \to \mathbf{x} < \mathbf{a} \to \neg\,\mathrm{Div}\,\mathbf{c}\mathrm{S}(\mathrm{S}\,\mathrm{OP}\,\mathbf{xy}\cdot\mathbf{b})$. Given the definition of $\mathbf{h}$, it suffices to prove

$$\vdash_P \mathbf{y} < \mathbf{z} \to \mathbf{x} < \mathbf{a} \to \mathrm{S}\,\mathrm{OP}\,\mathbf{xy} < \mathbf{fzx}_1 \dots \mathbf{x}_n \wedge \dot{1} < \mathrm{S}\,\mathrm{OP}\,\mathbf{xy} \wedge \forall\mathbf{y}(\mathbf{y} < \mathbf{z} \to \mathrm{S}\,\mathrm{OP}\,\mathbf{xy} \neq \mathrm{S}\,\mathrm{OP}\,\mathbf{ay}).$$

This easily follows from the definition of $\mathbf{f}$ and the properties of OP. □

3.3 Coding function symbols. A binary function symbol $\mathbf{f}$ of an extension P of PA will be called a *coding function symbol in P* if $\vdash_P \mathbf{f}xy \leq x - \dot{1}$, and if whenever $\mathbf{a}$ is a term of an extension by definitions P' of P in which $\mathbf{x}$ and $\mathbf{z}$ do not occur, $\vdash_{P'} \exists\mathbf{x}\forall\mathbf{y}(\mathbf{y} < \mathbf{z} \to \mathbf{fxy} = \mathbf{a})$. We have just proved that there exists a coding function symbol in any good extension of PA which is recursive on $L(\mathrm{N})$ in PA, namely B. *From now on we fix a good extension P of* PA *and a coding function symbol* B *in P.*

In the sequel, applications of the theorem on sequences will appear as definitions of the form $\mathbf{fzx}_1 \dots \mathbf{x}_n = \mu\mathbf{x}\mathbf{A}$ where $\vdash_P \mathbf{A} \leftrightarrow \forall\mathbf{y}(\mathbf{y} < \mathbf{z} \to \mathrm{B}\mathbf{xy} = \mathbf{gx}_1 \dots \mathbf{x}_n)$ for some defined function symbol $\mathbf{g}$. It will usually be obvious how to define $\mathbf{g}$ suitably.

We abbreviate $\mathrm{B}\mathbf{a}\mathrm{S}\mathbf{b}$ by $(\mathbf{a})_{\mathbf{b}}$, and $((\mathbf{a})_{\mathbf{b}})_{\mathbf{c}}$ by $(\mathbf{a})_{\mathbf{b},\mathbf{c}}$. If n is a natural number, we also write $(\mathbf{a})_n$ instead of $(\mathbf{a})_{\dot{n}}$. In the same way that coding functions were used to define recursive n-ary sequence functions, we can use coding function symbols to discuss sequences of numbers in PA. We now introduce defined symbols for this purpose.

(i) $\mathrm{Len}\,x = \mathrm{B}x\dot{0}$;

(ii) $\mathrm{Sq}\,x \leftrightarrow \forall y(y < x \to \exists z(z < \mathrm{S}\,\mathrm{Len}\,x \wedge \mathrm{B}xz \neq \mathrm{B}yz))$;

(iii) $x \in y \leftrightarrow \mathrm{Sq}\,y \wedge \exists z(z < \mathrm{Len}\,y \wedge x = (y)_z)$;

(iv) $*xy = \mu z(\mathrm{Len}\,z = \mathrm{Len}\,x + \mathrm{Len}\,y \wedge \forall w(w < \mathrm{Len}\,x \to (z)_w = (x)_w) \wedge \forall w(w < \mathrm{Len}\,y \to (z)_{w+\mathrm{Len}\,x} = (y)_w))$;

(v) $\mathrm{Ini}\,xy = \mu z(\mathrm{Len}\,z = y \wedge \forall w(w < y \to (z)_w = (x)_w))$;

(vi) $z = \mathrm{Rmv}\,xy \leftrightarrow (y \neq \dot{0} \wedge \mathrm{Sq}\,z \wedge \mathrm{Len}\,z = \mathrm{Len}\,x - \dot{1} \wedge \forall w(w < \mathrm{Len}\,x - \dot{1} \to w < y \to (z)_w = (x)_w) \wedge \forall w(w < \mathrm{Len}\,x - \dot{1} \to y \leq w \to (z)_w = (x)_{\mathrm{S}w})) \vee (y = \dot{0} \wedge z = x)$.

We abbreviate $*\mathbf{ab}$ by $(\mathbf{a} * \mathbf{b})$. Observe that Len, Sq, ∈, ∗, Ini, and Rmv are recursive on $L(\mathrm{N})$ with B. The fundamental property of Sq is the theorem

$$\vdash_P \mathrm{Sq}\,x \to \mathrm{Sq}\,y \to \forall z(z < \mathrm{S}\,\mathrm{Len}\,x \to \mathrm{B}xz = \mathrm{B}yz) \to x = y \tag{35}$$

which we now derive. We have $\vdash_P \mathrm{Sq}\,x \to y < x \to \exists z(z < \mathrm{S}\,\mathrm{Len}\,x \wedge \mathrm{B}xz \neq \mathrm{B}yz)$ by definition of Sq. By rudimentary operations, $\vdash_P \neg\exists z(z < \mathrm{S}\,\mathrm{Len}\,x \wedge \mathrm{B}xz \neq \mathrm{B}yz) \leftrightarrow \forall z(z < \mathrm{S}\,\mathrm{Len}\,x \to \mathrm{B}xz = \mathrm{B}yz)$. From these by the tautology theorem,

$$\vdash_P \mathrm{Sq}\,x \to \forall z(z < \mathrm{S}\,\mathrm{Len}\,x \to \mathrm{B}xz = \mathrm{B}yz) \to \neg(y < x), \tag{36}$$

whence by the substitution rule and the symmetry theorem,

$$\vdash_P \mathrm{Sq}\,y \to \forall z(z < \mathrm{S}\,\mathrm{Len}\,y \to \mathrm{B}xz = \mathrm{B}yz) \to \neg(x < y). \tag{37}$$

By the substitution theorem and the definition of Len, $\vdash_P \forall z(z < \mathrm{S\,Len}\,x \to \mathrm{B}xz = \mathrm{B}yz) \to \dot{0} < \mathrm{S\,Len}\,x \to \mathrm{Len}\,x = \mathrm{Len}\,y$. Since $\vdash_P \dot{0} < \mathrm{S\,Len}\,x$, we obtain $\vdash_P \forall z(z < \mathrm{S\,Len}\,x \to \mathrm{B}xz = \mathrm{B}yz) \to \mathrm{Len}\,x = \mathrm{Len}\,y$, whence

$$\vdash_P \forall z(z < \mathrm{S\,Len}\,x \to \mathrm{B}xz = \mathrm{B}yz) \to \forall z(z < \mathrm{S\,Len}\,y \to \mathrm{B}xz = \mathrm{B}yz) \tag{38}$$

by the equality theorem. From (36), (37), (38), and N9 by the tautology theorem, we obtain (35).

Other properties of these symbols are listed below. All of them are immediate consequences of the definitions.

(i) $\vdash_P \mathrm{Len}(x * y) = \mathrm{Len}\,x + \mathrm{Len}\,y$;

(ii) $\vdash_P \mathrm{Sq}(x * y)$;

(iii) $\vdash_P x * (y * z) = (x * y) * z$;

(iv) $\vdash_P \mathrm{Sq}\,y \to \mathrm{Sq}\,z \to x \in y * z \leftrightarrow x \in y \vee x \in z$;

(v) $\vdash_P \mathrm{Len\,Ini}\,xy = y$;

(vi) $\vdash_P \mathrm{Sq\,Ini}\,xy$;

(vii) $\vdash_P y = \dot{0} \to \mathrm{Rmv}\,xy = x$;

(viii) $\vdash_P y \neq \dot{0} \to \mathrm{Len\,Rmv}\,xy = \mathrm{Len}\,x - \dot{1}$;

(ix) $\vdash_P y \neq \dot{0} \to \mathrm{Sq\,Rmv}\,xy$.

(x) $\vdash_P \mathrm{Sq}\,x \to \mathrm{Len}\,x = \dot{0} \to x = \dot{0}$;

The proof of (x) uses the fact that $\vdash_P \mathrm{B}xy \leq x - \dot{1}$.

3.4 Sequences in PA. We now show how sequences can be defined in P. Let $\mathbf{a}$ be a term of P and $\mathbf{x}, \mathbf{x}_1, \ldots, \mathbf{x}_n$, $\mathbf{y}$, $\mathbf{z}$, and $\mathbf{w}$ distinct variables such that $\mathbf{x}_1, \ldots, \mathbf{x}_n$, and $\mathbf{y}$ include the variables occurring in $\mathbf{a}$. Define a function symbol $\mathbf{g}$ by $\mathbf{z} = \mathbf{gywx}_1 \ldots \mathbf{x}_n \leftrightarrow (\mathbf{y} = \dot{0} \wedge \mathbf{z} = \mathbf{w}) \vee (\mathbf{y} \neq \dot{0} \wedge \mathbf{z} = \mathbf{a}[\mathbf{y}|\mathbf{y} - \dot{1}])$, and let $\mathbf{f}$ be defined by $\mathbf{fzwx}_1 \ldots \mathbf{x}_n = \mu\mathbf{x}\forall\mathbf{y}(\mathbf{y} < \mathbf{z} \to \mathrm{B}\mathbf{xy} = \mathbf{gywx}_1 \ldots \mathbf{x}_n)$. We then define an $(n+1)$-ary function symbol $\mathbf{h}$ by $\mathbf{hwx}_1 \ldots \mathbf{x}_n = \mathbf{f}\mathrm{S}\mathbf{wwx}_1 \ldots \mathbf{x}_n$. Intuitively, it is understood that $\mathbf{hwx}_1 \ldots \mathbf{x}_n$ is a number for the sequence of $\mathbf{a}$ as $\mathbf{y}$ varies from $\dot{0}$ to $\mathbf{w} - \dot{1}$ (being the empty sequence if $\mathbf{w} = \dot{0}$). Observe that $\mathbf{h}$ is recursive on any language containing $L(\mathrm{N})$, B, and the nonlogical symbols of $\mathbf{a}$. In the particular case when $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are exactly the variables occurring in $\mathbf{a}$ other than $\mathbf{y}$ in alphabetical order, we shall abbreviate the term $\mathbf{hbx}_1 \ldots \mathbf{x}_n$ to $\langle \mathbf{a} \rangle_{\mathbf{y}<\mathbf{b}}$. Note that $\mathbf{y}$ does not occur in $\langle \mathbf{a} \rangle_{\mathbf{y}<\mathbf{b}}$ unless it occurs in $\mathbf{b}$.

Suppose now that $\mathbf{f}$ is the $(n+1)$-ary function symbol defined by $y = \mathbf{f}xx_1 \ldots x_n \leftrightarrow (x = \dot{0} \wedge y = x_1) \vee \cdots \vee (x = n \dot{-} 1 \wedge y = x_n) \vee (\neg(x = \dot{0} \vee \cdots \vee x = n \dot{-} 1) \wedge y = \dot{0})$. We introduce the n-ary function symbol $\lozenge_n$ by $\lozenge_n x_1 \ldots x_n = \langle \mathbf{f}xx_1 \ldots x_n \rangle_{x<\dot{n}}$. By ch. III §5.2 (ix), we have

$$\vdash_P \lozenge_n x_1 \ldots x_n = \mu x(\mathrm{B}x\dot{0} = \dot{n} \wedge \mathrm{B}x\dot{1} = x_1 \wedge \cdots \wedge \mathrm{B}x\dot{n} = x_n).$$

We use $\langle \mathbf{a}_1, \ldots, \mathbf{a}_n \rangle$ as an abbreviation of $\lozenge_n \mathbf{a}_1 \ldots \mathbf{a}_n$. All the symbols $\lozenge_n$ are recursive on $L(\mathrm{N})$ with B.

Let $\mathbf{x}, \mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{y}$, and $\mathbf{z}$ be distinct variables and $\mathbf{a}$, $\mathbf{b}$ terms of P such that $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are the variables occurring in $\mathbf{a}$ in alphabetical order except $\mathbf{x}$ and the variables occurring in $\mathbf{b}$ in alphabetical order except $\mathbf{x}$. Define $\mathbf{f}$ by $\mathbf{w} = \mathbf{fzxx}_1 \ldots \mathbf{x}_n \leftrightarrow (\mathbf{x} < \mathbf{z} \wedge \mathbf{w} = \mathbf{a}) \vee (\neg(\mathbf{x} < \mathbf{z}) \wedge \mathbf{w} = \mathbf{b}[\mathbf{y}|\mathbf{x} - \mathbf{z}])$ and $\mathbf{g}$ by $\mathbf{w} = \mathbf{gzxx}_1 \ldots \mathbf{x}_n \leftrightarrow (\mathbf{x} < \mathbf{z} \to \mathbf{w} = \mathbf{a}) \vee (\neg(\mathbf{x} < \mathbf{z}) \wedge \mathbf{w} = \mathbf{a}[\mathbf{x}|\mathrm{S}\mathbf{x}])$. Then

(i) $\vdash_P \mathbf{y} < \mathbf{z} \to (\langle \mathbf{a} \rangle_{\mathbf{x}<\mathbf{z}})_{\mathbf{y}} = \mathbf{a}[\mathbf{x}|\mathbf{y}]$

(ii) $\vdash_P \mathrm{Len}\langle \mathbf{a} \rangle_{\mathbf{x}<\mathbf{z}} = \mathbf{z}$;

(iii) $\vdash_P \mathrm{Sq}\langle \mathbf{a} \rangle_{\mathbf{x}<\mathbf{z}}$;

(iv) $\vdash_P \forall\mathbf{x}(\mathbf{x} < \mathbf{z} \to \mathbf{a} = \mathbf{b}) \to \langle \mathbf{a} \rangle_{\mathbf{x}<\mathbf{z}} = \langle \mathbf{b} \rangle_{\mathbf{x}<\mathbf{z}}$;

(v) $\vdash_P \mathrm{Ini}\langle \mathbf{a} \rangle_{\mathbf{x}<\mathbf{z}}\mathbf{y} = \langle \mathbf{a} \rangle_{\mathbf{x}<\mathbf{y}}$;

(vi) $\vdash_P \langle \mathbf{a} \rangle_{\mathbf{x}<\mathbf{y}} * \langle \mathbf{b} \rangle_{\mathbf{x}<\mathbf{z}} = \langle \mathbf{fyxx}_1 \ldots \mathbf{x}_n \rangle_{\mathbf{x}<\mathbf{y}+\mathbf{z}}$;

(vii) $\vdash_P \mathbf{y} < \mathbf{z} \to \mathrm{Rmv}\langle \mathbf{a} \rangle_{\mathbf{x}<\mathbf{z}}\mathrm{S}\mathbf{y} = \langle \mathbf{gyxx}_1 \ldots \mathbf{x}_n \rangle_{\mathbf{x}<\mathbf{z}-\dot{1}}$;

and for all natural numbers k and n,

(viii) $\vdash_P \langle \mathbf{a} \rangle_{\mathbf{x}<\mathrm{S}\dot{n}} = \langle \mathbf{a}[\mathbf{x}|\dot{0}], \ldots, \mathbf{a}[\mathbf{x}|\dot{n}] \rangle$;

(ix) if $1 \leq k \leq n$, $\vdash_P \mathrm{B}\langle x_1, \ldots, x_n \rangle \dot{k} = x_k$;

(x) $\vdash_P \mathrm{Len}\langle x_1, \dots, x_n \rangle = \dot{n}$;

(xi) $\vdash_P \mathrm{Sq}\langle x_1, \dots, x_n \rangle$;

(xii) if $1 \le k \le n$, $\vdash_P \mathrm{Ini}\langle x_1, \dots, x_n \rangle \dot{k} = \langle x_1, \dots, x_k \rangle$;

(xiii) if $1 \le k \le n$, $\vdash_P \mathrm{Rmv}\langle x_1, \dots, x_n \rangle \dot{k} = \langle x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n \rangle$;

(xiv) $\vdash_P \lozenge_0 = \dot{0}$.

Again, these are straightforward consequences of the definitions.

3.5 The recursion principle.

Recursion Principle 1. Let P be a good extension of PA, B a coding function symbol in P, and $\mathbf{g}$ an $(n+2)$-ary function symbol of P. Then there is a defined $(n+1)$-ary function symbol $\mathbf{f}$ of P which is recursive on $L(\mathrm{N})$ with B and $\mathbf{g}$ such that $\vdash_P \mathbf{f}xx^n = \mathbf{g}\langle \mathbf{f}xx^n \rangle_{x<x} xx^n$.

Proof. Let $\mathbf{A}$ be $\mathrm{Len}\, y = x \wedge \forall z(z < x \to \mathrm{B}y\mathrm{S}z = \mathbf{g}\,\mathrm{Ini}\, yzzx^n)$. We prove $\vdash_P \exists y\mathbf{A}$. Using the induction axioms, it will suffice to prove

$$\vdash_P \exists y \mathbf{A}[x|\dot{0}] \text{ and} \tag{39}$$
$$\vdash_P \exists y \mathbf{A} \to \exists y \mathbf{A}[x|\mathrm{S}x]. \tag{40}$$

Since $\vdash_P \mathrm{Len}\lozenge_0 = \dot{0}$, we clearly have $\vdash_P \mathbf{A}[x, y|\dot{0}, \lozenge_0]$, whence (39). We denote by $\mathbf{a}$ the term $y * \langle \mathbf{g}\,\mathrm{Ini}\, yxxx^n \rangle$ and we prove $\vdash_P \mathbf{A} \to \mathbf{A}[x, y|\mathrm{S}x, \mathbf{a}]$. We can then infer (40) by the substitution axioms and the $\exists$-introduction rule. By properties of $*$ and Ini, we have

$$\vdash_P \mathrm{Len}\, y = x \to z < \mathrm{S}x \to \mathrm{Ini}\, yz = \mathrm{Ini}\,\mathbf{a}z \tag{41}$$

and $\vdash_P \mathrm{Len}\, y = x \to z < x \to \mathrm{B}y\mathrm{S}z = \mathrm{B}\mathbf{a}\mathrm{S}z$. From these we obtain

$$\vdash_P \mathrm{Len}\, y = x \to z < x \to \mathrm{B}y\mathrm{S}z = \mathbf{g}\,\mathrm{Ini}\, yzzx^n \to \mathrm{B}\mathbf{a}\mathrm{S}z = \mathbf{g}\,\mathrm{Ini}\,\mathbf{a}zzx^n \tag{42}$$

by the equality theorem. Also, by the choice of $\mathbf{a}$, we have

$$\vdash_P \mathrm{Len}\, y = x \to \mathrm{Len}\,\mathbf{a} = \mathrm{S}x \text{ and} \tag{43}$$
$$\vdash_P \mathrm{Len}\, y = x \to \mathrm{B}\mathbf{a}\mathrm{S}x = \mathbf{g}\,\mathrm{Ini}\, yxxx^n. \tag{44}$$

From (44) and (41) we obtain $\vdash_P \mathrm{Len}\, y = x \to z = x \to \mathrm{B}\mathbf{a}\mathrm{S}z = \mathbf{g}\,\mathrm{Ini}\,\mathbf{a}zzx^n$. From this, (43), and (42), we obtain (40).

Thus we can legitimately define a function symbol $\mathbf{f}'$ by $\mathbf{f}'xx^n = \mu y\mathbf{A}$. The actual definition of $\mathbf{f}$ is then $\mathbf{f}xx^n = \mathrm{B}\mathbf{f}'\mathrm{S}xx^n\mathrm{S}x$. Since Ini is recursive on $L(\mathrm{N})$ with B, $\mathbf{f}$ is recursive on $L(\mathrm{N})$ with B and $\mathbf{g}$. By the definition of $\mathbf{f}'$, we have

$$\vdash_P \mathrm{B}\mathbf{f}'\mathrm{S}xx^n\mathrm{S}x = \mathbf{g}\,\mathrm{Ini}\,\mathbf{f}'\mathrm{S}xx^n xxx^n \text{ and} \tag{45}$$
$$\vdash_P \mathrm{Len}\,\mathbf{f}'\mathrm{S}xx^n = \mathrm{S}x. \tag{46}$$

The latter implies $\vdash_P z < \mathrm{S}x \to \mathrm{B}\,\mathrm{Ini}\,\mathbf{f}'\mathrm{S}xx^n x\mathrm{S}z = \mathrm{B}\mathbf{f}'\mathrm{S}xx^n\mathrm{S}z$. By the definition of $\mathbf{f}'$, $\vdash_P z < \mathrm{S}x \to \mathrm{B}\mathbf{f}'\mathrm{S}xx^n\mathrm{S}z = \mathrm{B}\mathbf{f}'\mathrm{S}zx^n\mathrm{S}z$, whence $\vdash_P z < \mathrm{S}x \to \mathrm{B}\,\mathrm{Ini}\,\mathbf{f}'\mathrm{S}xx^n x\mathrm{S}z = \mathbf{f}zx^n$ by the previous formula and the definition of $\mathbf{f}$. But by properties of sequences, $\vdash_P z < \mathrm{S}x \to \mathrm{B}\langle \mathbf{f}xx^n \rangle_{x<x}\mathrm{S}z = \mathbf{f}zx^n$, and so

$$\vdash_P z < \mathrm{S}x \to \mathrm{B}\,\mathrm{Ini}\,\mathbf{f}'\mathrm{S}xx^n x\mathrm{S}z = \mathrm{B}\langle \mathbf{f}xx^n \rangle_{x<x}\mathrm{S}z. \tag{47}$$

We also have $\vdash_P \mathrm{Len}\,\mathrm{Ini}\,\mathbf{f}'\mathrm{S}xx^n x = x$ and $\vdash_P \mathrm{Len}\langle \mathbf{f}xx^n \rangle_{x<x} = x$, and together with (47) we obtain $\vdash_P \mathrm{Ini}\,\mathbf{f}'\mathrm{S}xx^n x = \langle \mathbf{f}xx^n \rangle_{x<x}$ by (35) of §3.3. From this and (45) by the equality theorem, $\vdash_P \mathrm{B}\mathbf{f}'\mathrm{S}xx^n\mathrm{S}x = \mathbf{g}\langle \mathbf{f}xx^n \rangle_{x<x} xx^n$, which is the desired conclusion. □

We remark that if $\mathbf{f}$ and $\mathbf{f}'$ both satisfy the conclusion of the recursion principle, then an easy application of the principle of complete induction shows that $\vdash_P \mathbf{f}xx^n = \mathbf{f}'xx^n$, so in that sense $\mathbf{f}$ is "unique". Sometimes we use $\mathbf{f}xx^n = \mathbf{g}\langle \mathbf{f}xx^n \rangle_{x<x} xx^n$ as if it were the defining axiom of $\mathbf{f}$. It is also possible to define predicate symbols by recursion. If $\mathbf{p}$ is an n-ary predicate symbol in a numerical theory T such that $\vdash_T \dot{0} \neq \dot{1}$, an n-ary function symbol $\mathrm{X}_\mathbf{p}$ can be defined by $y = \mathrm{X}_\mathbf{p} x_1 \dots x_n \leftrightarrow (\mathbf{p}x_1 \dots x_n \wedge y = \dot{0}) \vee (\neg\mathbf{p}x_1 \dots x_n \wedge y = \dot{1})$.

RECURSION PRINCIPLE 2. Let P be a good extension of PA, B a coding function symbol in P, and $\mathbf{q}$ an $(n+2)$-ary predicate symbol of P. Then there is a defined $(n+1)$-ary predicate symbol $\mathbf{p}$ of P which is recursive on $L(\mathrm{N})$ with B and $\mathbf{q}$ such that $\vdash_P \mathbf{p}xx^n \leftrightarrow \mathbf{q}\langle \mathrm{X}_{\mathbf{p}}xx^n\rangle_{x<x}xx^n$.

Proof. Define $\mathbf{f}$ by the recursion principle so that $\vdash_P \mathbf{f}xx^n = \mathrm{X}_{\mathbf{q}}\langle \mathbf{f}xx^n\rangle_{x<x}xx^n$, and define $\mathbf{p}$ by $\mathbf{p}xx^n \leftrightarrow \mathbf{f}xx^n = \dot{0}$. Then clearly $\vdash_P \mathbf{f}xx^n = \mathrm{X}_{\mathbf{p}}xx^n$ and $\vdash_P \mathbf{p}xx^n \leftrightarrow \mathbf{q}\langle \mathrm{X}_{\mathbf{p}}xx^n\rangle_{x<x}xx^n$. □

3.6 Examples of recursion. We end this chapter by discussing a few examples of application of the recursion principle. First, it is useful to know that if $\mathbf{f}$ is defined so that $\vdash_P \mathbf{f}xx^n = \mathbf{g}\langle \mathbf{f}xx^n\rangle_{x<x}xx^n$, then we can "unroll" this formula to obtain $\vdash_P \mathbf{f}\dot{k}x^n = \mathbf{a}$ where $\mathbf{f}$ does not occur in $\mathbf{a}$. We show this by induction on k. If $k = 0$, then $\vdash_P \langle \mathbf{f}xx^n\rangle_{x<\dot{0}} = \dot{0}$ and so $\vdash_P \mathbf{f}\dot{0}x^n = \mathbf{g}\dot{0}\dot{0}x^n$. Now if $k \geq 1$ and for all $r < k$, $\vdash_P \mathbf{f}\dot{r}x^n = \mathbf{a}_r$ for a term $\mathbf{a}_r$ in which $\mathbf{f}$ does not occur, then $\vdash_P \mathbf{f}\dot{k}x^n = \mathbf{g}\langle \mathbf{a}_0, \ldots, \mathbf{a}_{k-1}\rangle \dot{k}x^n$.

Let $\mathbf{f}$ be an $(n+1)$-ary function symbol in P. Define the $(n+3)$-ary function symbol $\mathbf{g}$ by

$$\mathbf{w} = \mathbf{g}zyxx^n \leftrightarrow (y = \dot{0} \wedge w = x) \vee (y \neq \dot{0} \wedge w = \mathrm{B}zy).$$

By the recursion principle, we can define an $(n+2)$-ary function symbol $\mathbf{f}'$ so that $\vdash_P \mathbf{f}'yxx^n = \mathbf{g}\langle \mathbf{f}'yxx^n\rangle_{y<y}yxx^n$. Now $\vdash_P y \neq \dot{0} \to \mathrm{B}\langle \mathbf{f}'yxx^n\rangle_{y<y}y = \mathbf{f}'(y - \dot{1})xx^n$, and hence from the definition of $\mathbf{g}$ we obtain $\vdash_P \mathbf{f}'\dot{0}xx^n = x$ and $\vdash_P y \neq \dot{0} \to \mathbf{f}'yxx^n = \mathbf{f}\mathbf{f}'(y - \dot{1})xx^n x^n$. For these reasons the function symbol $\mathbf{f}'$ so defined is called the *iteration* of $\mathbf{f}$, and we abbreviate $\mathbf{f}'\mathbf{b}\mathbf{a}\mathbf{a}_1 \ldots \mathbf{a}_n$ to $\mathbf{f}^{\mathbf{b}}\mathbf{a}\mathbf{a}_1 \ldots \mathbf{a}_n$.

[include further examples]

Chapter Five
Arithmetical Theories

§1 The Hilbert–Bernays–Löb derivability conditions

1.1 Löb's theorem. Let U be a numerical first-order theory and let T be an arithmetized extension of U. Let $\mathbf{D}$ be a formula of U and $\mathbf{x}$ a variable. We say that $\mathbf{D}$ *with* $\mathbf{x}$ *satisfies in* U *the derivability conditions for* T if no variable other than $\mathbf{x}$ is free in $\mathbf{D}$ and if for every closed formulae $\mathbf{A}$ and $\mathbf{B}$ of T

(i) $\vdash_U \mathbf{D}[\mathbf{x}|\ulcorner\mathbf{A}\urcorner]$ whenever $\mathbf{A}$ is a theorem of T;

(ii) $\vdash_U \mathbf{D}[\mathbf{x}|\ulcorner\mathbf{A} \to \mathbf{B}\urcorner] \to \mathbf{D}[\mathbf{x}|\ulcorner\mathbf{A}\urcorner] \to \mathbf{D}[\mathbf{x}|\ulcorner\mathbf{B}\urcorner]$;

(iii) $\vdash_U \mathbf{D}[\mathbf{x}|\ulcorner\mathbf{A}\urcorner] \to \mathbf{D}[\mathbf{x}|\ulcorner\mathbf{D}[\mathbf{x}|\ulcorner\mathbf{A}\urcorner]\urcorner]$.

Löb's Theorem. Suppose that T is diagonalizable, and let $\mathbf{A}$ be a closed formula of T. Let $\mathbf{D}$ with $\mathbf{x}$ satisfy in T the derivability conditions for T. If $\vdash_T \mathbf{D}[\mathbf{x}|\ulcorner\mathbf{A}\urcorner] \to \mathbf{A}$, then $\vdash_T \mathbf{A}$.

Proof. Let f be a diagonal function for $L(T)$ that is representable in T and let $\mathbf{D}'$ be a variant of $\mathbf{D}$ in which $\mathbf{x}_f$ is substitutible for $\mathbf{x}$. Using the fixed point theorem, we find a closed formula $\mathbf{B}$ of T such that $\vdash_T \mathbf{B} \leftrightarrow \mathbf{D}'[\mathbf{x}|\ulcorner\mathbf{B}\urcorner] \to \mathbf{A}$, and so by the variant theorem,

$$\vdash_T \mathbf{B} \leftrightarrow \mathbf{D}[\mathbf{x}|\ulcorner\mathbf{B}\urcorner] \to \mathbf{A}. \tag{1}$$

From this by the tautology theorem and (i), we obtain $\vdash_T \mathbf{D}[\mathbf{x}|\ulcorner\mathbf{B} \to \mathbf{D}[\mathbf{x}|\ulcorner\mathbf{B}\urcorner] \to \mathbf{A}\urcorner]$, whence

$$\vdash_T \mathbf{D}[\mathbf{x}|\ulcorner\mathbf{B}\urcorner] \to \mathbf{D}[\mathbf{x}|\ulcorner\mathbf{D}[\mathbf{x}|\ulcorner\mathbf{B}\urcorner]\urcorner] \to \mathbf{D}[\mathbf{x}|\ulcorner\mathbf{A}\urcorner]$$

by (ii) and the tautology theorem. As a tautological consequence of this, (iii), and the hypothesis, we obtain

$$\vdash_T \mathbf{D}[\mathbf{x}|\ulcorner\mathbf{B}\urcorner] \to \mathbf{A}, \tag{2}$$

so $\vdash_T \mathbf{B}$ by (1) and the tautology theorem. By (i), $\vdash_T \mathbf{D}[\mathbf{x}|\ulcorner\mathbf{B}\urcorner]$, whence $\vdash_T \mathbf{A}$ by (2) and the detachment rule. ☐

1.2. Löb's theorem has the following consequence, which is a very general version of Gödel's second incompleteness theorem.

Corollary. Suppose that T is diagonalizable. Let $\mathbf{D}$ with $\mathbf{x}$ satisfy in T the derivability conditions for T. If $\vdash_T \neg\mathbf{D}[\mathbf{x}|\ulcorner\mathbf{A}\urcorner]$ for some closed formula $\mathbf{A}$ of T, then T is inconsistent.

Proof. From $\vdash_T \neg\mathbf{D}[\mathbf{x}|\ulcorner\mathbf{A}\urcorner]$ we obtain $\vdash_T \mathbf{D}[\mathbf{x}|\ulcorner\mathbf{A}\urcorner] \to \mathbf{A}$ by the tautology theorem. Hence by Löb's theorem, $\vdash_T \mathbf{A}$. By (i), this implies $\vdash_T \mathbf{D}[\mathbf{x}|\ulcorner\mathbf{A}\urcorner]$. So T is inconsistent. ☐

If the formula $\mathbf{D}[\mathbf{x}|\ulcorner\mathbf{A}\urcorner]$ is taken to mean "$\mathbf{A}$ is derivable in T", then any formula of the form $\neg\mathbf{D}[\mathbf{x}|\ulcorner\mathbf{A}\urcorner]$ means "T is consistent". With this interpretation in mind, we might say of a formula $\mathbf{C}$ of T that it *expresses the consistency of* T if for some formula $\mathbf{D}$ of T which with $\mathbf{x}$ satisfies in T the derivability conditions for T and for some closed formula $\mathbf{A}$ of T, $\vdash_T \mathbf{C} \to \neg\mathbf{D}[\mathbf{x}|\ulcorner\mathbf{A}\urcorner]$. Thus, the corollary can be rephrased as follows: if T is diagonalizable and if some formula $\mathbf{C}$ expressing the consistency of T is a theorem of T, then T is inconsistent.

There are two more steps to take before this result becomes meaningful. First, we must find concrete examples of first-order theories for which the hypotheses of the corollary can be proved in a constructive manner: the remaining of this chapter is devoted to the resolution of this issue. We should also intuitively justify the interpretation of $\mathbf{D}[\mathbf{x}|\ulcorner\mathbf{A}\urcorner]$ as "$\mathbf{A}$ is derivable in T" in these cases.

§2 Arithmetical languages and theories

2.1 Arithmetical languages. In chapter IV we have seen that many recursive functions and predicates that we defined in chapter III can be "introduced" in PA by means of suitable defining axioms. Our goal is to pursue this formalization further in order to be able to express statements about first-order languages and first-order theories in PA. We begin with a formal analogue to numerotations of first-order languages.

Let L be a numerical first-order language and n a natural number. An *arithmetical language* $\mathfrak{L}$ *in* L *with* n *parameters* is given by

(i′) an n-ary predicate symbol Ω of L;

(ii′) an $(n+1)$-ary function symbol Vr of L;

(iii′) a $(n+2)$-ary predicate symbol Func of L;

(iv′) a $(n+2)$-ary predicate symbol Pred of L;

(v′) n-ary function symbols $\dot\vee$, $\dot\neg$, $\dot\exists$, $\dot=$, $\ddot0$, and $\dot{\mathrm{S}}$ of L.

We agree to the convention that the symbols forming an arithmetical language will always be abbreviated to Ω, Vr, Func, Pred, $\dot\vee$, $\dot\neg$, $\dot\exists$, $\dot=$, $\ddot0$, and $\dot{\mathrm{S}}$, the only possible variation being the uniform adjunction of superscripts or subscripts. Accordingly, the name of that arithmetical language will be $\mathfrak{L}$ with the same superscripts and subscripts. For example, the arithmetical language $\mathfrak{L}_1$ is composed of the symbols Ω_1, Vr_1, Func_1, Pred_1, $\dot\vee_1$, $\dot\neg_1$, $\dot\exists_1$, $\dot=_1$, $\ddot0_1$, and $\dot{\mathrm{S}}_1$.

Let P be a numerical first-order theory. An *arithmetical language* $\mathfrak{L}$ *in* P *with* n *parameters* is an arithmetical language in $L(P)$ with n parameters such that

(i) $\vdash_P \exists x_1 \ldots \exists x_n \Omega x^n$;

(ii) $\vdash_P \Omega x^n \to x < y \to \mathrm{Vr}\, x x^n < \mathrm{Vr}\, y x^n$;

(iii) $\vdash_P \Omega x^n \to \neg(\mathrm{Func}\,\mathrm{Vr}\, x x^n y x^n \vee \mathrm{Pred}\,\mathrm{Vr}\, x x^n y x^n)$;

(iv) $\vdash_P \Omega x^n \to \neg\,\mathrm{Func}\, x y x^n \vee \neg\,\mathrm{Pred}\, x z x^n$;

(v) $\vdash_P \Omega x^n \to y \neq z \to \mathrm{Func}\, x y x^n \to \neg\,\mathrm{Func}\, x z x^n$;

(vi) $\vdash_P \Omega x^n \to y \neq z \to \mathrm{Pred}\, x y x^n \to \neg\,\mathrm{Pred}\, x z x^n$;

(vii) $\vdash_P \Omega x^n \to \neg(\mathrm{Vr}\, x x^n = \dot\vee x^n \vee \mathrm{Vr}\, x x^n = \dot\neg x^n \vee \mathrm{Vr}\, x x^n = \dot\exists x^n)$;

(viii) $\vdash_P \Omega x^n \to \neg(\mathrm{Func}\,\dot\vee x^n x x^n \vee \mathrm{Func}\,\dot\neg x^n x x^n \vee \mathrm{Func}\,\dot\exists x^n x x^n)$;

(ix) $\vdash_P \Omega x^n \to \neg(\mathrm{Pred}\,\dot\vee x^n x x^n \vee \mathrm{Pred}\,\dot\neg x^n x x^n \vee \mathrm{Pred}\,\dot\exists x^n x x^n)$;

(x) $\vdash_P \Omega x^n \to \neg(\dot\vee x^n = \dot\neg x^n \vee \dot\vee x^n = \dot\exists x^n \vee \dot\neg x^n = \dot\exists x^n)$; and

(xi) $\vdash_P \Omega x^n \to \mathrm{Pred}\,\dot= x^n \dot2 x^n$.

An arithmetical language $\mathfrak{L}$ in P with n parameters is *numerical* if moreover

(xii) $\vdash_P \Omega x^n \to \mathrm{Func}\,\ddot0 x^n \dot0 x^n$ and

(xiii) $\vdash_P \Omega x^n \to \mathrm{Func}\,\dot{\mathrm{S}} x^n \dot1 x^n$.

In practice, we are mostly interested in arithmetical languages with no parameters. In this case, Ω is a truth value and condition (i) is reduced to $\vdash_P \Omega$, so that we can forget about Ω in (ii)–(xiii). In an arithmetical language with n parameters, Ω should be thought of as the parameter space: replacing the variable parameters by closed terms in the parameter space yields an arithmetical language without paremeters. Of course, an arithmetical language with n parameters is much more than a collection of arithmetical language without parameters. To simplify the notations, we shall usually omit these parameters and discuss all arithmetical languages as arithmetical languages without parameters, but it will be clear that no such restriction is necessary. When we do not specify the number of parameters of an arithmetical language, it must be understood that our discussion applies to arithmetical languages with any number of parameters by introducing the symbol Ω as needed. When, on the other hand, some development is specific to arithmetical languages with a specific number of parameters, we mention it explicitely.

If $\mathfrak{L}$ and $\mathfrak{L}'$ are arithmetical languages (with as many parameters) in P, we say that $\mathfrak{L}'$ is an *extension* of $\mathfrak{L}$ in P if $\vdash_P \Omega \leftrightarrow \Omega'$, $\vdash_P \mathrm{Vr}\, x = \mathrm{Vr}'\, x$, $\vdash_P \mathrm{Func}\, x y \to \mathrm{Func}'\, x y$, $\vdash_P \mathrm{Pred}\, x y \to \mathrm{Pred}'\, x y$, and for every $\mathbf{f}$ among $\dot\vee$, $\dot\neg$, $\dot\exists$, $\dot=$, $\ddot0$, and $\dot{\mathrm{S}}$, $\vdash_P \mathbf{f} = \mathbf{f}'$.

2.2. From now on we let P be a good extension of PA and we fix a coding function symbol B in P. Let $\mathfrak{L}$ be an arithmetical language in P. We use this data to define new nonlogical symbols in P. For convenience, we first define $\mathrm{Disj}\,xy = \langle \dot\vee, x, y\rangle$; $\mathrm{Neg}\,x = \langle \dot\neg, x\rangle$; $\mathrm{Inst}\,xy = \langle \dot\exists, x, y\rangle$; $\mathrm{Imp}\,xy = \langle \dot\vee, \langle \dot\neg, x\rangle, y\rangle$; $\mathrm{Cnj}\,xy = \langle \dot\neg, \langle \dot\vee, \langle \dot\neg, x\rangle, \langle \dot\neg, y\rangle\rangle\rangle$; $\mathrm{Gen}\,xy = \langle \dot\neg, \langle \dot\exists, x, \langle \dot\neg, y\rangle\rangle\rangle$; $\mathrm{Eqv}\,xy = \langle \dot\neg, \langle \dot\vee, \langle \dot\neg, \langle \dot\vee, \langle \dot\neg, x\rangle, y\rangle\rangle, \langle \dot\neg, \langle \dot\vee, \langle \dot\neg, y\rangle, x\rangle\rangle\rangle\rangle$. These definitions must be modified in the obvious way if $\mathfrak{L}$ has parameters. For example, Disj is an $(n+2)$-ary function symbol whose actual definition is $\mathrm{Disj}\,xyx^n = \langle \dot\vee x^n, x, y\rangle$. In what follows we define predicate symbols with formulae of the form $\mathbf{p}\mathbf{x}_1 \ldots \mathbf{x}_n \leftrightarrow \mathbf{A}$. In order to restore the parameters, these should be read $\mathbf{p}\mathbf{x}_1 \ldots \mathbf{x}_n \leftrightarrow \Omega \wedge \mathbf{A}$. We also define function symbols with formulae of the form $\mathbf{y} = \mathbf{f}\mathbf{x}_1 \ldots \mathbf{x}_n \leftrightarrow \mathbf{A}$, which should be read $\mathbf{y} = \mathbf{f}\mathbf{x}_1 \ldots \mathbf{x}_n \leftrightarrow (\Omega \wedge \mathbf{A}) \vee (\neg\Omega \wedge \mathbf{y} = \dot 0)$.

(i) $\mathrm{Sym}\,x \leftrightarrow x = \dot\vee \vee x = \dot\neg \vee x = \dot\exists \vee \exists y(\mathrm{Func}\,xy \vee \mathrm{Pred}\,xy \vee x = \mathrm{Vr}\,y)$;

(ii) $\mathrm{Vble}\,x \leftrightarrow x = \langle (x)_0\rangle \wedge \exists y(y < x \wedge \mathrm{Vr}\,y = (x)_0)$;

(iii) $\mathrm{Tm}\,x \leftrightarrow \mathrm{Vble}\,x \vee \mathrm{Sq}\,x \wedge \mathrm{Len}\,x \neq \dot 0 \wedge \mathrm{Func}(x)_0(\mathrm{Len}\,x - \dot 1) \wedge \forall y(y < \mathrm{Len}\,x - \dot 1 \rightarrow \mathrm{Tm}(x)_{\mathrm{S}y})$;

(iv) $\mathrm{Atfm}\,x \leftrightarrow \mathrm{Sq}\,x \wedge \mathrm{Len}\,x \neq \dot 0 \wedge \mathrm{Pred}(x)_0(\mathrm{Len}\,x - \dot 1) \wedge \forall y(y < \mathrm{Len}\,x - \dot 1 \rightarrow \mathrm{Tm}(x)_{\mathrm{S}y})$;

(v) $\mathrm{Fm}\,x \leftrightarrow \mathrm{Atfm}\,x \vee (x = \mathrm{Disj}(x)_1(x)_2 \wedge \mathrm{Fm}(x)_1 \wedge \mathrm{Fm}(x)_2) \vee (x = \mathrm{Neg}(x)_1 \wedge \mathrm{Fm}(x)_1) \vee (x = \mathrm{Inst}(x)_1(x)_2 \wedge \mathrm{Vble}(x)_1 \wedge \mathrm{Fm}(x)_2)$;

(vi) $\mathrm{Des}\,x \leftrightarrow \mathrm{Tm}\,x \vee \mathrm{Fm}\,x$;

(vii) $\mathrm{Occ}\,xy \leftrightarrow \mathrm{Des}\,x \wedge \mathrm{Des}\,y \wedge (x = y \vee \exists z(z < \mathrm{Len}\,x - \dot 1 \wedge \mathrm{Occ}(x)_{\mathrm{S}z}y))$;

(viii) $\mathrm{Fr}\,xy \leftrightarrow \mathrm{Des}\,x \wedge \mathrm{Vble}\,y \wedge ((\mathrm{Tm}\,x \vee \mathrm{Atfm}\,x) \wedge \mathrm{Occ}\,xy) \vee (x = \mathrm{Disj}(x)_1(x)_2 \wedge \mathrm{Fr}(x)_1 y \vee \mathrm{Fr}(x)_2 y) \vee (x = \mathrm{Neg}(x)_1 \wedge \mathrm{Fr}(x)_1 y) \vee (x = \mathrm{Inst}(x)_1(x)_2 \wedge \mathrm{Fr}(x)_2 y \wedge y \neq (x)_1)$;

(ix) $\mathrm{Cl}\,x \leftrightarrow \mathrm{Des}\,x \wedge \forall y(y < x \rightarrow \neg\,\mathrm{Fr}\,xy)$;

(x) $\mathrm{Subtl}\,xyz \leftrightarrow \mathrm{Des}\,x \wedge \mathrm{Vble}\,y \wedge \mathrm{Tm}\,z \wedge (x = \mathrm{Disj}(x)_1(x)_2 \wedge \mathrm{Subtl}(x)_1 yz \wedge \mathrm{Subtl}(x)_2 yz) \vee (x = \mathrm{Neg}(x)_1 \wedge \mathrm{Subtl}(x)_1 yz) \vee (x = \mathrm{Inst}(x)_1(x)_2 \wedge y \neq (x)_1 \wedge \mathrm{Subtl}(x)_2 yz \wedge \neg\,\mathrm{Fr}(x)_2 y \vee \neg\,\mathrm{Fr}\,z(x)_1) \vee \mathrm{Tm}\,x \vee \mathrm{Atfm}\,x \vee y = (x)_1$;

(xi) $w = \mathrm{Sub}\,xyz \leftrightarrow (\mathrm{Vble}\,x \wedge x \in y \wedge w = (z)_{\mathbf{f}xy}) \vee (\neg\,\mathrm{Vble}\,x \wedge w = \mu w(\mathrm{Len}\,w = \mathrm{Len}\,x \wedge (w)_0 = (x)_0 \wedge \forall x'(x' < \mathrm{Len}\,x - \dot 1 \rightarrow (w)_{\mathrm{S}x'} = \mathrm{Sub}(x)_{\mathrm{S}x'}\,\mathrm{Rmv}\,y\mathbf{g}xy\,\mathrm{Rmv}\,z\mathbf{g}xy))) \vee (\mathrm{Vble}\,x \wedge \neg(x \in y) \wedge w = x)$ where $\mathbf{f}xy = \mu z(\neg(x \in y) \vee x = (y)_z)$ and $\mathbf{g}xy = \mu z((x)_0 \neq \dot\exists \vee \neg((x)_1 \in y) \vee z \neq \dot 0 \wedge (x)_1 = \mathrm{B}yz)$;

(xii) $\Sigma\mathrm{ub}\,xy \leftrightarrow \mathrm{Sq}\,x \wedge \mathrm{Sq}\,y \wedge \mathrm{Len}\,x = \mathrm{Len}\,y \wedge \forall z(z < \mathrm{Len}\,x \rightarrow \mathrm{Vble}(x)_z \wedge \mathrm{Tm}(y)_z \wedge \forall w(w < z \rightarrow (x)_z \neq (x)_w))$;

(xiii) $\mathrm{Pax}\,x \leftrightarrow \mathrm{Fm}\,x \wedge x = \mathrm{Imp}(x)_2(x)_2$;

(xiv) $\mathrm{Sax}\,x \leftrightarrow \mathrm{Fm}\,x \wedge x = \mathrm{Imp}(x)_{1,1}\,\mathrm{Inst}(x)_{2,1}(x)_{2,2} \wedge \exists z(z < x \wedge \mathrm{Tm}\,z \wedge \mathrm{Subtl}(x)_{2,2}(x)_{2,1}z \wedge (x)_{1,1} = \mathrm{Sub}(x)_{2,2}(x)_{2,1}z)$;

(xv) $\mathrm{Iax}\,x \leftrightarrow x = \langle \dot=, (x)_1, (x)_1\rangle \wedge \mathrm{Vble}(x)_1$;

(xvi) $\mathrm{Feax}\,x \leftrightarrow \exists y(y < x \wedge \mathrm{B}^y x\dot 3 = \langle \dot=, (\mathrm{B}^y x\dot 3)_1, (\mathrm{B}^y x\dot 3)_2\rangle \wedge \mathrm{Len}(\mathrm{B}^y x\dot 3)_1 = \mathrm{S}y \wedge \mathrm{Len}(\mathrm{B}^y x\dot 3)_2 = \mathrm{S}y \wedge \mathrm{Func}(\mathrm{B}^y x\dot 3)_{1,0}y \wedge (\mathrm{B}^y x\dot 3)_{1,0} = (\mathrm{B}^y x\dot 3)_{2,0} \wedge \forall z(z < y \rightarrow \mathrm{Vble}(\mathrm{B}^y x\dot 3)_{1,\mathrm{S}z} \wedge \mathrm{Vble}(\mathrm{B}^y x\dot 3)_{2,\mathrm{S}z} \wedge \mathrm{B}^z x\dot 3 = \mathrm{Imp}\langle \dot=, (\mathrm{B}^z x\dot 3)_{1,1,1}, (\mathrm{B}^z x\dot 3)_{1,1,2}\rangle \mathrm{B}^{\mathrm{S}z} x\dot 3 \wedge (\mathrm{B}^z x\dot 3)_{1,1,1} = (\mathrm{B}^y x\dot 3)_{1,\mathrm{S}z} \wedge (\mathrm{B}^z x\dot 3)_{1,1,2} = (\mathrm{B}^y x\dot 3)_{2,\mathrm{S}z}))$;

(xvii) $\mathrm{Peax}\,x \leftrightarrow \exists y(y < x \wedge \mathrm{B}^y x\dot 3 = \mathrm{Imp}(\mathrm{B}^y x\dot 3)_{1,1}(\mathrm{B}^y x\dot 3)_2 \wedge \mathrm{Len}(\mathrm{B}^y x\dot 3)_{1,1} = \mathrm{S}y \wedge \mathrm{Len}(\mathrm{B}^y x\dot 3)_2 = \mathrm{S}y \wedge \mathrm{Pred}(\mathrm{B}^y x\dot 3)_{1,1,0}y \wedge (\mathrm{B}^y x\dot 3)_{1,1,0} = (\mathrm{B}^y x\dot 3)_{2,0} \wedge \forall z(z < y \rightarrow \mathrm{Vble}(\mathrm{B}^y x\dot 3)_{1,1,\mathrm{S}z} \wedge \mathrm{Vble}(\mathrm{B}^y x\dot 3)_{2,\mathrm{S}z} \wedge \mathrm{B}^z x\dot 3 = \mathrm{Imp}\langle \dot=, (\mathrm{B}^z x\dot 3)_{1,1,1}, (\mathrm{B}^z x\dot 3)_{1,1,2}\rangle \mathrm{B}^{\mathrm{S}z} x\dot 3 \wedge (\mathrm{B}^z x\dot 3)_{1,1,1} = (\mathrm{B}^y x\dot 3)_{1,1,\mathrm{S}z}$
$\wedge (\mathrm{B}^z x\dot 3)_{1,1,2} = (\mathrm{B}^y x\dot 3)_{2,\mathrm{S}z}))$;

(xviii) $\mathrm{Ax}\,x \leftrightarrow \mathrm{Pax}\,x \vee \mathrm{Sax}\,x \vee \mathrm{Iax}\,x \vee \mathrm{Feax}\,x \vee \mathrm{Peax}\,x$;

(xix) $\mathrm{Ctr}\,xy \leftrightarrow \mathrm{Fm}\,x \wedge y = \mathrm{Disj}\,xx$;

(xx) $\mathrm{Exp}\,xy \leftrightarrow \mathrm{Fm}\,x \wedge x = \mathrm{Disj}(x)_1 y$;

(xxi) $\mathrm{Assoc}\,xy \leftrightarrow \mathrm{Fm}\,x \wedge x = \mathrm{Disj}\,\mathrm{Disj}(x)_{1,1}(x)_{1,2}(x)_2 \wedge y = \mathrm{Disj}(x)_{1,1}\,\mathrm{Disj}(x)_{1,2}(x)_2$;

(xxii) $\mathrm{Cut}\,xyz \leftrightarrow \mathrm{Fm}\,y \wedge \mathrm{Fm}\,z \wedge y = \mathrm{Disj}(y)_1(y)_2 \wedge z = \mathrm{Imp}(y)_1(z)_2 \wedge x = \mathrm{Disj}(y)_2(z)_2$;

(xxiii) $\mathrm{Intr}\,xy \leftrightarrow \mathrm{Fm}\,x \wedge x = \mathrm{Imp}\,\mathrm{Inst}(x)_{1,1,1}(x)_{1,1,2}(x)_2 \wedge y = \mathrm{Imp}(x)_{1,1,2}(x)_2 \wedge \neg\,\mathrm{Fr}(x)_2(x)_{1,1,1}$;

(xxiv) $\mathrm{Inf}\,xy \leftrightarrow \mathrm{Sq}\,y \wedge \exists z(z < \mathrm{Len}\,y \wedge (\mathrm{Ctr}\,x\mathrm{B}y\mathrm{S}z \vee \mathrm{Exp}\,x\mathrm{B}y\mathrm{S}z \vee \mathrm{Assoc}\,x\mathrm{B}y\mathrm{S}z \vee \exists w(w < \mathrm{Len}\,y \wedge \mathrm{Cut}\,x\mathrm{B}y\mathrm{S}z\mathrm{B}y\mathrm{S}w) \vee \mathrm{Intr}\,x\mathrm{B}y\mathrm{S}z))$;

(xxv) $y = \mathrm{Num}\,x \leftrightarrow (x = \dot 0 \wedge y = \ddot 0) \vee (x \neq \dot 0 \wedge y = \langle \dot{\mathrm{S}}, \mathrm{Num}(x - \dot 1)\rangle)$.

All these defined symbols are subject to the same notational convention as arithmetical languages: we require that they inherit the superscripts and subscripts of the arithmetical language from which they are defined. Thus from the arithmetical language $\mathfrak{L}^*$ are defined Vble*, Tm*, etc. We observe that all of them except (i) are recursive on $L(P)$, and in fact on any extension of $L(\mathrm{N})$ on which B and the symbols of $\mathfrak{L}$ are recursive.

We now list some consequences of the definitions.

(i′) $\vdash_P x \leq \mathrm{Vr}\,x$;

(ii′) $\vdash_P \mathrm{Vble}\,x \leftrightarrow x = \langle (x)_0 \rangle \wedge \exists y(\mathrm{Vr}\,y = (x)_0)$;

(iii′) $\vdash_P \mathrm{Tm}\,x \to \neg\,\mathrm{Fm}\,x$;

(iv′) $\vdash_P \mathrm{Occ}\,xy \to y \leq x$;

(v′) $\vdash_P \mathrm{Vble}\,x \to \mathrm{Occ}\,xy \to x = y$;

(vi′) $\vdash_P \mathrm{Occ}\,xy \to \mathrm{Occ}\,yz \to \mathrm{Occ}\,xz$;

(vii′) $\vdash_P \mathrm{Fr}\,xy \to \mathrm{Occ}\,xy$;

(viii′) $\vdash_P \mathrm{Tm}\,x \to \mathrm{Vble}\,y \to \mathrm{Tm}\,z \to \mathrm{Subtl}\,xyz$;

(ix′) $\vdash_P \mathrm{Fm}\,x \to \mathrm{Vble}\,y \to \mathrm{Tm}\,z \to \forall w(\mathrm{Vble}\,w \to \mathrm{Occ}\,xw \to \neg\,\mathrm{Occ}\,zw) \to \mathrm{Subtl}\,xyz$;

(x′) $\vdash_P \mathrm{Fm}\,x \to \mathrm{Vble}\,y \to \mathrm{Tm}\,z \to x = \mathrm{Inst}(x)_1(x)_2 \to \mathrm{Occ}\,z(x)_1 \to \mathrm{Subtl}\,xyz \to \neg\,\mathrm{Fr}\,xy$;

(xi′) $\vdash_P \mathrm{Tm}\,x \to \Sigma\mathrm{ub}\,yz \to \mathrm{Tm}\,\mathrm{Sub}\,xyz$;

(xii′) $\vdash_P \mathrm{Fm}\,x \to \Sigma\mathrm{ub}\,yz \to \mathrm{Fm}\,\mathrm{Sub}\,xyz$;

(xiii′) $\vdash_P \mathrm{Des}\,x \to \Sigma\mathrm{ub}\,yz \to \forall w(w < \mathrm{Len}\,y \to \neg\,\mathrm{Fr}\,x(y)_w) \to \mathrm{Sub}\,xyz = x$;

(xiv′) $\vdash_P \mathrm{Des}\,x \to \Sigma\mathrm{ub}\,yz \to w < \mathrm{Len}\,y \to \neg\,\mathrm{Fr}\,x(y)_w \to \mathrm{Sub}\,xyz = \mathrm{Sub}\,x\,\mathrm{Rmv}\,ySw\,\mathrm{Rmv}\,zSw$;

(xv′) $\vdash_P \mathrm{Des}\,x \to \Sigma\mathrm{ub}\,yz \to \forall w(\mathrm{Vble}\,w \to \mathrm{Fr}\,xw \to w \in y) \to \forall w(w < \mathrm{Len}\,z \to \mathrm{Cl}(z)_w) \to \mathrm{Cl}\,\mathrm{Sub}\,xyz$;

(xvi′) $\vdash_P \mathrm{Des}\,x \to \Sigma\mathrm{ub}\,yz \to \mathrm{Vble}\,y' \to \mathrm{Tm}\,z' \to \forall w(w < \mathrm{Len}\,y \to \mathrm{Subtl}\,x(y)_w(z)_w) \to \forall w(w < \mathrm{Len}\,y \to \mathrm{Occ}(z)_w y' \to \mathrm{Subtl}\,x(y)_w z') \to \mathrm{Subtl}\,\mathrm{Sub}\,xyzy'z'$;

(xvii′) $\vdash_P \mathrm{Des}\,x \to \Sigma\mathrm{ub}\,yy' \to \Sigma\mathrm{ub}\,y'z \to \forall w(w < \mathrm{Len}\,y \to \mathrm{Subtl}\,x(y)_w(y')_w \wedge \neg\,\mathrm{Fr}\,x(y')_w) \to \mathrm{Sub}\,\mathrm{Sub}\,xyy'y'z = \mathrm{Sub}\,xyz$;

By the recursion principle we can define a function symbol $\mathbf{h}'$ so that $\vdash_P \mathbf{h}'\dot{0}xyz = x$ and $\vdash_P w \neq \dot{0} \to \mathbf{h}'wxyz = \mathrm{Sub}\,\mathbf{h}'(w-1)xyz\langle \mathrm{B}yw\rangle\langle \mathrm{B}zw\rangle$, and we set $\mathbf{h}xyz = \mathbf{h}'\,\mathrm{Len}\,yxyz$.

(xviii′) $\vdash_P \mathrm{Des}\,x \to \Sigma\mathrm{ub}\,yz \to \forall w \forall w'(w' < \mathrm{Len}\,y \to w < w' \to \neg\,\mathrm{Occ}(z)_w(y)_{w'}) \to \mathbf{h}xyz = \mathrm{Sub}\,xyz$;

(xix′) $\vdash_P \mathrm{Des}\,x \to \Sigma\mathrm{ub}\,yz \to \Sigma\mathrm{ub}\,y'z' \to \forall w \forall w'(w < \mathrm{Len}\,y \to w' < \mathrm{Len}\,y' \to \neg\,\mathrm{Fr}(z')_{w'}(y)_w \wedge \neg\,\mathrm{Fr}(z)_w(y')_{w'}) \to \mathrm{Sub}\,\mathrm{Sub}\,xyzy'z' = \mathrm{Sub}\,\mathrm{Sub}\,xy'z'yz$;

(xx′) $\vdash_P \mathrm{Ax}\,x \to \mathrm{Fm}\,x$;

and if $\mathfrak{L}$ is numerical,

(xxi′) $\vdash_P \mathrm{Tm}\,\mathrm{Num}\,x$;

(xxii′) $\vdash_P \mathrm{Cl}\,\mathrm{Num}\,y$.

[insert a few examples of derivations]

Proposition. If $\mathfrak{L}'$ is an extension of $\mathfrak{L}$, then for every n-ary predicate symbol $\mathbf{p}$ among (i)–(xxv), $\vdash_P \mathbf{p}x_1 \ldots x_n \to \mathbf{p}'x_1 \ldots x_n$, $\vdash_P \mathrm{Sub}\,xyz = \mathrm{Sub}'\,xyz$, and $\vdash_P \mathrm{Num}\,x = \mathrm{Num}'\,x$.

The proof is straightforward using the defining axioms and the principle of complete induction when the symbol is defined by recursion.

2.3 Arithmetical theories. An *arithmetical theory* $\mathfrak{T}$ *in* P *with* n *parameters* consists of an arithmetical language $\mathfrak{L}(\mathfrak{T})$ in P with n parameters, called the *language of* $\mathfrak{T}$, and an $(n+1)$-ary predicate symbol Nlax of P such that $\vdash_P \Omega x^n \to \mathrm{Nlax}\,xx^n \to \mathrm{Fm}\,xx^n$, where Fm is defined from $\mathfrak{L}(\mathfrak{T})$. We agree that when an arithmetical theory in T has the name $\mathfrak{T}$ possibly with superscripts or subscripts, then the associated predicate symbol will be Nlax with the same superscripts and subscripts, and, unless otherwise specified, its language will be $\mathfrak{L}$ with the same superscripts and subscripts. An arithmetical theory is *numerical* if its language is numerical. Our remarks on parameters in §2.1 also apply to arithmetical theories.

If $\mathfrak{T}$ is an arithmetical theory in P, we define

(i) $\mathrm{Der}\,xy \leftrightarrow \mathrm{Sq}\,y \wedge \mathrm{Len}\,y \neq \dot{0} \wedge \forall z(z < \mathrm{Len}\,y \to \mathrm{Ax}(y)_z \vee \mathrm{Nlax}(y)_z \vee \mathrm{Inf}(y)_z\,\mathrm{Ini}\,yz)$
$\wedge\, x = \mathrm{B}y\,\mathrm{Len}\,y$;

(ii) $\mathrm{Thm}\,x \leftrightarrow \exists y\,\mathrm{Der}\,xy$;

(iii) $\mathrm{Con} \leftrightarrow \exists x(\mathrm{Fm}\,x \wedge \neg\,\mathrm{Thm}\,x)$;

(iv) $\mathrm{Cm} \leftrightarrow \forall x(\mathrm{Fm}\,x \to \mathrm{Cl}\,x \to \mathrm{Thm}\,x \vee \mathrm{Thm}\,\mathrm{Neg}\,x)$;

(v) $\mathrm{Hk} \leftrightarrow \forall x(\mathrm{Fm}\,x \to \mathrm{Cl}\,x \to x = \mathrm{Inst}(x)_1(x)_2$
$\to \exists y(\mathrm{Func}\,y\dot{0} \wedge \mathrm{Thm}\,\mathrm{Imp}\,\mathrm{Inst}(x)_2(x)_1\,\mathrm{Sub}(x)_2\langle(x)_1\rangle\langle\langle y\rangle\rangle))$.

As for arithmetical languages, the above defined symbols inherit the superscripts and subscripts of the name of the arithmetical theory from which they are defined. Note that the predicate symbols Con, Cm, and Hk express the consistency, the completeness, and the Henkin property of $\mathfrak{T}$, respectively. It is clear that Der is recursive on $L(\mathrm{N})$ with B and the symbols of $\mathfrak{T}$. Moreover, if B and the symbols of $\mathfrak{L}(\mathfrak{T})$ are recursive on an extension L of $L(\mathrm{N})$ and if Nlax is recursively enumerable on L, then Der and Thm are recursively enumerable on L.

We list a few easily verified results about arithmetical theories.

(i′) $\vdash_P \mathrm{Ax}\,x \to \mathrm{Thm}\,x$;

(ii′) $\vdash_P \mathrm{Nlax}\,x \to \mathrm{Thm}\,x$;

(iii′) $\vdash_P \mathrm{Der}\,xy \to \forall z(z < \mathrm{Len}\,y \to \mathrm{Thm}(y)_z)$;

(iv′) $\vdash_P \mathrm{Der}\,xy \to z < \mathrm{Len}\,y \to \mathrm{Der}(y)_z\,\mathrm{Ini}\,y\mathrm{S}z$.

If $\mathfrak{T}$ and $\mathfrak{T}'$ are arithmetical theories (with as many parameters) in P, we say that $\mathfrak{T}'$ is an *extension* of $\mathfrak{T}$ in P if $\mathfrak{L}(\mathfrak{T}')$ is an extension of $\mathfrak{L}(\mathfrak{T})$, and if $\vdash_P \mathrm{Nlax}\,x \to \mathrm{Thm}'\,x$.

2.4 Extension and restriction. Let $\mathfrak{T}$ be a arithmetical theory in P with n paremeters. We associate to $\mathfrak{T}$ two useful arithmetical theories $\mathfrak{T}[]$ and $\mathfrak{T}{\restriction}$ with $n+1$ parameters in a recursive extension by definitions of P. We shall abbreviate $\mathrm{Vr}[]\mathbf{aba}_1\ldots\mathbf{a}_n$ to $\mathrm{Vr}[\mathbf{b}]\mathbf{aa}_1\ldots\mathbf{a}_n$, $\mathrm{Vr}{\restriction}\mathbf{aba}_1\ldots\mathbf{a}_n$ to $\mathrm{Vr}_{\restriction\mathbf{b}}\,\mathbf{aa}_1\ldots\mathbf{a}_n$, and similarly for other symbols. As usual, we take $n = 0$ to simplify the notations. The symbols of $\mathfrak{T}[]$ are defined by: $\Omega[x] \leftrightarrow \Omega \wedge \mathrm{Sq}\,x \wedge \forall y(y < \mathrm{Len}\,x \to \mathrm{Fm}(x)_y)$; $\mathrm{Vr}[z]x = \mathrm{Vr}\,x$; $\mathrm{Func}[z]xy \leftrightarrow \mathrm{Func}\,xy$; $\mathrm{Pred}[z]xy \leftrightarrow \mathrm{Pred}\,xy$; for $\mathbf{f}$ among $\dot{\vee}$, $\dot{\neg}$, $\dot{\exists}$, $\dot{=}$, $\ddot{0}$, and $\dot{\mathrm{S}}$, $\mathbf{f}[z] = \mathbf{f}$; $\mathrm{Nlax}[z]x \leftrightarrow \mathrm{Nlax}\,x \vee x \in z$.

The arithmetical theory $\mathfrak{T}{\restriction}$ has the same language as $\mathfrak{T}[]$, except for $\Omega_{\restriction z} \leftrightarrow \Omega$, and we define $\mathrm{Nlax}_{\restriction z}\,x \leftrightarrow \mathrm{Nlax}\,x \wedge x < z$. It is clear that $\mathfrak{T}[]$ and $\mathfrak{T}{\restriction}$ are arithmetical theories in a recursive extension by definitions of P. The fundamental properties of these arithmetical theories in relation to $\mathfrak{T}$ are expressed below.

(i) $\vdash_P \Omega[y] \to \mathrm{Thm}[y]x \to \forall z(z < \mathrm{Len}\,y \to \mathrm{Thm}(y)_z) \to \mathrm{Thm}\,x$;

(ii) $\vdash_P \mathrm{Thm}\,x \leftrightarrow \exists z\,\mathrm{Thm}_{\restriction z}\,x$;

(iii) $\vdash_P \mathrm{Con} \leftrightarrow \forall z\,\mathrm{Con}_{\restriction z}$.

[(i) is never used]
Note that $\vdash_P \mathrm{Ax}_{\restriction z}\,x \leftrightarrow \mathrm{Ax}\,x$ and $\vdash_P \mathrm{Inf}_{\restriction z}\,xy \leftrightarrow \mathrm{Inf}\,xy$. The implication from right to left in (ii) is an easy consequence of this and the definition of Thm. The other implication is derived from $\vdash_P \mathrm{Der}\,xy \to \mathrm{Der}_{\restriction y}\,xy$ which again is immediate from the definition of Der. Then we have $\vdash_P \neg\,\mathrm{Thm}\,x \leftrightarrow \forall z \neg\,\mathrm{Thm}_{\restriction z}\,x$, whence (iii).

2.5 Change of numerotation. Let $\mathfrak{L}$ and $\mathfrak{L}'$ be arithmetical languages in P with n parameters. An $(n+1)$-ary function symbol $\mathbf{f}$ of P is called a *change of numerotation from $\mathfrak{L}$ to $\mathfrak{L}'$* if

(i) $\vdash_P \Omega x^n \leftrightarrow \Omega' x^n$;

(ii) $\vdash_P \Omega x^n \to \mathrm{Sym}\,xx^n \to \mathbf{f}xx^n = \mathbf{f}yx^n \to x = y$;

(iii) $\vdash_P \Omega x^n \to \mathrm{Vr}'\,xx^n = \mathbf{f}\,\mathrm{Vr}\,xx^n x^n$;

(iv) $\vdash_P \Omega x^n \to \mathrm{Func}\,xyx^n \leftrightarrow \mathrm{Func}'\,\mathbf{f}xx^n yx^n$;

(v) $\vdash_P \Omega x^n \to \mathrm{Pred}\,xyx^n \leftrightarrow \mathrm{Pred}'\,\mathbf{f}xx^n yx^n$;

(vi) for $\mathbf{g}$ among $\dot{\vee}$, $\dot{\neg}$, $\dot{\exists}$, $\dot{=}$, $\ddot{0}$, and $\dot{\mathrm{S}}$, $\vdash_P \Omega x^n \to \mathbf{g}' x^n = \mathbf{fg}x^n x^n$;

The condition (ii) allows the definition of an $(n+1)$-ary function symbol $\mathbf{f}^{-1}$ by $y = \mathbf{f}^{-1}xx^n \leftrightarrow (\Omega x^n \wedge \mathrm{Sym}\, yx^n \wedge x = \mathbf{f}yx^n) \vee (\neg(\Omega x^n \wedge \exists z(\mathrm{Sym}\, zx^n \wedge \mathbf{f}zx^n = x)) \wedge y = \dot{0})$. Using (iii)–(vi) we deduce that $\vdash_P \Omega x^n \to \mathrm{Sym}'\, xx^n \leftrightarrow \exists y(\mathrm{Sym}\, yx^n \wedge x = \mathbf{f}yx^n)$. From this it is easy to check that $\mathbf{f}^{-1}$ is a change of numerotation from $\mathfrak{L}'$ to $\mathfrak{L}$. Define temporarily $\mathbf{h_f}$ by recursion so that

$$\vdash_P \mathbf{h_f}xx^n = \mu y(\mathrm{Len}\, y = \mathrm{Len}\, x \wedge (y)_0 = \mathbf{f}(x)_0 x^n \wedge \forall z(z < \mathrm{Len}\, x \mathbin{\dot{-}} \dot{1} \to (y)_{\mathrm{S}z} = \mathbf{h_f}(x)_{\mathrm{S}z}x^n)).$$

We then define $\mathbf{f}_*$ by $y = \mathbf{f}_*xx^n \leftrightarrow (\mathrm{Des}\, x \wedge y = \mathbf{h_f}xx^n) \vee (\neg\,\mathrm{Des}\, x \wedge y = \dot{0})$ and $\mathbf{f}^*$ by $\mathbf{f}^*xx^n = (\mathbf{f}^{-1})_*xx^n$.

Let $\mathfrak{L}$ be an arithmetical language in P and $\mathbf{f}$ a function symbol of P such that (ii) holds. We can then define an arithmetical language $\mathfrak{L}_\mathbf{f}$ in an extension by definitions of P as follows: $\Omega_\mathbf{f} \leftrightarrow \Omega$; $\mathrm{Vr}_\mathbf{f}\, x = \mathbf{f}\,\mathrm{Vr}\, x$; $\mathrm{Func}_\mathbf{f}\, xy \leftrightarrow \mathrm{Func}\, \mathbf{f}^{-1}xy$; $\mathrm{Pred}_\mathbf{f}\, xy \leftrightarrow \mathrm{Pred}\, \mathbf{f}^{-1}xy$; and for $\mathbf{g}$ among $\dot{\vee}$, $\dot{\neg}$, $\dot{\exists}$, $\dot{=}$, $\ddot{0}$, and $\dot{\mathrm{S}}$, $\mathbf{g_f} = \mathbf{fg}$. Then $\mathbf{f}$ is a change of numerotation from $\mathfrak{L}$ to $\mathfrak{L}_\mathbf{f}$.

Let $\mathfrak{T}$ and $\mathfrak{T}'$ be arithmetical theories in P. A function symbol $\mathbf{f}$ of P is a *change of numerotation from $\mathfrak{T}$ to $\mathfrak{T}'$* if it is a change of numerotation from $\mathfrak{L}(\mathfrak{T})$ to $\mathfrak{L}(\mathfrak{T}')$ and if $\vdash_P \mathrm{Nlax}'\, x \leftrightarrow \mathrm{Nlax}\, \mathbf{f}^*x$.

As before, if $\mathfrak{T}$ is an arithmetical theory in P with n parameters and $\mathbf{f}$ an $(n+1)$-ary function symbol of P satisfying (ii), we can define an arithmetical theory $\mathfrak{T}_\mathbf{f}$ in an extension by definitions of P so that $\mathbf{f}$ is a change of numerotation from $\mathfrak{T}$ to $\mathfrak{T}_\mathbf{f}$: let the language of $\mathfrak{T}_\mathbf{f}$ be $\mathfrak{L}_\mathbf{f}$ and define $\mathrm{Nlax}_\mathbf{f}\, x \leftrightarrow \mathrm{Nlax}\, \mathbf{f}^*x$.

We say that $\mathfrak{L}$ and $\mathfrak{L}'$ (resp. $\mathfrak{T}$ and $\mathfrak{T}'$) *differ by a change of numerotation* in P if there is a change of numerotation from $\mathfrak{L}$ to $\mathfrak{L}'$ (resp. from $\mathfrak{T}$ to $\mathfrak{T}'$) in an extension by definitions of P. We should now make a long list of theorems of the form $\vdash_P \mathrm{Fm}\, x \leftrightarrow \mathrm{Fm}'\, \mathbf{f}_*x$, but since we do not intend to derive any of them, we simply state the final result which will be used in §5.2.

Proposition. If $\mathfrak{T}$ and $\mathfrak{T}'$ differ by a change of numerotation in P, then $\vdash_P \mathrm{Con} \leftrightarrow \mathrm{Con}'$.

2.6 The formalized substitution rule. With the definition of an arithmetical theory in P and its associated defined symbols, it is now possible to express and derive in P formal versions of the general results on first-order theories for arithmetical theories: the derived rules, the tautology theorem, the equivalence and equality theorems, the theorems on definitions, the interpretation theorem, etc. For instance, the tautology theorem becomes the formula $\mathrm{Taut}\, xy \to \mathrm{Thm}[y]x$ for some suitably defined predicate symbol Taut meaning "the formula x is a tautological consequence of the formulae $z \in y$". This is obviously a very laborious task, so we shall be content with deriving only those few results which will be required later on. These include a few instances of the tautology theorem, the substitution rule, and the interpretation theorem. We begin by proving a result expressing that an application of a rule of inference to theorems yields a theorem.

Lemma 1. Let $\mathfrak{T}$ be an arithmetical theory in P.

(i) $\vdash_T \mathrm{Thm}\, x \to \mathrm{Ctr}\, yx \to \mathrm{Thm}\, y$.
(ii) $\vdash_T \mathrm{Thm}\, x \to \mathrm{Exp}\, yx \to \mathrm{Thm}\, y$.
(iii) $\vdash_T \mathrm{Thm}\, x \to \mathrm{Assoc}\, yx \to \mathrm{Thm}\, y$.
(iv) $\vdash_T \mathrm{Thm}\, x \to \mathrm{Thm}\, y \to \mathrm{Cut}\, zxy \to \mathrm{Thm}\, z$.
(v) $\vdash_T \mathrm{Thm}\, x \to \mathrm{Intr}\, yx \to \mathrm{Thm}\, y$.
(vi) $\vdash_T \mathrm{Inf}\, xy \to \forall z(z < \mathrm{Len}\, y \to \mathrm{Thm}(y)_z) \to \mathrm{Thm}\, x$.

Proof. We shall only prove (i); (ii)–(v) are completely analogous and (vi) is an easy consequence of (i)–(v). By the substitution axioms and the $\exists$-introduction rule, it will suffice to prove $\vdash_P \mathrm{Der}\, xz \to \mathrm{Ctr}\, yx \to \mathrm{Der}\, y(z * \langle y\rangle)$. Let P' be obtained from P by the adjunction of new constants $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ and the axioms $\mathrm{Der}\, \mathbf{e}_1\mathbf{e}_3$ and $\mathrm{Ctr}\, \mathbf{e}_2\mathbf{e}_1$. We must then prove $\vdash_P \mathrm{Der}\, \mathbf{e}_2(\mathbf{e}_3 * \langle \mathbf{e}_2\rangle)$. We let $\mathbf{a}$ stand for $\mathbf{e}_3 * \langle \mathbf{e}_2\rangle$. Then $\vdash_{P'} \mathrm{Sq}\, \mathbf{a}$, $\vdash_{P'} \mathrm{Len}\, \mathbf{a} \neq \dot{0}$, and $\vdash_{P'} \mathbf{e}_2 = \mathrm{B}\mathbf{a}\,\mathrm{Len}\, \mathbf{a}$. Since $\vdash_{P'} x < \mathbf{a} \to x < \mathrm{Len}\, \mathbf{e}_3 \vee x = \mathrm{Len}\, \mathbf{e}_3$, it remain to prove

$$\vdash_{P'} x < \mathrm{Len}\, \mathbf{e}_3 \to \mathrm{Ax}(\mathbf{a})_x \vee \mathrm{Nlax}(\mathbf{a})_x \vee \mathrm{Inf}(\mathbf{a})_x\, \mathrm{Ini}\, \mathbf{a}x, \text{ and} \tag{1}$$

$$\vdash_{P'} x = \mathrm{Len}\, \mathbf{e}_3 \to \mathrm{Inf}(\mathbf{a})_x\, \mathrm{Ini}\, \mathbf{a}x. \tag{2}$$

Now (1) follows at once from the axiom $\mathrm{Der}\, \mathbf{e}_1\mathbf{e}_3$. From this axiom we also derive $\vdash_{P'} \mathbf{e}_1 = \mathrm{B}\mathbf{e}_3\, \mathrm{Len}\, \mathbf{e}_3$. From this and the axiom $\mathrm{Ctr}\, \mathbf{e}_2\mathbf{e}_1$, we obtain $\vdash_{P'} \mathrm{Inf}\, \mathbf{e}_2\mathbf{e}_3$. But $\vdash_{P'} \mathbf{e}_2 = (\mathbf{a})_{\mathrm{Len}\, \mathbf{e}_3}$ and $\vdash_{P'} \mathbf{e}_3 = \mathrm{Ini}\, \mathbf{a}\, \mathrm{Len}\, \mathbf{e}_3$, and hence (2) by the equality theorem. □

LEMMA 2. If $\mathfrak{T}'$ is an extension of $\mathfrak{T}$, then $\vdash_P \mathrm{Thm}\, x \to \mathrm{Thm}'\, x$ and $\vdash_P \mathrm{Con}' \to \mathrm{Con}$.

Proof. The second assertion is a direct consequence of the first. By (vi) of lemma 1 and the substitution rule,

$$\vdash_P \mathrm{Inf}'\, \mathrm{B}ySz\, \mathrm{Ini}\, yz \to \forall w(w < \mathrm{Len}\, \mathrm{Ini}\, yz \to \mathrm{Thm}'\, \mathrm{B}\, \mathrm{Ini}\, yzSw) \to \mathrm{Thm}'\, \mathrm{B}ySz.$$

Using $\vdash_P \mathrm{Len}\, \mathrm{Ini}\, yz = z$ and $\vdash_P w < z \to \mathrm{B}\, \mathrm{Ini}\, yzSw = \mathrm{B}ySw$, we obtain $\vdash_P \mathrm{Inf}'\, \mathrm{B}ySz\, \mathrm{Ini}\, yz \to \forall w(w < z \to \mathrm{Thm}'\, \mathrm{B}ySw) \to \mathrm{Thm}'\, \mathrm{B}ySz$, whence

$$\vdash_P z < \mathrm{Len}\, y \to \mathrm{Inf}'\, \mathrm{B}ySz\, \mathrm{Ini}\, yz \to \forall w(w < z \to w < \mathrm{Len}\, y \to \mathrm{Thm}'\, \mathrm{B}ySw) \to \mathrm{Thm}'\, \mathrm{B}ySz. \tag{3}$$

By definition of Der, we have $\vdash_P \mathrm{Der}\, xy \to z < \mathrm{Len}\, y \to \mathrm{Ax}\, \mathrm{B}ySz \vee \mathrm{Nlax}\, \mathrm{B}ySz \vee \mathrm{Inf}\, \mathrm{B}ySz\, \mathrm{Ini}\, yz$. Since $\mathfrak{L}(\mathfrak{T}')$ is an extension of $\mathfrak{L}(\mathfrak{T})$, we have $\vdash_P \mathrm{Ax}\, \mathrm{B}ySz \to \mathrm{Thm}'\, \mathrm{B}ySz$ and $\vdash_P \mathrm{Inf}\, \mathrm{B}ySz\, \mathrm{Ini}\, yz \to \mathrm{Inf}'\, \mathrm{B}ySz\, \mathrm{Ini}\, yz$, and since $\mathfrak{T}'$ is an extension of $\mathfrak{T}$, $\vdash_P \mathrm{Nlax}\, \mathrm{B}ySz \to \mathrm{Thm}'\, \mathrm{B}ySz$. Thus

$$\vdash_P \mathrm{Der}\, xy \to z < \mathrm{Len}\, y \to \mathrm{Thm}'\, \mathrm{B}ySz \vee \mathrm{Inf}'\, \mathrm{B}ySz\, \mathrm{Ini}\, yz. \tag{4}$$

From (3) and (4) by the tautology theorem, $\vdash_P \mathrm{Der}\, xy \to \forall w(w < z \to w < \mathrm{Len}\, y \to \mathrm{Thm}'\, \mathrm{B}ySw) \to z < \mathrm{Len}\, y \to \mathrm{Thm}'\, \mathrm{B}ySz$. By the $\forall$-introduction rule and the principle of complete induction, $\vdash_P \mathrm{Der}\, xy \to \forall z(z < \mathrm{Len}\, y \to \mathrm{Thm}'\, \mathrm{B}ySz)$, whence $\vdash_P \mathrm{Der}\, xy \to \mathrm{Len}\, y - \dot{1} < \mathrm{Len}\, y \to \mathrm{Thm}'\, x$ by the substitution theorem. But $\vdash_P \mathrm{Der}\, xy \to \mathrm{Len}\, y - \dot{1} < \mathrm{Len}\, y$, and hence we find $\vdash_P \mathrm{Der}\, xy \to \mathrm{Thm}'\, x$. We conclude using the $\exists$-introduction rule. ☐

Remark. For any arithmetical theories $\mathfrak{T}$ and $\mathfrak{T}'$ in P, we have $\vdash_P \forall x(\mathrm{Nlax}\, x \to \mathrm{Thm}'\, x) \to \mathrm{Thm}\, x \to \mathrm{Thm}'\, x$, which by the deduction theorem is equivalent to $\vdash_{P[\mathrm{Nlax}\, x \to \mathrm{Thm}'\, x]} \mathrm{Thm}\, x \to \mathrm{Thm}'\, x$. This is, in fact, a consequence of the lemma, for $\mathfrak{T}'$ is certainly an extension of $\mathfrak{T}$ in $P[\mathrm{Nlax}\, x \to \mathrm{Thm}'\, x]$ which is also a good extension of PA. A similar argument can be made in many other situations.

We now use lemma 1 to derive some elementary formalized rules.

(i) $\vdash_P \mathrm{Thm}\, \mathrm{Disj}\, xy \to \mathrm{Thm}\, \mathrm{Disj}\, yx$;

(ii) $\vdash_P \mathrm{Thm}\, \mathrm{Imp}\, xy \to \mathrm{Thm}\, x \to \mathrm{Thm}\, y$;

(iii) $\vdash_P \mathrm{Thm}\, \mathrm{Imp}\, xy \to \mathrm{Thm}\, \mathrm{Imp}\, \mathrm{Neg}\, y\, \mathrm{Neg}\, x$;

(iv) $\vdash_P \mathrm{Thm}\, \mathrm{Imp}\, \mathrm{Neg}\, xy \to \mathrm{Thm}\, \mathrm{Imp}\, \mathrm{Neg}\, yx$;

(v) $\vdash_P \mathrm{Thm}\, x \vee \mathrm{Thm}\, y \to \mathrm{Thm}\, \mathrm{Disj}\, xy$;

(vi) $\vdash_P \mathrm{Thm}\, \mathrm{Cnj}\, xy \leftrightarrow \mathrm{Thm}\, x \wedge \mathrm{Thm}\, y$;

(vii) $\vdash_P \mathrm{Thm}\, \mathrm{Neg}\, \mathrm{Disj}\, xy \leftrightarrow \mathrm{Thm}\, \mathrm{Cnj}\, \mathrm{Neg}\, x\, \mathrm{Neg}\, y$;

(viii) $\vdash_P \mathrm{Thm}\, \mathrm{Eqv}\, xy \to \mathrm{Thm}\, x \leftrightarrow \mathrm{Thm}\, y$;

(ix) $\vdash_P \mathrm{Vble}\, x \to \mathrm{Thm}\, y \to \mathrm{Thm}\, \mathrm{Gen}\, xy$;

(x) $\vdash_P \mathrm{Thm}\, x \to \mathrm{Thm}\, \mathrm{Neg}\, x \to \mathrm{Fm}\, y \to \mathrm{Thm}\, y$;

(xi) $\vdash_P \mathrm{Con} \leftrightarrow \forall x(\mathrm{Fm}\, x \to \neg\mathrm{Thm}\, x \vee \neg\mathrm{Thm}\, \mathrm{Neg}\, x)$;

(xii) $\vdash_P \mathrm{Con} \leftrightarrow \forall x(\mathrm{Fm}\, x \to \mathrm{Cl}\, x \to \neg\mathrm{Thm}\, x \vee \neg\mathrm{Thm}\, \mathrm{Neg}\, x)$;

(xiii) $\vdash_P \mathrm{Con} \to \mathrm{Cm} \to \mathrm{Fm}\, x \to \mathrm{Cl}\, x \to \mathrm{Thm}\, \mathrm{Neg}\, x \leftrightarrow \neg\mathrm{Thm}\, x$.

We remark that (i) is a formalized commutativity rule. In order to derive it, we first remember how to prove commutativity rule: $\mathbf{B} \vee \mathbf{A}$ is the conclusion of a cut rule with premises $\mathbf{A} \vee \mathbf{B}$ and $\neg\mathbf{A} \vee \mathbf{A}$, and the latter is a propositional axiom. We have $\vdash_P \mathrm{Fm}\, x \to \mathrm{Pax}\, \mathrm{Imp}\, xx$ by definition of Pax, and $\vdash_P \mathrm{Fm}\, x \to \mathrm{Fm}\, y \to \mathrm{Cut}\, \mathrm{Disj}\, yx\, \mathrm{Disj}\, xy\, \mathrm{Imp}\, xx$ by the definitions of Cut. From the former, $\vdash_P \mathrm{Fm}\, x \to \mathrm{Thm}\, \mathrm{Imp}\, xx$, and from the latter and (iv) of lemma 1, $\vdash_P \mathrm{Fm}\, x \to \mathrm{Fm}\, y \to \mathrm{Thm}\, \mathrm{Disj}\, xy \to \mathrm{Thm}\, \mathrm{Imp}\, xx \to \mathrm{Thm}\, \mathrm{Disj}\, yx$. Now $\vdash_P \mathrm{Thm}\, \mathrm{Disj}\, xy \to \mathrm{Fm}\, x \wedge \mathrm{Fm}\, y$, so by the tautology theorem we find $\vdash_P \mathrm{Thm}\, \mathrm{Disj}\, xy \to \mathrm{Thm}\, \mathrm{Disj}\, yx$. Derivations of (ii)–(x) are found in the same way, using lemma 1. We derive (x) as a further example. By definition of Exp, $\vdash_P \mathrm{Fm}\, x \to \mathrm{Fm}\, y \to \mathrm{Exp}\, \mathrm{Disj}\, yxx \wedge \mathrm{Exp}\, \mathrm{Disj}\, y\, \mathrm{Neg}\, x\, \mathrm{Neg}\, x$, whence $\vdash_P \mathrm{Thm}\, x \to \mathrm{Thm}\, \mathrm{Neg}\, x \to \mathrm{Fm}\, y \to \mathrm{Thm}\, \mathrm{Disj}\, yx \wedge \mathrm{Thm}\, \mathrm{Disj}\, y\, \mathrm{Neg}\, x$ by (ii) of lemma 1. Using the formalized commutativity rule, we obtain

$$\vdash_P \mathrm{Thm}\, x \to \mathrm{Thm}\, \mathrm{Neg}\, x \to \mathrm{Fm}\, y \to \mathrm{Thm}\, \mathrm{Disj}\, xy \wedge \mathrm{Thm}\, \mathrm{Disj}\, \mathrm{Neg}\, xy. \tag{5}$$

By definition of Cut, $\vdash_P \mathrm{Fm}\,x \to \mathrm{Fm}\,y \to \mathrm{Cut}\,\mathrm{Disj}\,yy\,\mathrm{Disj}\,xy\,\mathrm{Disj}\,\mathrm{Neg}\,xy$, and by definition of Ctr, $\vdash_P \mathrm{Fm}\,y \to \mathrm{Ctr}\,y\,\mathrm{Disj}\,yy$. From these by (i) and (iv) of lemma 1, $\vdash_P \mathrm{Thm}\,x \to \mathrm{Thm}\,y \to \mathrm{Thm}\,\mathrm{Disj}\,xy \to \mathrm{Thm}\,\mathrm{Neg}\,xy \to \mathrm{Thm}\,y$. Together with (5), we obtain (x).

Let $\mathbf{a}$ be a variable-free term of P such that $\vdash_P \mathrm{Fm}\,\mathbf{a} \wedge \mathrm{Cl}\,\mathbf{a}$, for example $\langle \dot{\exists}, \mathrm{Vr}\,\dot{0}, \langle \dot{=}, \langle \mathrm{Vr}\,\dot{0}\rangle, \langle \mathrm{Vr}\,\dot{0}\rangle\rangle\rangle$. By the substitution theorem, $\vdash_P \forall x(\mathrm{Fm}\,x \to \mathrm{Cl}\,x \to \neg\mathrm{Thm}\,x \vee \neg\mathrm{Thm}\,\mathrm{Neg}\,x) \to \mathrm{Fm}\,\mathbf{a} \to \mathrm{Cl}\,\mathbf{a} \to \neg\mathrm{Thm}\,\mathbf{a} \vee \neg\mathrm{Thm}\,\mathrm{Neg}\,\mathbf{a}$, whence by the tautology theorem, $\vdash_P \forall x(\mathrm{Fm}\,x \to \mathrm{Cl}\,x \to \neg\mathrm{Thm}\,x \vee \neg\mathrm{Thm}\,\mathrm{Neg}\,x) \to \mathrm{Fm}\,\mathbf{a} \wedge \neg\mathrm{Thm}\,\mathbf{a} \vee \neg\mathrm{Thm}\,\mathrm{Neg}\,\mathbf{a}$. From this by the substitution axioms and the tautology theorem, we obtain the implication from right to left in (xii). The implication from left to right in (xi) follows from (x). Taken together, these implications prove (xi) and (xii). Finally, (xiii) follows from (xi) and the definition of Cm.

We call (ii) and its instances the *formalized detachment rule.*

Lemma 3. Let $\mathfrak{T}$ be an arithmetical theory in P. Then

$$\vdash_P \mathrm{Thm}\,x \to \mathrm{Vble}\,y \to \mathrm{Tm}\,z \to \mathrm{Subtl}\,xyz \to \mathrm{Thm}\,\mathrm{Sub}\,x\langle y\rangle\langle z\rangle.$$

Proof. From the definition of Sax and Sub, we infer $\vdash_P \mathrm{Fm}\,x \to \mathrm{Vble}\,y \to \mathrm{Tm}\,z \to \mathrm{Subtl}\,xyz \to \mathrm{Sax}\,\mathrm{Imp}\,\mathrm{Neg}\,\mathrm{Sub}\,x\langle y\rangle\langle z\rangle\,\mathrm{Inst}\,y\,\mathrm{Neg}\,x$. Hence by (iv) and the definition of Gen, $\vdash_P \mathrm{Fm}\,x \to \mathrm{Vble}\,y \to \mathrm{Tm}\,z \to \mathrm{Subtl}\,xyz \to \mathrm{Thm}\,\mathrm{Imp}\,\mathrm{Gen}\,yx\,\mathrm{Sub}\,x\langle y\rangle\langle z\rangle$. By the formalized detachment rule, $\vdash_P \mathrm{Fm}\,x \to \mathrm{Vble}\,y \to \mathrm{Tm}\,z \to \mathrm{Subtl}\,xyz \to \mathrm{Thm}\,\mathrm{Gen}\,yx \to \mathrm{Thm}\,\mathrm{Sub}\,x\langle y\rangle\langle z\rangle$, and by (ix), $\vdash_P \mathrm{Fm}\,x \to \mathrm{Vble}\,y \to \mathrm{Tm}\,z \to \mathrm{Subtl}\,xyz \to \mathrm{Thm}\,x \to \mathrm{Thm}\,\mathrm{Sub}\,x\langle y\rangle\langle z\rangle$. Since $\vdash_P \mathrm{Thm}\,x \to \mathrm{Fm}\,x$, the desired result follows by the tautology theorem. ☐

Lemma 4. Let $\mathfrak{T}$ be an arithmetical theory in P and let $\mathbf{h}$ be the function symbol defined in §2.2. Then

$$\vdash_P \mathrm{Thm}\,x \to \Sigma\mathrm{ub}\,yz \to \forall w(w < \mathrm{Len}\,y \to \mathrm{Subtl}\,x(y)_w(z)_w) \\ \to \forall w\forall w'(w < \mathrm{Len}\,y \to w' < w \to \neg\mathrm{Occ}(z)_{w'}(y)_w) \to \mathrm{Thm}\,\mathbf{h}xyz.$$

Proof. Let $\mathbf{h}'$ be defined as in §2.2. We shall prove $\vdash_P \mathbf{A}$ where $\mathbf{A}$ is

$$\vdash_P \mathrm{Thm}\,x \to \Sigma\mathrm{ub}\,yz \to \forall w(w < \mathrm{Len}\,y \to \mathrm{Subtl}\,x(y)_w(z)_w) \\ \to \forall w\forall w'(w < \mathrm{Len}\,y \to w' < w \to \neg\mathrm{Occ}(z)_{w'}(y)_w) \to w < \mathrm{Len}\,y \\ \to \mathrm{Subtl}\,\mathbf{h}'wxyz(y)_w(z)_w \wedge \mathrm{Thm}\,\mathbf{h}'\mathrm{S}wxyz$$

using the induction axioms. The conclusion follows by substituting $\mathrm{Len}\,y - \dot{1}$ for w in $\mathbf{A}$. Now since $\vdash_P \mathbf{h}'\dot{0}xyz = x$ and $\vdash_P \dot{0} < \mathrm{Len}\,y \to \forall w(w < \mathrm{Len}\,y \to \mathrm{Subtl}\,x(y)_w(z)_w) \to \mathrm{Subtl}\,x(y)_0(z)_0$, the first conjunct of $\mathbf{A}[w|\dot{0}]$ is derivable. Since $\vdash_P \mathbf{h}'\dot{1}xyz = \mathrm{Sub}\,x\langle(y)_0\rangle\langle(z)_0\rangle$, the second is derivable as well by lemma 3, so we have $\vdash_P \mathbf{A}[w|\dot{0}]$. Recall that $\vdash_P \mathbf{h}'\mathrm{S}wxyz = \mathrm{Sub}\,\mathbf{h}'wxyz\langle(y)_w\rangle\langle(z)_w\rangle$. If we substitute $\mathbf{h}'wxyz$ for x, $\langle(y)_w\rangle$ for y, $\langle(z)_w\rangle$ for z, $(y)_{\mathrm{S}w}$ for y', and $(z)_{\mathrm{S}w}$ for z' in (xvi′) of §2.2, we obtain

$$\vdash_P \mathrm{Des}\,\mathbf{h}'wxyz \to \Sigma\mathrm{ub}\langle(y)_w\rangle\langle(z)_w\rangle \to \mathrm{Vble}(y)_{\mathrm{S}w} \to \mathrm{Tm}(z)_{\mathrm{S}w} \\ \to \mathrm{Subtl}\,\mathbf{h}'wxyz(y)_w(z)_w \to (\mathrm{Occ}(z)_w(y)_{\mathrm{S}w} \to \mathrm{Subtl}\,x(y)_w(z)_{\mathrm{S}w}) \\ \to \mathrm{Subtl}\,\mathbf{h}'\mathrm{S}wxyz(y)_{\mathrm{S}w}(z)_{\mathrm{S}w}. \quad (6)$$

Now it is clear that $\vdash_P \mathbf{A} \to \mathrm{S}w < \mathrm{Len}\,y \to \mathrm{Des}\,\mathbf{h}'wxyz \wedge \Sigma\mathrm{ub}\langle(y)_w\rangle\langle(z)_w\rangle \wedge \mathrm{Vble}(y)_{\mathrm{S}w} \wedge \mathrm{Tm}(z)_{\mathrm{S}w} \wedge (\mathrm{Occ}(z)_w(y)_{\mathrm{S}w} \to \mathrm{Subtl}\,x(y)_w(z)_{\mathrm{S}w})$, so by (6) and the tautology theorem,

$$\vdash_P \mathbf{A} \to \mathrm{S}w < \mathrm{Len}\,y \to \mathrm{Subtl}\,\mathbf{h}'\mathrm{S}wxyz(y)_{\mathrm{S}w}(z)_{\mathrm{S}w}. \quad (7)$$

By this and lemma 3, we have $\vdash_P \mathbf{A} \to \mathrm{S}w < \mathrm{Len}\,y \to \mathrm{Thm}\,\mathrm{Sub}\,\mathbf{h}'\mathrm{S}wxyz\langle(y)_{\mathrm{S}w}\rangle\langle(z)_{\mathrm{S}w}\rangle$, but by definition of $\mathbf{h}'$ this gives

$$\vdash_P \mathbf{A} \to \mathrm{S}w < \mathrm{Len}\,y \to \mathrm{Thm}\,\mathbf{h}'\mathrm{SS}wxyz. \quad (8)$$

From (7) and (8) by the tautology theorem, $\vdash_P \mathbf{A} \to \mathbf{A}[w|\mathrm{S}w]$. By the induction axioms, $\vdash_P \mathbf{A}$. ☐

Formalized Substitution Rule. Let $\mathfrak{T}$ be an arithmetical theory in P. Then

$$\vdash_P \mathrm{Thm}\,x \to \Sigma\mathrm{ub}\,yz \to \forall w(w < \mathrm{Len}\,y \to \mathrm{Subtl}\,x(y)_w(z)_w) \to \mathrm{Thm}\,\mathrm{Sub}\,xyz.$$

Proof. Let $\mathbf{h}$ be the function symbol defined in §2.2. Define $\mathbf{f}xyz = \mu w(\mathrm{Len}\,w = \mathrm{Len}\,y \wedge \forall x'(x' < \mathrm{Len}\,y \to (w)_{x'} = \langle \mathrm{Vr}((x * (y * z)) + \mathrm{S}x')\rangle))$. Then clearly $\vdash_P \mathrm{Len}\,\mathbf{f}xyz = \mathrm{Len}\,y$ and

$$\vdash_P \forall w(w < \mathrm{Len}\,y \to \mathrm{Vble}(\mathbf{f}xyz)_w \wedge \forall w'(w' < w \to (\mathbf{f}xyz)_w \neq (\mathbf{f}xyz)_{w'})), \tag{9}$$

and so, by definition of Σub,

$$\vdash_P \Sigma\mathrm{ub}\,yz \to \Sigma\mathrm{ub}\,y\mathbf{f}xyz \wedge \Sigma\mathrm{ub}\,\mathbf{f}xyzz. \tag{10}$$

and by (v′) of §2.2, $\vdash_P w < \mathrm{Len}\,y \to w' < \mathrm{Len}\,y \to \mathrm{Occ}(\mathbf{f}xyz)_{w'}(\mathbf{f}xyz)_w \to w = w'$, whence

$$\vdash_P w < \mathrm{Len}\,y \to w' < \mathrm{Len}\,y \to \forall w(w < \mathrm{Len}\,y \to \mathrm{Subtl}\,x(y)_w(z)_w) \\ \to \mathrm{Occ}(\mathbf{f}xyz)_{w'}(\mathbf{f}xyz)_w \to \mathrm{Subtl}\,x(y)_w(z)_{w'}. \tag{11}$$

By (i′), (iv′), and (v′) of §2.2 and the definition of $\mathbf{f}$, we have

$$\vdash_P w < \mathrm{Len}\,y \to w' < \mathrm{Len}\,y \to \neg\,\mathrm{Occ}\,x(\mathbf{f}xyz)_w \wedge \neg\,\mathrm{Occ}(\mathbf{f}xyz)_w(y)_{w'} \wedge \neg\,\mathrm{Occ}(z)_w(\mathbf{f}xyz)_{w'}, \tag{12}$$

whence by (v′) and (viii′) of §2.2,

$$\vdash_P \mathrm{Des}\,x \to \Sigma\mathrm{ub}\,yz \to w < \mathrm{Len}\,y \to \mathrm{Subtl}\,x(y)_w(\mathbf{f}xyz)_w. \tag{13}$$

From (10), (12), and (13) by (vii′) and (xvii′) of §2.2,

$$\vdash_P \mathrm{Des}\,x \to \Sigma\mathrm{ub}\,yz \to \mathrm{Sub}\,\mathrm{Sub}\,xy\mathbf{f}xyz\mathbf{f}xyzz = \mathrm{Sub}\,xyz. \tag{14}$$

From (10) and (12) by (xviii′) of §2.2,

$$\vdash_P \mathrm{Des}\,x \to \Sigma\mathrm{ub}\,yz \to \mathrm{Sub}\,xy\mathbf{f}xyz = \mathbf{h}xy\mathbf{f}xyz \text{ and} \tag{15}$$

$$\vdash_P \mathrm{Des}\,x \to \Sigma\mathrm{ub}\,yz \to \mathrm{Sub}\,\mathrm{Sub}\,xy\mathbf{f}xyz\mathbf{f}xyzz = \mathbf{hh}xy\mathbf{f}xyz\mathbf{f}xyzz. \tag{16}$$

From (14) and (16),

$$\vdash_P \mathrm{Des}\,x \to \Sigma\mathrm{ub}\,yz \to \mathrm{Sub}\,xyz = \mathbf{hh}xy\mathbf{f}xyz\mathbf{f}xyzz. \tag{17}$$

In view of this we need only prove

$$\vdash_P \mathrm{Thm}\,x \to \Sigma\mathrm{ub}\,yz \to \forall w(w < \mathrm{Len}\,y \to \mathrm{Subtl}\,x(y)_w(z)_w) \to \mathrm{Thm}\,\mathbf{hh}xy\mathbf{f}xyz\mathbf{f}xyzz. \tag{18}$$

If we substitute $\mathbf{f}xyz$ for z, $(\mathbf{f}xyz)_{w'}$ for y', and $(z)_{w'}$ for z' in (xvi′) of §2.2, we obtain

$$\vdash_P \mathrm{Des}\,x \to \Sigma\mathrm{ub}\,y\mathbf{f}xyz \to \mathrm{Vble}(\mathbf{f}xyz)_{w'} \to \mathrm{Tm}(z)_{w'} \\ \to \forall w(w < \mathrm{Len}\,y \to \mathrm{Subtl}\,x(y)_w(\mathbf{f}xyz)_w) \to \forall w(w < \mathrm{Len}\,y \to \mathrm{Occ}(\mathbf{f}xyz)_w(\mathbf{f}xyz)_{w'} \\ \to \mathrm{Subtl}\,x(y)_w(z)_{w'}) \to \mathrm{Subtl}\,\mathrm{Sub}\,xy\mathbf{f}xyz(\mathbf{f}xyz)_{w'}(z)_{w'}$$

From this by (9), (10), (11), and (13),

$$\vdash_P \mathrm{Des}\,x \to \Sigma\mathrm{ub}\,yz \to \forall w(w < \mathrm{Len}\,y \to \mathrm{Subtl}\,x(y)_w(z)_w) \\ \to w' < \mathrm{Len}\,y \to \mathrm{Subtl}\,\mathbf{h}xy\mathbf{f}xyz(\mathbf{f}xyz)_{w'}(z)_{w'}. \tag{19}$$

From (12), (13) and (19), we obtain (18) by two instances of lemma 4. ☐

2.7 Arithmetical interpretations. We have introduced so far formalizations of the notions of first-order language and first-order theory. We are now going to introduce a notion corresponding to interpretations. For obvious reasons we shall only formalize interpretations that have 0 parameters. Let P be a good extension of PA and $\mathfrak{L}$ and $\mathfrak{L}'$ arithmetical languages in P with n parameters such that $\vdash_P \Omega x^n \leftrightarrow \Omega' x^n$. An *arithmetical interpretation* $\mathfrak{I}$ *in* P *of* $\mathfrak{L}$ *in* $\mathfrak{L}'$ consists of an n-ary function symbol $\mathfrak{U}_{\mathfrak{I}}$ of P called the *universe* of $\mathfrak{I}$ and an $(n+1)$-ary function symbol $()_{\mathfrak{I}}$ of P such that

(i) $\vdash_P \Omega x^n \to \mathrm{Pred}'\,\mathfrak{U}_{\mathfrak{I}} x^n \dot{1} x^n$;

(ii) $\vdash_P \Omega x^n \to ()_{\mathfrak{I}}\,\mathrm{Vr}\,xx^n x^n = \mathrm{Vr}'\,xx^n$;

(iii) $\vdash_P \Omega x^n \to \mathrm{Func}\, x y x^n \to \mathrm{Func}' ()_{\mathfrak{I}} x x^n y x^n$;

(iv) $\vdash_P \Omega x^n \to \mathrm{Pred}\, x y x^n \to \mathrm{Pred}' ()_{\mathfrak{I}} x x^n y x^n$.

We take n to be 0 from now on, and we abbreviate $()_{\mathfrak{I}}\mathbf{a}$ to $(\mathbf{a})_{\mathfrak{I}}$. We define a unary function symbol $\mathbf{f}$ by recursion so that

$$\begin{aligned}\vdash_P y = \mathbf{f}x \leftrightarrow \mathrm{Sq}\, y &\\ &\wedge (\mathrm{Tm}\, x \wedge \mathrm{Len}\, y = \mathrm{Len}\, x \wedge (y)_0 = ((x)_0)_{\mathfrak{I}} \wedge \forall z(z < \mathrm{Len}\, x - \dot{1} \to (y)_{\mathrm{S}z} = \mathbf{f}(x)_{\mathrm{S}z}))\\ &\vee (\mathrm{Atfm}\, x \wedge \mathrm{Len}\, y = \mathrm{Len}\, x \wedge (y)_0 = ((x)_0)_{\mathfrak{I}} \wedge \forall z(z < \mathrm{Len}\, x - \dot{1} \to (y)_{\mathrm{S}z} = \mathbf{f}(x)_{\mathrm{S}z}))\\ &\vee (\mathrm{Fm}\, x \wedge x = \mathrm{Disj}(x)_1(x)_2 \wedge y = \mathrm{Disj}'\, \mathbf{f}(x)_1 \mathbf{f}(x)_2)\\ &\vee (\mathrm{Fm}\, x \wedge x = \mathrm{Neg}(x)_1 \wedge y = \mathrm{Neg}'\, \mathbf{f}(x)_1)\\ &\vee (\mathrm{Fm}\, x \wedge x = \mathrm{Inst}(x)_1(x)_2 \wedge y = \mathrm{Inst}'((x)_1)_{\mathfrak{I}}\, \mathrm{Cnj}'\langle \mathfrak{U}_{\mathfrak{I}}, ((x)_1)_{\mathfrak{I}}\rangle \mathbf{f}(x)_2)\\ &\vee (\neg\, \mathrm{Des}\, x \wedge y = \dot{0}).\end{aligned}$$

Define the unary function symbol $\mathbf{g}$ by $\mathbf{g}x = \mu y(\mathrm{Cl}'\, x \vee \mathrm{Fr}'\, xy)$, and the binary function symbol $\mathbf{h}$ by recursion so that $\vdash_P z = \mathbf{h}xy \leftrightarrow ((y = \dot{0} \vee \mathrm{Cl}'\, x \vee \neg\, \mathrm{Fm}'\, x) \wedge z = x) \vee (y \neq \dot{0} \wedge \neg\, \mathrm{Cl}'\, x \wedge \mathrm{Fm}'\, x \wedge z = \mathrm{Imp}'\langle \mathfrak{U}_{\mathfrak{I}}, \mathbf{gh}x(y - \dot{1})\rangle \mathbf{h}x(y - \dot{1}))$.

We then define a unary function symbol Int by $\mathrm{Int}\, x = \mathbf{hf}xx$. Intuitively, $\mathrm{Int}\, x$ corresponds to the interpretation of the designator x by $\mathfrak{I}$.

An arithmetical interpretation $\mathfrak{I}$ of $\mathfrak{L}$ in $\mathfrak{L}(\mathfrak{T}')$ is an *arithmetical interpretation of* $\mathfrak{L}$ *in* $\mathfrak{T}'$ if $\vdash_P \mathrm{Thm}'\, \mathrm{Inst}'\langle \mathrm{Vr}'\, \dot{0}\rangle\langle \mathfrak{U}_{\mathfrak{I}}, \langle \mathrm{Vr}'\, \dot{0}\rangle\rangle$ and $\vdash_P \mathrm{Func}\, xy \to \mathrm{Thm}'\, \mathbf{h}\mu z(\mathrm{Len}\, z = \mathrm{SS}y \wedge (z)_0 = \mathfrak{U}_{\mathfrak{I}} \wedge (z)_1 = (x)_{\mathfrak{I}} \wedge \forall w(w < y \to (z)_{\mathrm{SS}w} = \langle \mathrm{Vr}'\, w\rangle))y$.

Let $\mathfrak{I}$ be an arithmetical interpretation of $\mathfrak{L}(\mathfrak{T})$ in $\mathfrak{L}(\mathfrak{T}')$. We define a new arithmetical theory $\mathfrak{T}_{\mathfrak{I}}$: the language of $\mathfrak{T}_{\mathfrak{I}}$ is $\mathfrak{L}(\mathfrak{T})$ and $\mathrm{Nlax}_{\mathfrak{I}}$ is given by

$$\mathrm{Nlax}_{\mathfrak{I}}\, x \leftrightarrow \mathrm{Nlax}\, x \wedge \mathrm{Thm}'\, \mathrm{Int}\, x.$$

ARITHMETICAL INTERPRETATION THEOREM. If $\mathfrak{I}$ is an arithmetical interpretation of $\mathfrak{L}(\mathfrak{T})$ in $\mathfrak{T}'$ and if $\vdash_P \mathrm{Iax}\, x \vee \mathrm{Feax}\, x \vee \mathrm{Peax}\, x \to \mathrm{Thm}'\, \mathrm{Int}\, x$, then $\vdash_P \mathrm{Thm}_{\mathfrak{I}}\, x \to \mathrm{Thm}'\, \mathrm{Int}\, x$.

Proof. Just believe it. □

COROLLARY. If $\mathfrak{I}$ is an arithmetical interpretation of $\mathfrak{L}(\mathfrak{T})$ in $\mathfrak{T}'$ and if $\vdash_P \mathrm{Iax}\, x \vee \mathrm{Feax}\, x \vee \mathrm{Peax}\, x \to \mathrm{Thm}'\, \mathrm{Int}\, x$, then $\vdash_P \mathrm{Con}' \to \mathrm{Con}_{\mathfrak{I}}$.

Proof. We have $\vdash_P \mathrm{Cl}\, x \to \mathrm{Cl}'\, \mathrm{Int}\, x$, $\vdash_P \mathrm{Fm}\, x \to \mathrm{Cl}\, x \to \mathrm{Int}\, \mathrm{Neg}\, x = \mathrm{Neg}\, \mathrm{Int}\, x$, and by the arithmetical interpretation theorem, $\vdash_P \neg\, \mathrm{Thm}'\, \mathrm{Int}\, x \to \neg\, \mathrm{Thm}_{\mathfrak{I}}\, x$. From these formulae we obtain $\vdash_P \mathrm{Fm}\, x \to \mathrm{Cl}\, x \to \neg\, \mathrm{Thm}'\, \mathrm{Int}\, x \vee \neg\, \mathrm{Thm}'\, \mathrm{Neg}\, \mathrm{Int}\, x \to \mathrm{Con}_{\mathfrak{I}}$ by the substitution axioms. Thus by the substitution theorem, $\vdash_P \forall x(\mathrm{Fm}\, x \to \mathrm{Cl}\, x \to \neg\, \mathrm{Thm}'\, x \vee \neg\, \mathrm{Thm}'\, \mathrm{Neg}\, x) \to \mathrm{Con}_{\mathfrak{I}}$, whence $\vdash_P \mathrm{Con}' \to \mathrm{Con}_{\mathfrak{I}}$ by (xii) of §2.6. □

An arithmetical interpretation $\mathfrak{I}$ of $\mathfrak{L}(\mathfrak{T})$ in $\mathfrak{T}$ is an *arithmetical interpretation of* $\mathfrak{T}$ *in* $\mathfrak{T}'$ if $\vdash_P \mathrm{Iax}\, x \vee \mathrm{Feax}\, x \vee \mathrm{Peax}\, x \vee \mathrm{Nlax}\, x \to \mathrm{Thm}'\, \mathrm{Int}\, x$. In this case $\vdash_P \mathrm{Nlax}_{\mathfrak{I}}\, x \leftrightarrow \mathrm{Nlax}\, x$ and the arithmetical interpretation theorem shows that $\vdash_P \mathrm{Thm}\, x \to \mathrm{Thm}'\, \mathrm{Int}\, x$, and the corollary that $\vdash_P \mathrm{Con}' \to \mathrm{Con}$.

PROPOSITION. If $\vdash_P (\dot{=})_{\mathfrak{I}} = \dot{=}'$, then $\vdash_P \mathrm{Iax}\, x \vee \mathrm{Feax}\, x \vee \mathrm{Peax}\, x \to \mathrm{Thm}'\, \mathrm{Int}\, x$.

Proof. [...exercise!] □

§3 Extensional application

3.1 Conventions. For the rest of this chapter, we fix a coding function β, a good extension P of PA, and B a coding function symbol in P representing β in P. The letter L with superscripts or subscripts is used of a first-order language arithmetized from a numerotation by the coding function β. This numerotation is written σ and its associated functions and predicates vr, func, and pred, with the same superscripts and subscripts. The letter T with superscripts or subscripts is used of a first-order theory whose language is denoted by the letter L with the same superscripts and subscripts. An arithmetical language or theory, unless otherwise stated, has no parameters.

3.2 Describing languages. Let $\mathfrak{L}$ be an arithmetical language in P. We let $L(\mathfrak{L})$ be the extension of $L(\mathrm{N})$ containing B, Len, Sq, $\in$, $*$, Ini, Rmv, $\langle\rangle_n$ for each n, the symbols of $\mathfrak{L}$, and the symbols (i)–(xxv) of §2.2. In applications, we wish to interpret the symbols of $L(\mathfrak{L})$ as describing an actual first-order language L. We say that $\mathfrak{L}$ *describes L in P* if

(i) Vr represents vr in P;

(ii) Func positively represents func in P;

(iii) Pred positively represents pred in P;

(iv) $\dot\vee$, $\dot\neg$, $\dot\exists$, and $\doteq$ respectively represent $\sigma(\vee)$, $\sigma(\neg)$, $\sigma(\exists)$, and $\sigma(=)$ in P, and in the case that L is numerical, $\dot{\dot 0}$ and $\dot{\mathrm{S}}$ represent $\sigma(\dot 0)$ and $\sigma(\mathrm{S})$ in P, respectively.

We say that $\mathfrak{L}$ *represents L in P* if (i) and (iv) holds and if Func and Pred respectively represent func and pred in P. Note that if $\mathfrak{L}$ describes L and L is numerical, then $\mathfrak{L}$ is numerical.

We would like to know how the nonlogical symbols of $L(\mathfrak{L})$ behave when $\mathfrak{L}$ describes or represents L. For instance, there is an obvious similarity between these symbols and the predicates of ch. III §3.1, which we suggestively indicated by using the same names with a capital letter. We shall now prove that, if $\mathfrak{L}$ describes (resp. represents) L, almost anything (resp. anything) that should be true is true.

Given a first-order language L arithmetized from a numerotation, we define a numerical realization α of $L(\mathfrak{L})$: α extends ν; B_α is β; Len_α is len; $\mathrm{Sq}^\pm_\alpha$ are sq; $\in^\pm_\alpha$ are $\in$; $*_\alpha$ is $*$; Ini_α is ini; Rmv_α is rmv; $(\langle\rangle_n)_\alpha$ is the n-ary sequence function; Vr_α is vr; Func^+_α is func; Pred^+_α is pred; $\dot\vee_\alpha$ is $\sigma(\vee)$; $\dot\neg_\alpha$ is $\sigma(\neg)$; $\dot\exists_\alpha$ is $\sigma(\exists)$; $\doteq_\alpha$ is $\sigma(=)$; $\dot{\dot 0}_\alpha$ is $\sigma(\dot 0)$; $\dot{\mathrm{S}}_\alpha$ is $\sigma(\mathrm{S})$; Sym^+_α is sym; if $\mathbf{p}$ is among (ii)–(x) or (xiii)–(xxiii), $\mathbf{p}^+_\alpha$ is the uncapitalized counterpart of $\mathbf{p}$ of ch. III §3.1; Vble^-_α is vble_L; Atfm^-_α is $\mathrm{atfm}_L \vee \neg\,\mathrm{fm}_L$; $\mathrm{Occ}^-_\alpha(a,b) \leftrightarrow \mathrm{occ}_L(a,b) \vee \neg\,\mathrm{des}_L(a) \vee \neg\,\mathrm{des}_L(b)$; $\mathrm{Fr}^-_\alpha(a,b) \leftrightarrow \mathrm{fr}_L(a,b) \vee \neg\,\mathrm{des}_L(a)$; Cl^-_α is $\mathrm{cl}_L \vee \neg\,\mathrm{des}_L$; $\mathrm{Subtl}^-_\alpha(a,b,c) \leftrightarrow \mathrm{subtl}_L(a,b,c) \vee \neg\,\mathrm{des}_L(a) \vee \neg\,\mathrm{tm}_L(c)$; if $\mathbf{p}$ ends in x, then $\mathbf{p}^-_\alpha$ is $\mathbf{p}^+_\alpha \vee \neg\,\mathrm{fm}_L$; if $\mathbf{p}$ is Ctr, Exp, Assoc, or Intr, $\mathbf{p}^-_\alpha(a,b) \leftrightarrow \mathbf{p}^+_\alpha(a,b) \vee \neg\,\mathrm{fm}_L(a) \vee \neg\,\mathrm{fm}_L(b)$; $\mathrm{Cut}^-_\alpha(a,b,c) \leftrightarrow \mathrm{cut}_L(a,b,c) \vee \neg\,\mathrm{fm}_L(a) \vee \neg\,\mathrm{fm}_L(b) \vee \neg\,\mathrm{fm}_L(c)$; Sub_α and Num_α are the functions sub and num defined in the proof of the theorem of ch. III §3.2; $\Sigma\mathrm{ub}^+_\alpha(a,b)$ if and only if $a = \langle a_1,\dots,a_n\rangle$, $b = \langle b_1,\dots,b_m\rangle$, $n = m$, $\mathrm{vble}_L(a_i)$ for all i, $\mathrm{tm}_L(b_i)$ for all i, and the a_i are pairwise distinct; $\Sigma\mathrm{ub}^-_\alpha(a,b)$ if and only if $\Sigma\mathrm{ub}^+_\alpha(a,b)$ or $\neg\,\mathrm{tm}_L((b)_i)$ for some $i < \mathrm{len}(b)$; $\mathrm{Inf}^+_\alpha(a,b)$ if and only if $\mathrm{sq}(b)$ and a is the expression number of the conclusion of a rule of inference for L whose premises have expression numbers in b; unspecified n-ary predicates $\mathbf{p}^-_\alpha$ are $\top^n$. Let τ be the numerical realization of $L(\mathfrak{L})$ for which $\mathbf{f}_\tau$ is $\mathbf{f}_\alpha$, $\mathrm{Sym}^\pm_\tau$ is $\mathrm{Sym}^\pm_\alpha$, and $\mathbf{p}^\pm_\tau$ are both $\mathbf{p}^+_\alpha$ for $\mathbf{p}$ other than Sym.

THEOREM. If $\mathfrak{L}$ describes L, α is faithful in P. If $\mathfrak{L}$ represents L, τ is faithful in P.

Proof. This is an application of the theorem on RE-formulae. We start from the realization ν of $L(\mathrm{N})$ which is known to be faithful in P, and we extend it one symbol at a time starting from B which is assumed to represent β, applying the theorem on RE-formulae to the previous extension to obtain the representability conditions for the new symbol. In all cases the relevant RE-formula or PR-formula is given by the defining axiom of the symbol. □

3.3 A property of Sub. Let $\mathfrak{L}$ describe L in P. There is one more result which is not purely extensional in nature that we must establish. It concerns the function symbol Sub when the third argument is variable. Let $\mathbf{u}$ be a designator of L, $\mathbf{x}_1$, …, $\mathbf{x}_n$ distinct variables, and $\mathbf{b}_1$, …, $\mathbf{b}_n$ terms of P. We define a term $\hat{\mathbf{u}}\{\mathbf{x}_1,\dots,\mathbf{x}_n|\mathbf{b}_1,\dots,\mathbf{b}_n\}$ of P by induction on the length of $\mathbf{u}$ as follows. If $\mathbf{u}$ is $\mathbf{x}_i$ for some i, $\hat{\mathbf{u}}\{\mathbf{x}_1,\dots,\mathbf{x}_n|\mathbf{b}_1,\dots,\mathbf{b}_n\}$ is $\mathbf{b}_i$. Otherwise, let $\mathbf{u}$ be $\mathbf{s}\mathbf{u}_1\dots\mathbf{u}_m$ where $\mathbf{s}$ is a symbol of index m and $\mathbf{u}_1$, …, $\mathbf{u}_m$ are designators. Then $\hat{\mathbf{u}}\{\mathbf{x}_1,\dots,\mathbf{x}_n|\mathbf{b}_1,\dots,\mathbf{b}_n\}$ is

$$\langle\dot\sigma(\mathbf{s}), \hat{\mathbf{u}}_1\{\mathbf{x}_1,\dots,\mathbf{x}_n|\mathbf{b}_1,\dots,\mathbf{b}_n\},\dots,\hat{\mathbf{u}}_m\{\mathbf{x}_1,\dots,\mathbf{x}_n|\mathbf{b}_1,\dots,\mathbf{b}_n\}\rangle,$$

unless $\mathbf{u}$ is $\exists\mathbf{x}_i\mathbf{A}$ for some i, in which case it is

$$\langle\dot\sigma(\exists), \langle\dot\sigma(\mathbf{x}_i)\rangle, \hat{\mathbf{A}}\{\mathbf{x}_1,\dots,\mathbf{x}_{i-1},\mathbf{x}_{i+1},\dots,\mathbf{x}_n|\mathbf{b}_1,\dots,\mathbf{b}_{i-1},\mathbf{b}_{i+1},\dots,\mathbf{b}_n\}\rangle.$$

For example, if $\mathbf{u}$ is $\mathbf{f}x\mathbf{g}y$, then $\hat{\mathbf{u}}\{x,z|\mathbf{b}_1,\mathbf{b}_2\}$ is $\langle\dot\sigma(\mathbf{f}),\mathbf{b}_1,\langle\dot\sigma(\mathbf{g}),\langle\dot\sigma(y)\rangle\rangle\rangle$. When $n = 0$, an easy induction on the length of $\mathbf{u}$, using the equality theorem and the fact that $\langle\rangle_n$ represents the n-ary sequence function, shows that

$$\vdash_P \ulcorner\mathbf{u}\urcorner = \hat{\mathbf{u}}.$$

We now prove that

$$\vdash_P \mathrm{Sub}\ulcorner\mathbf{u}\urcorner\langle\ulcorner\mathbf{x}_1\urcorner,\dots,\ulcorner\mathbf{x}_n\urcorner\rangle\langle\mathbf{b}_1,\dots,\mathbf{b}_n\rangle = \hat{\mathbf{u}}\{\mathbf{x}_1,\dots,\mathbf{x}_n|\mathbf{b}_1,\dots,\mathbf{b}_n\} \tag{1}$$

by induction on the length of $\mathbf{u}$. We let $\mathbf{f}$ and $\mathbf{g}$ be the defined function symbols of P as in §2.2. We can extend α by setting $\mathbf{f}_\alpha(a,b) = \mu i(\neg(a \in b) \vee a = (b)_i)$ and $\mathbf{g}_\alpha(a,b) = \mu i((a)_0 \neq \sigma(\exists) \vee \neg((a)_1 \in b) \vee (i \neq 0 \wedge (a)_1 = \beta(b,i)))$. Then α remains faithful by the theorem on RE-formulae. Suppose first that $\mathbf{u}$ is $\mathbf{x}_{k+1}$ for some $k < n$. Then by faithfulness, $\vdash_P \mathrm{Vble}\ulcorner\mathbf{u}\urcorner$, $\vdash_P \ulcorner\mathbf{u}\urcorner \in \langle\ulcorner\mathbf{x}_1\urcorner,\dots,\ulcorner\mathbf{x}_n\urcorner\rangle$, and $\vdash_P \mathbf{f}\ulcorner\mathbf{u}\urcorner\langle\ulcorner\mathbf{x}_1\urcorner,\dots,\ulcorner\mathbf{x}_n\urcorner\rangle = \dot{k}$, so by definition of Sub,

$$\vdash_P \mathrm{Sub}\ulcorner\mathbf{u}\urcorner\langle\ulcorner\mathbf{x}_1\urcorner,\dots,\ulcorner\mathbf{x}_n\urcorner\rangle\langle\mathbf{b}_1,\dots,\mathbf{b}_n\rangle = (\langle\mathbf{b}_1,\dots,\mathbf{b}_n\rangle)_{\dot{k}},$$

and since $\vdash_P (\langle\mathbf{b}_1,\dots,\mathbf{b}_n\rangle)_{\dot{k}} = \mathbf{b}_{k+1}$, this proves (1).

Suppose now that $\mathbf{u}$ is a variable which is not among the $\mathbf{x}_i$. By faithfulness, $\vdash_P \mathrm{Vble}\ulcorner\mathbf{u}\urcorner$ and $\vdash_P \neg(\ulcorner\mathbf{u}\urcorner \in \langle\ulcorner\mathbf{x}_1\urcorner,\dots,\ulcorner\mathbf{x}_n\urcorner\rangle)$, so the definition of Sub gives

$$\vdash_P \mathrm{Sub}\ulcorner\mathbf{u}\urcorner\langle\ulcorner\mathbf{x}_1\urcorner,\dots,\ulcorner\mathbf{x}_n\urcorner\rangle\langle\mathbf{b}_1,\dots,\mathbf{b}_n\rangle = \ulcorner\mathbf{u}\urcorner,$$

whence (1).

Suppose that $\mathbf{u}$ is $\mathbf{s}\mathbf{u}_1\dots\mathbf{u}_m$ where $\mathbf{s}$ is a symbol of L of index m that is not a variable and $\mathbf{u}_1,\dots,\mathbf{u}_m$ are designators of L. We consider two cases. Suppose first that either $\mathbf{s}$ is not $\exists$ or $\mathbf{u}_1$ is not among the $\mathbf{x}_i$. Then by faithfulness $\vdash_P \mathbf{g}\ulcorner\mathbf{u}\urcorner\langle\ulcorner\mathbf{x}_1\urcorner,\dots,\ulcorner\mathbf{x}_n\urcorner\rangle = \dot{0}$, and so by definition of Rmv, $\vdash_P \mathrm{Rmv}\,x\mathbf{g}\ulcorner\mathbf{u}\urcorner\langle\ulcorner\mathbf{x}_1\urcorner,\dots,\ulcorner\mathbf{x}_n\urcorner\rangle = x$. Using the representabiliy of len by Len, the definition of Sub gives

$$\begin{aligned}\vdash_P \mathrm{Sub}\ulcorner\mathbf{u}\urcorner\langle\ulcorner\mathbf{x}_1\urcorner,\dots,\ulcorner\mathbf{x}_n\urcorner\rangle\langle\mathbf{b}_1,\dots,\mathbf{b}_n\rangle &= \mu w(\mathrm{Len}\,w = \mathrm{S}\dot{m} \wedge (w)_0 = \dot{\sigma}(\mathbf{s}) \\ &\wedge \forall x'(x' < \dot{m} \to (w)_{\mathrm{S}x'} = \mathrm{Sub}(\ulcorner\mathbf{u}\urcorner)_{\mathrm{S}x'}\langle\ulcorner\mathbf{x}_1\urcorner,\dots,\ulcorner\mathbf{x}_n\urcorner\rangle\langle\mathbf{b}_1,\dots,\mathbf{b}_n\rangle)).\end{aligned} \tag{2}$$

For $k < m$, $\vdash_P (\ulcorner\mathbf{u}\urcorner)_{\mathrm{S}\dot{k}} = \ulcorner\mathbf{u}_k\urcorner$ by representability, so from (2), (ix) of ch. III §5.2, and the induction hypothesis, we obtain

$$\begin{aligned}\vdash_P \mathrm{Sub}\ulcorner\mathbf{u}\urcorner\langle\ulcorner\mathbf{x}_1\urcorner,\dots,\ulcorner\mathbf{x}_n\urcorner\rangle\langle\mathbf{b}_1,\dots,\mathbf{b}_n\rangle &= \mu w(\mathrm{Len}\,w = \mathrm{S}\dot{m} \wedge (w)_0 = \dot{\sigma}(\mathbf{s}) \\ &\wedge (w)_{\dot{1}} = \hat{\mathbf{u}}_1\{\mathbf{x}_1,\dots,\mathbf{x}_n|\mathbf{b}_1,\dots,\mathbf{b}_n\} \wedge \dots \wedge (w)_{\dot{m}} = \hat{\mathbf{u}}_m\{\mathbf{x}_1,\dots,\mathbf{x}_n|\mathbf{b}_1,\dots,\mathbf{b}_n\}).\end{aligned}$$

But the right-hand side is just the definition of $\langle\rangle_{m+1}$, so that

$$\vdash_P \mathrm{Sub}\ulcorner\mathbf{u}\urcorner\langle\ulcorner\mathbf{x}_1\urcorner,\dots,\ulcorner\mathbf{x}_n\urcorner\rangle\langle\mathbf{b}_1,\dots,\mathbf{b}_n\rangle = \langle\dot{\sigma}(\mathbf{s}), \hat{\mathbf{u}}_1\{\mathbf{x}_1,\dots,\mathbf{x}_n|\mathbf{b}_1,\dots,\mathbf{b}_n\},\dots,\hat{\mathbf{u}}_m\{\mathbf{x}_1,\dots,\mathbf{x}_n|\mathbf{b}_1,\dots,\mathbf{b}_n\}\rangle.$$

Now the term on the right is exactly $\hat{\mathbf{u}}\{\mathbf{x}_1,\dots,\mathbf{x}_n|\mathbf{b}_1,\dots,\mathbf{b}_n\}$. It remains to consider the case where $\mathbf{s}$ is $\exists$ and $\mathbf{u}_1$ is $\mathbf{x}_{k+1}$ for some $k < n$. By faithfulness, $\vdash_P \mathbf{g}\ulcorner\mathbf{u}\urcorner\langle\ulcorner\mathbf{x}_1\urcorner,\dots,\ulcorner\mathbf{x}_n\urcorner\rangle = \mathrm{S}\dot{k}$. By properties of Rmv, $\vdash_P \mathrm{Rmv}\langle x_1,\dots,x_n\rangle\mathrm{S}\dot{k} = \langle x_1,\dots,x_k,x_{k+2},\dots,x_n\rangle$. From this by induction hypothesis,

$$\begin{aligned}&\vdash_P \mathrm{Sub}\ulcorner\mathbf{u}_r\urcorner\,\mathrm{Rmv}\langle\ulcorner\mathbf{x}_1\urcorner,\dots,\ulcorner\mathbf{x}_n\urcorner\rangle\mathbf{g}\ulcorner\mathbf{u}\urcorner\langle\ulcorner\mathbf{x}_1\urcorner,\dots,\ulcorner\mathbf{x}_n\urcorner\rangle\,\mathrm{Rmv}\langle\mathbf{b}_1,\dots,\mathbf{b}_n\rangle\mathbf{g}\ulcorner\mathbf{u}\urcorner\langle\ulcorner\mathbf{x}_1\urcorner,\dots,\ulcorner\mathbf{x}_n\urcorner\rangle \\ &\qquad = \hat{\mathbf{u}}_r\{\mathbf{x}_1,\dots,\mathbf{x}_k,\mathbf{x}_{k+2},\dots,\mathbf{x}_n|\mathbf{b}_1,\dots,\mathbf{b}_k,\mathbf{b}_{k+2},\dots,\mathbf{b}_n\}.\end{aligned} \tag{3}$$

When we inject this in the definition of Sub, we obtain as above with (ix) of ch. III §5.2

$$\begin{aligned}\vdash_P \mathrm{Sub}\ulcorner\mathbf{u}\urcorner\langle\ulcorner\mathbf{x}_1\urcorner,\dots,\ulcorner\mathbf{x}_n\urcorner\rangle\langle\mathbf{b}_1,\dots,\mathbf{b}_n\rangle &= \mu w(\mathrm{Len}\,w = \mathrm{S}\dot{m} \wedge (w)_0 = \dot{\sigma}(\mathbf{s}) \\ &\wedge (w)_{\dot{1}} = \hat{\mathbf{u}}_1\{\mathbf{x}_1,\dots,\mathbf{x}_k,\mathbf{x}_{k+2},\dots,\mathbf{x}_n|\mathbf{b}_1,\dots,\mathbf{b}_k,\mathbf{b}_{k+2},\dots,\mathbf{b}_n\} \\ &\wedge \dots \wedge (w)_{\dot{m}} = \hat{\mathbf{u}}_m\{\mathbf{x}_1,\dots,\mathbf{x}_k,\mathbf{x}_{k+2},\dots,\mathbf{x}_n|\mathbf{b}_1,\dots,\mathbf{b}_k,\mathbf{b}_{k+2},\dots,\mathbf{b}_n\}).\end{aligned}$$

Simplifying the right-hand side with the definition of $\langle\rangle_{m+1}$, we find the desired result.

In the above proof, we had to use that $\vdash_P \mathrm{B}\langle x_1,\dots,x_n\rangle\dot{k} = x_k$, that $\vdash_P \mathrm{Rmv}\,x\dot{0} = x$, and that $\vdash_P \mathrm{Rmv}\langle x_1,\dots,x_n\rangle\dot{k} = \langle x_1,\dots,x_{k-1},x_{k+1},\dots,x_n\rangle$, and in this only does it differ from a proof of the representability of sub by Sub obtained from the theorem on RE-formulae. These are intensional properties of B and Rmv of which only numerical instances can be derived from representability conditions.

3.4 Describing theories. Let $\mathfrak{T}$ be an arithmetical theory in P. We say that $\mathfrak{T}$ *describes T in P* if $\mathfrak{L}(\mathfrak{T})$ describes $L(T)$ and if Nlax positively represents nlax_T in P. We say that $\mathfrak{T}$ *represents T in P* if $\mathfrak{L}(\mathfrak{T})$ represents $L(T)$ and if Nlax represents nlax_T in P.

We let $L(\mathfrak{T})$ be the extension of $L(\mathfrak{L}(\mathfrak{T}))$ containing the symbols Nlax, Der, and Thm. Let α and τ be the numerical realizations of $L(\mathfrak{L}(\mathfrak{T}))$ defined in §3.2. We extend them to numerical realizations of $L(\mathfrak{T})$, which we continue to denote by α and τ, as follows: Nlax^+_α and $\mathrm{Nlax}^\pm_\tau$ are nlax_T; Nlax^-_α is $\top^1$; Der^+_α and $\mathrm{Der}^\pm_\tau$ are der_T; Der^-_α is $\top^2$; Thm^+_α and Thm^+_τ are thm_T; Thm^-_α and Thm^-_τ are $\top^1$.

Theorem. If $\mathfrak{T}$ describes T, α is faithful in P. If $\mathfrak{T}$ represents T, τ is faithful in P.

Proof. By the theorem on RE-formulae using the defining axioms of the symbols of $L(\mathfrak{T})$. □

3.5 Describing interpretations. Let $\mathfrak{L}$ and $\mathfrak{L}'$ be arithmetical languages in P with no parameters. We say that an arithmetical interpretation $\mathfrak{I}$ in P of $\mathfrak{L}$ in $\mathfrak{L}'$ *describes in P* an interpretation I of L in L' if

(i) $\mathfrak{L}$ describes L in P and $\mathfrak{L}'$ describes L' in P;

(ii) $\mathfrak{U}_\mathfrak{I}$ represents $\sigma'(\mathrm{U}_I)$ in P;

(iii) for every variable, function symbol, or predicate symbol $\mathbf{s}$ of L, $\vdash_P (\dot\sigma(\mathbf{s}))_\mathfrak{I} = \dot\sigma'(\mathbf{s}_I)$.

Let $f(a)$ denote the expression number of the interpretation of the designator with expression number a if indeed a is the expression number of a designator, and 0 otherwise. We now have two numerical realizations, one of $L(\mathfrak{L})$ associated with $\mathfrak{L}$ and L, and one of $L(\mathfrak{L}')$ associated with $\mathfrak{L}'$ and L'. We let α extend them both and we further let $(\mathfrak{U}_\mathfrak{I})_\alpha$ be $\sigma'(\mathrm{U}_I)$ and Int_α be f. Using (i)–(iii) and the defining axiom of Int, an application of the theorem on RE-formulae shows that if $\mathfrak{I}$ describes an interpretation I of L in L', then α is faithful in P, i.e.,

$$\vdash_P \mathrm{Int}\ulcorner\mathbf{a}\urcorner = \ulcorner\mathbf{a}_I\urcorner' \text{ and} \tag{4}$$

$$\vdash_P \mathrm{Int}\ulcorner\mathbf{A}\urcorner = \ulcorner\mathbf{A}^I\urcorner'. \tag{5}$$

Proposition. Suppose that $\mathfrak{T}$ describes (resp. represents) T, that $\mathfrak{T}'$ describes T', that $\mathfrak{I}$ is an arithmetical interpretation of $\mathfrak{L}(\mathfrak{T})$ in $\mathfrak{L}(\mathfrak{T}')$, and that $\mathfrak{I}$ describes an interpretation I of T in T'. Then $\mathfrak{T}_\mathfrak{I}$ describes (resp. represents) T.

Proof. Let $\mathbf{A}$ be a nonlogical axiom of T. Then $\vdash_{T'} \mathbf{A}^I$ and hence $\vdash_P \mathrm{Thm}'\ulcorner\mathbf{A}^I\urcorner'$. Since $\vdash_P \ulcorner\mathbf{A}^I\urcorner' = \mathrm{Int}\ulcorner\mathbf{A}\urcorner$, we obtain $\vdash_P \mathrm{Thm}'\,\mathrm{Int}\ulcorner\mathbf{A}\urcorner$. By definition of $\mathrm{Nlax}_\mathfrak{I}$, $\vdash_P \mathrm{Nlax}_\mathfrak{I}\ulcorner\mathbf{A}\urcorner$. If moreover $\mathfrak{T}$ represents T and a is not the expression number of a nonlogial axiom of T, then $\vdash_P \neg\mathrm{Nlax}\,\dot a$ and so by definition of $\mathrm{Nlax}_\mathfrak{I}$, $\vdash_P \neg\mathrm{Nlax}_\mathfrak{I}\,\dot a$. □

3.6 Generalization through interpretations. The definitions given in this section can be naturally generalized with no more than a notational cost. Consider an arithmetical language $\mathfrak{L}$ in P and a first-order language L. Let I be an interpretation of P in an arbitrary first-order theory P' such that $=_I$ is $=$ and such that different symbols of P have different interpretations (the latter is not an actual restriction, since we can always "duplicate" function or predicate symbols of P' using extensions by definitions). We also assume that $\dot 0_I$ is $\dot 0$ and that S_I is S (again, not a restriction). We shall say that $\mathfrak{L}$ *describes L with respect to I* if the interpretations by I of the representability conditions (i)–(iv) of §3.2 are theorems of P'. Given our assumptions on $=_I$, $\dot 0_I$, and S_I, it is equivalent to require that Vr_I represents vr in P' and similarly for (ii)–(iv). Recall that, by the interpretation extension theorem, I can be extended to $L(\mathfrak{L})$ and still be an interpretation in an extension by definitions of P' (since there are constants in P', for example $\dot 0_I$). We can then define a numerical realization α_I on the symbols $\mathbf{s}_I$ for $\mathbf{s}$ a nonlogical symbols of $L(\mathfrak{L})$ by letting $(\mathbf{s}_I)^\pm_{\alpha_I}$ be $\mathbf{s}^\pm_\alpha$ (here we use that different symbols of P have different interpretations; otherwise α_I might not be well-defined). Then α_I is faithful in P'. Indeed, α is faithful in $P[\Gamma]$ where Γ is the collection of the representability conditions (i)–(iv), and I is an interpretation of $P[\Gamma]$ in P', whence the result by the interpretation theorem.

We define similary the expressions "$\mathfrak{L}$ represents L with respect to I" and "$\mathfrak{T}$ describes or represents T with respect to I", and we have analogous results of faithfulness. We shall only use these generalized notions to give more applicability to the result of §4.3.

§4 The theorems on consistency proofs

4.1 The Main Lemma. Let P be a good extension of PA. Suppose that $L(\mathrm{N})$ is arithmetized from a numerotation and that $\mathfrak{L}$ is an arithmetical language in P that describes $L(\mathrm{N})$. We build an arithmetical theory

$\mathfrak{T}_{\mathrm{N}}$ in (an extension by definitions of) P by letting $\mathfrak{L}(\mathfrak{T}_{\mathrm{N}})$ be $\mathfrak{L}$ and by defining $\mathrm{Nlax}_{\mathrm{N}}$ by

$$\mathrm{Nlax}_{\mathrm{N}}\, x \leftrightarrow x = \dot{n}_1 \vee \cdots \vee x = \dot{n}_9$$

where $n_1, \ldots, n_9$ are the expression numbers of the axioms N1–N9. Since $\mathfrak{L}$ describes $L(\mathrm{N})$, $\vdash_P \mathrm{Nlax}_{\mathrm{N}}\, x \to \mathrm{Fm}\, x$, and so $\mathfrak{T}_{\mathrm{N}}$ is indeed an arithmetical theory in P. It is evident that $\mathrm{Nlax}_{\mathrm{N}}$ represents $\mathrm{nlax}_{\mathrm{N}}$ in P, so in particular $\mathfrak{T}_{\mathrm{N}}$ describes N.

The next theorem is a formalized statement of the theorem on RE-formulae for the numerical realization v.

Main Lemma. Let $\mathbf{A}$ be an RE-formula of $L(\mathrm{N})$ and let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be distinct variables including the variables free in $\mathbf{A}$. Then

$$\vdash_P \mathbf{A} \to \mathrm{Thm}_{\mathrm{N}}\, \mathrm{Sub}\ulcorner\mathbf{A}\urcorner\langle\ulcorner\mathbf{x}_1\urcorner, \ldots, \ulcorner\mathbf{x}_n\urcorner\rangle\langle\mathrm{Num}\,\mathbf{x}_1, \ldots, \mathrm{Num}\,\mathbf{x}_n\rangle.$$

Proof. Before embarking on the proof itself, we recall a few facts that will be used. Let $\mathbf{y}_1, \ldots, \mathbf{y}_k$ be any distinct variables. By (1) of §3.3, we have

$$\vdash_P \mathrm{Sub}\ulcorner\mathbf{u}\urcorner\langle\ulcorner\mathbf{y}_1\urcorner, \ldots, \ulcorner\mathbf{y}_k\urcorner\rangle\langle\mathrm{Num}\,\mathbf{y}_1, \ldots, \mathrm{Num}\,\mathbf{y}_k\rangle = \hat{\mathbf{u}}\{\mathbf{y}_1, \ldots, \mathbf{y}_k \,|\, \mathrm{Num}\,\mathbf{y}_k, \ldots, \mathrm{Num}\,\mathbf{y}_k\}. \tag{1}$$

In view of this and the equality theorem, we need only prove the theorem for a particular choice of the variables $\mathbf{x}_1, \ldots, \mathbf{x}_n$. By (xxi′) of §2.2, $\vdash_P \Sigma\mathrm{ub}\langle\ulcorner\mathbf{y}_1\urcorner, \ldots, \ulcorner\mathbf{y}_n\urcorner\rangle\langle\mathrm{Num}\,\mathbf{y}_1, \ldots, \mathrm{Num}\,\mathbf{y}_n\rangle$, and by (ix′) and (xxii′) of §2.2, $\vdash_P \mathrm{Fm}\, x \to \mathrm{Vble}\, y \to \mathrm{Subtl}\, x y\, \mathrm{Num}\, z$, so for any formula $\mathbf{B}$ of $L(\mathrm{N})$, since $\vdash_P \mathrm{Fm}\ulcorner\mathbf{B}\urcorner$ and $\vdash_P \mathrm{Vble}\ulcorner\mathbf{x}\urcorner$, the formalized substitution rule yields

$$\vdash_P \mathrm{Thm}_{\mathrm{N}}\ulcorner\mathbf{B}\urcorner \to \mathrm{Thm}_{\mathrm{N}}\, \mathrm{Sub}\ulcorner\mathbf{B}\urcorner\langle\ulcorner\mathbf{y}_1\urcorner, \ldots, \ulcorner\mathbf{y}_k\urcorner\rangle\langle\mathrm{Num}\,\mathbf{y}_1, \ldots, \mathrm{Num}\,\mathbf{y}_k\rangle.$$

If $\mathbf{A}$ and $\mathbf{B}$ are formulae of $L(\mathrm{N})$ such that $\vdash_{\mathrm{N}} \mathbf{A} \leftrightarrow \mathbf{B}$, then

$$\vdash_P \mathrm{Thm}_{\mathrm{N}}\, \mathrm{Sub}\ulcorner\mathbf{A} \leftrightarrow \mathbf{B}\urcorner\langle\ulcorner\mathbf{y}_1\urcorner, \ldots, \ulcorner\mathbf{y}_k\urcorner\rangle\langle\mathrm{Num}\,\mathbf{y}_1, \ldots, \mathrm{Num}\,\mathbf{y}_k\rangle$$

by positive representability and the formalized substitution rule. Directly from the definition of Sub, we see that $\vdash_P \mathrm{Sub}\, \mathrm{Eqv}\, x x' y z = \mathrm{Eqv}\, \mathrm{Sub}\, x y z\, \mathrm{Sub}\, x' y z$, and so we obtain

$$\begin{aligned}\vdash_P \mathrm{Thm}_{\mathrm{N}}\, &\mathrm{Sub}\ulcorner\mathbf{A}\urcorner\langle\ulcorner\mathbf{y}_1\urcorner, \ldots, \ulcorner\mathbf{y}_k\urcorner\rangle\langle\mathrm{Num}\,\mathbf{y}_1, \ldots, \mathrm{Num}\,\mathbf{y}_k\rangle \\ &\leftrightarrow \mathrm{Thm}_{\mathrm{N}}\, \mathrm{Sub}\ulcorner\mathbf{B}\urcorner\langle\ulcorner\mathbf{y}_1\urcorner, \ldots, \ulcorner\mathbf{y}_k\urcorner\rangle\langle\mathrm{Num}\,\mathbf{y}_1, \ldots, \mathrm{Num}\,\mathbf{y}_k\rangle\end{aligned}$$

by (viii) of §2.6. This applies in particular if $\mathbf{A}$ and $\mathbf{B}$ are numerically equivalent. Thus we may suppose that $\mathbf{A}$ is a strict RE-formula. This is our hypothesis from now on, and we prove the theorem by induction on the length of $\mathbf{A}$.

Suppose that $\mathbf{A}$ is $\mathbf{x} = \dot{0}$. Since $\vdash_{\mathrm{N}} \dot{0} = \dot{0}$, we have $\vdash_P \mathrm{Thm}_{\mathrm{N}}\langle\dot{=}, \ddot{0}, \ddot{0}\rangle$ because $\mathfrak{T}_{\mathrm{N}}$ describes N, whence $\vdash_P \mathrm{Thm}_{\mathrm{N}}\langle\dot{=}, \mathrm{Num}\,\dot{0}, \ddot{0}\rangle$ by definition of Num. From this by the equality theorem, $\vdash_P \mathbf{x} = \dot{0} \to \mathrm{Thm}_{\mathrm{N}}\langle\dot{=}, \mathrm{Num}\,\mathbf{x}, \ddot{0}\rangle$, which is the desired result by (1).

Suppose that $\mathbf{A}$ is $\mathbf{y} = \mathrm{S}\mathbf{x}$. Since $\vdash_{\mathrm{N}} x = x$ we obtain $\vdash_P \mathrm{Thm}_{\mathrm{N}}\ulcorner x = x\urcorner$ whence $\vdash_P \mathrm{Thm}_{\mathrm{N}}\langle\dot{=}, \mathrm{Num}\,\mathrm{S}\mathbf{x}, \mathrm{Num}\,\mathrm{S}\mathbf{x}\rangle$ by the formalized substitution rule. By definition of Num, $\vdash_P \mathrm{Num}\,\mathrm{S}\mathbf{x} = \langle\dot{\sigma}(\mathrm{S}), \mathrm{Num}\,\mathbf{x}\rangle$, and so $\vdash_P \mathrm{Thm}_{\mathrm{N}}\langle\dot{=}, \mathrm{Num}\,\mathrm{S}\mathbf{x}, \langle\dot{\sigma}(\mathrm{S}), \mathrm{Num}\,\mathbf{x}\rangle\rangle$ by the equality theorem. From this by the equality theorem, $\vdash_P \mathbf{y} = \mathrm{S}\mathbf{x} \to \mathrm{Thm}_{\mathrm{N}}\langle\dot{=}, \mathrm{Num}\,\mathbf{y}, \langle\dot{\sigma}(\mathrm{S}), \mathrm{Num}\,\mathbf{x}\rangle\rangle$.

Suppose that $\mathbf{A}$ is $\mathbf{z} = \mathbf{x} + \mathbf{y}$. We let $\mathbf{a}$ be $\langle\dot{=}, \mathrm{Num}(x + y), \langle\dot{\sigma}(+), \mathrm{Num}\,x, \mathrm{Num}\,y\rangle\rangle$. Since $\vdash_{\mathrm{N}} x = x + \dot{0}$, we have $\vdash_P \mathrm{Thm}_{\mathrm{N}}\ulcorner x = x + \dot{0}\urcorner$, whence $\vdash_P \mathrm{Thm}_{\mathrm{N}}\langle\dot{=}, \mathrm{Num}\,x, \langle\dot{\sigma}(+), \mathrm{Num}\,x, \ddot{0}\rangle\rangle$ by the formalized substitution rule. Since $\vdash_P \mathrm{Num}\,\dot{0} = \ddot{0}$, we obtain

$$\vdash_P \mathrm{Thm}_{\mathrm{N}}\, \mathbf{a}[y|\dot{0}]. \tag{2}$$

Since $\vdash_{\mathrm{N}} z = x + y \to \mathrm{S}z = x + \mathrm{S}y$, we have $\vdash_P \mathrm{Thm}_{\mathrm{N}}\ulcorner z = x + y \to \mathrm{S}z = x + \mathrm{S}y\urcorner$, and hence

$$\vdash_P \mathrm{Thm}_{\mathrm{N}}\, \mathrm{Imp}\langle\dot{=}, \mathrm{Num}\,z, \langle\dot{\sigma}(+), \mathrm{Num}\,x, \mathrm{Num}\,y\rangle\rangle\langle\dot{=}, \langle\dot{\mathrm{S}}, \mathrm{Num}\,z\rangle, \langle\dot{\sigma}(+), \mathrm{Num}\,x, \langle\dot{\mathrm{S}}, \mathrm{Num}\,y\rangle\rangle\rangle$$

by the formalized substitution rule, and

$$\vdash_P \mathrm{Thm}_{\mathrm{N}}\, \mathrm{Imp}\langle\dot{=}, \mathrm{Num}\,z, \langle\dot{\sigma}(+), \mathrm{Num}\,x, \mathrm{Num}\,y\rangle\rangle\langle\dot{=}, \mathrm{Num}\,\mathrm{S}z, \langle\dot{\sigma}(+), \mathrm{Num}\,x, \mathrm{Num}\,\mathrm{S}y\rangle\rangle$$

by the definition of Num. From this by the substitution rule, N4, and the formalized detachment rule, we obtain

$$\vdash_P \mathrm{Thm_N}\,\mathbf{a} \to \mathrm{Thm_N}\,\mathbf{a}[y|\mathrm{S}y]. \tag{3}$$

From (2) and (3) by the induction axioms, $\vdash_P \mathrm{Thm_N}\,\mathbf{a}$. By the equality theorem and the substitution rule, $\vdash_P \mathbf{z} = \mathbf{x} + \mathbf{y} \to \mathrm{Thm_N}\langle \dot{=}, \mathrm{Num}\,\mathbf{z}, \langle \dot{\sigma}(+), \mathrm{Num}\,\mathbf{x}, \mathrm{Num}\,\mathbf{y}\rangle\rangle$.

Suppose that $\mathbf{A}$ is $\mathbf{z} = \mathbf{x} \cdot \mathbf{y}$. We let $\mathbf{a}$ be $\langle \dot{=}, \mathrm{Num}(x \cdot y), \langle \dot{\sigma}(\cdot), \mathrm{Num}\,x, \mathrm{Num}\,y\rangle\rangle$. Since $\vdash_N \dot{0} = x \cdot \dot{0}$, we have $\vdash_P \mathrm{Thm_N}\ulcorner \dot{0} = x \cdot \dot{0}\urcorner$, whence $\vdash_P \mathrm{Thm_N}\langle \dot{=}, \ddot{0}, \langle \dot{\sigma}(\cdot), \mathrm{Num}\,x, \ddot{0}\rangle\rangle$ by the formalized substitution rule. Since $\vdash_P \mathrm{Num}\,\dot{0} = \ddot{0}$, we obtain

$$\vdash_P \mathrm{Thm_N}\,\mathbf{a}[y|\dot{0}]. \tag{4}$$

Since $\vdash_N w = z + x \to z = x \cdot y \to w = x \cdot \mathrm{S}y$, we have $\vdash_P \mathrm{Thm_N}\ulcorner w = z + x \to z = x \cdot y \to w = x \cdot \mathrm{S}y\urcorner$, and hence

$$\begin{aligned}\vdash_P &\mathrm{Thm_N}\,\mathrm{Imp}\langle \dot{=}, \mathrm{Num}\,w, \langle \dot{\sigma}(+), \mathrm{Num}\,z, \mathrm{Num}\,x\rangle\rangle \\ &\qquad \mathrm{Imp}\langle \dot{=}, \mathrm{Num}\,z, \langle \dot{\sigma}(\cdot), \mathrm{Num}\,x, \mathrm{Num}\,y\rangle\rangle\langle \dot{=}, \mathrm{Num}\,w, \langle \dot{\sigma}(\cdot), \mathrm{Num}\,x, \mathrm{Num}\,\mathrm{S}y\rangle\end{aligned}$$

by the formalized substitution rule and the definition of Num. Substituting $z + x$ for w and using the formalized detachement rule, we obtain

$$\begin{aligned}\vdash_P &\mathrm{Thm_N}\langle \dot{=}, \mathrm{Num}(z + x), \langle \dot{\sigma}(+), \mathrm{Num}\,z, \mathrm{Num}\,x\rangle\rangle \\ &\to \mathrm{Thm_N}\langle \dot{=}, \mathrm{Num}\,z, \langle \dot{\sigma}(\cdot), \mathrm{Num}\,x, \mathrm{Num}\,y\rangle\rangle \to \mathrm{Thm_N}\langle \dot{=}, \mathrm{Num}(z + x), \langle \dot{\sigma}(\cdot), \mathrm{Num}\,x, \mathrm{Num}\,\mathrm{S}y\rangle\end{aligned}$$

whence

$$\vdash_P \mathrm{Thm_N}\langle \dot{=}, \mathrm{Num}\,z, \langle \dot{\sigma}(\cdot), \mathrm{Num}\,x, \mathrm{Num}\,y\rangle\rangle \to \mathrm{Thm_N}\langle \dot{=}, \mathrm{Num}(z + x), \langle \dot{\sigma}(\cdot), \mathrm{Num}\,x, \mathrm{Num}\,\mathrm{S}y\rangle$$

by the previous case and the detachment rule. Substituting $x \cdot y$ for z and using N6, we obtain

$$\vdash_P \mathrm{Thm_N}\,\mathbf{a} \to \mathrm{Thm_N}\,\mathbf{a}[y|\mathrm{S}y]. \tag{5}$$

From (4) and (5) by the induction axioms, $\vdash_P \mathrm{Thm_N}\,\mathbf{a}$. By the equality theorem and the substitution rule, $\vdash_P \mathbf{z} = \mathbf{x} \cdot \mathbf{y} \to \mathrm{Thm_N}\langle \dot{=}, \mathrm{Num}\,\mathbf{z}, \langle \dot{\sigma}(\cdot), \mathrm{Num}\,\mathbf{x}, \mathrm{Num}\,\mathbf{y}\rangle\rangle$.

Suppose that $\mathbf{A}$ is $\mathbf{x} = \mathbf{y}$. From $\vdash_N x = x$ we obtain $\vdash_P \mathrm{Thm_N}\langle \dot{=}, \mathrm{Num}\,\mathbf{x}, \mathrm{Num}\,\mathbf{x}\rangle$ by the formalized substitution rule, whence $\vdash_P \mathbf{A} \to \mathrm{Thm_N}\langle \dot{=}, \mathrm{Num}\,\mathbf{x}, \mathrm{Num}\,\mathbf{y}\rangle$ by the equality theorem.

Suppose that $\mathbf{A}$ is $\mathbf{x} < \mathbf{y}$, and let $\mathbf{B}$ be $x < y \to \mathrm{Thm_N}\langle \dot{\sigma}(<), \mathrm{Num}\,x, \mathrm{Num}\,y\rangle$; by the substitution rule, it will suffice to prove $\vdash_P \mathbf{B}$. By N1 and the tautology theorem,

$$\vdash_P \mathbf{B}[y|\dot{0}] \tag{6}$$

Since $\vdash_N x < y \to x < \mathrm{S}y$ and $\vdash_N x = y \to x < \mathrm{S}y$, we have, using the formalized substitution rule and the definition of Num,

$$\begin{gathered}\vdash_P \mathrm{Thm_N}\,\mathrm{Imp}\langle \dot{\sigma}(<), \mathrm{Num}\,x, \mathrm{Num}\,y\rangle\langle \dot{\sigma}(<), \mathrm{Num}\,x, \mathrm{Num}\,\mathrm{S}y\rangle \text{ and} \\ \vdash_P \mathrm{Thm_N}\,\mathrm{Imp}\langle \dot{=}, \mathrm{Num}\,x, \mathrm{Num}\,y\rangle\langle \dot{\sigma}(<), \mathrm{Num}\,x, \mathrm{Num}\,\mathrm{S}y\rangle.\end{gathered}$$

By the formalized detachment rule, these become

$$\begin{gathered}\vdash_P \mathrm{Thm_N}\langle \dot{\sigma}(<), \mathrm{Num}\,x, \mathrm{Num}\,y\rangle \to \mathrm{Thm_N}\langle \dot{\sigma}(<), \mathrm{Num}\,x, \mathrm{Num}\,\mathrm{S}y\rangle \text{ and} \\ \vdash_P \mathrm{Thm_N}\langle \dot{=}, \mathrm{Num}\,x, \mathrm{Num}\,y\rangle \to \mathrm{Thm_N}\langle \dot{\sigma}(<), \mathrm{Num}\,x, \mathrm{Num}\,\mathrm{S}y\rangle.\end{gathered}$$

By the previous case, $\vdash_P x = y \to \mathrm{Thm_N}\langle \dot{=}, \mathrm{Num}\,x, \mathrm{Num}\,y\rangle$, and since $\vdash_P x < \mathrm{S}y \to x < y \lor x = y$, we obtain

$$\vdash_P \mathbf{B} \to \mathbf{B}[y|\mathrm{S}y] \tag{7}$$

by the tautology theorem. From (6) and (7) by the induction axioms, $\vdash_P \mathbf{B}$.

Suppose that $\mathbf{A}$ is $\mathbf{x} \neq \mathbf{y}$. Let $\mathbf{B}$ be the formula $x \neq y \to \mathrm{Thm_N}\,\mathrm{Neg}\langle \dot{=}, \mathrm{Num}\,x, \mathrm{Num}\,y\rangle$, to be derived. Since $\vdash_N \mathrm{S}x \neq \dot{0}$ we have $\vdash_P \mathrm{Thm_N}\ulcorner \mathrm{S}x \neq \dot{0}\urcorner$ whence $\vdash_P \mathrm{Thm_N}\,\mathrm{Neg}\langle \dot{=}, \mathrm{Num}\,\mathrm{S}y, \mathrm{Num}\,\dot{0}\rangle$ by the formalized substitution rule and the definition of Num. By the equality theorem and the $\exists$-introduction rule,

$\vdash_P \exists y(x = \mathrm{S}y) \to \mathrm{Thm}_{\mathrm{N}}\,\mathrm{Neg}\langle\dot{=}, \mathrm{Num}\,x, \mathrm{Num}\,\dot{0}\rangle$. Now $\vdash_P x \neq \dot{0} \to \exists y(x = \mathrm{S}y)$, whence $\vdash_P x \neq \dot{0} \to \mathrm{Thm}_{\mathrm{N}}\,\mathrm{Neg}\langle\dot{=}, \mathrm{Num}\,x, \mathrm{Num}\,\dot{0}\rangle$ by the tautology theorem, i.e., $\vdash_P \mathbf{B}[y|\dot{0}]$. By the induction axioms and the closure theorem, it remains to prove $\vdash_P \forall x\mathbf{B} \to \forall x\mathbf{B}[y|\mathrm{S}y]$. By the $\forall$-introduction rule and the tautology theorem, we need only prove

$$\vdash_P x = \dot{0} \to \forall x\mathbf{B} \to \mathbf{B}[y|\mathrm{S}y] \text{ and} \tag{8}$$

$$\vdash_P x \neq \dot{0} \to \forall x\mathbf{B} \to \mathbf{B}[y|\mathrm{S}y]. \tag{9}$$

Since $\vdash_{\mathrm{N}} x \neq \dot{0} \to \dot{0} \neq x$, we obtain

$$\vdash_P \mathrm{Thm}_{\mathrm{N}}\,\mathrm{Neg}\langle\dot{=}, \mathrm{Num}\,x, \mathrm{Num}\,\dot{0}\rangle \to \mathrm{Thm}_{\mathrm{N}}\,\mathrm{Neg}\langle\dot{=}, \mathrm{Num}\,\dot{0}, \mathrm{Num}\,x\rangle$$

by the formalized substitution rule, the definition of Num, and the formalized detachment rule. From this and $\vdash_P \mathbf{B}[y|\dot{0}]$, $\vdash_P x \neq \dot{0} \to \mathrm{Thm}_{\mathrm{N}}\,\mathrm{Neg}\langle\dot{=}, \mathrm{Num}\,\dot{0}, \mathrm{Num}\,x\rangle$, whence $\vdash_P \mathrm{S}y \neq \dot{0} \to \mathrm{Thm}_{\mathrm{N}}\,\mathrm{Neg}\langle\dot{=}, \mathrm{Num}\,\dot{0}, \mathrm{Num}\,\mathrm{S}y\rangle$ by the substitution rule. By an instance of N1 and the detachment rule, $\vdash_P \mathrm{Thm}_{\mathrm{N}}\,\mathrm{Neg}\langle\dot{=}, \mathrm{Num}\,\dot{0}, \mathrm{Num}\,\mathrm{S}y\rangle$ and hence by the equality theorem, $\vdash_P x = \dot{0} \to \mathrm{Thm}_{\mathrm{N}}\,\mathrm{Neg}\langle\dot{=}, \mathrm{Num}\,x, \mathrm{Num}\,\mathrm{S}y\rangle$, of which (8) is a tautological consequence. By N2, $\vdash_{\mathrm{N}} x \neq y \to \mathrm{S}x \neq \mathrm{S}y$, and so $\vdash_P \mathrm{Thm}_{\mathrm{N}}\,\mathrm{Neg}\langle\dot{=}, \mathrm{Num}\,x', \mathrm{Num}\,y\rangle \to \mathrm{Thm}_{\mathrm{N}}\,\mathrm{Neg}\langle\dot{=}, \mathrm{Num}\,\mathrm{S}x', \mathrm{Num}\,\mathrm{S}y\rangle$ by the formalized substitution rule, the definition of Num, and the formalized detachment rule. From this by the equality theorem,

$$\vdash_P x = \mathrm{S}x' \to \mathrm{Thm}_{\mathrm{N}}\,\mathrm{Neg}\langle\dot{=}, \mathrm{Num}\,x', \mathrm{Num}\,y\rangle \to \mathrm{Thm}_{\mathrm{N}}\,\mathrm{Neg}\langle\dot{=}, \mathrm{Num}\,x, \mathrm{Num}\,\mathrm{S}y\rangle. \tag{10}$$

By the substitution theorem,

$$\vdash_P \forall x\mathbf{B} \to x' \neq y \to \mathrm{Thm}_{\mathrm{N}}\,\mathrm{Neg}\langle\dot{=}, \mathrm{Num}\,x', \mathrm{Num}\,y\rangle. \tag{11}$$

From (10), (11), and $\vdash_P x = \mathrm{S}x' \to x \neq \mathrm{S}y \to x' \neq y$ we obtain

$$\vdash_P x = \mathrm{S}x' \to \forall x\mathbf{B} \to x \neq \mathrm{S}y \to \mathrm{Thm}_{\mathrm{N}}\,\mathrm{Neg}\langle\dot{=}, \mathrm{Num}\,x, \mathrm{Num}\,\mathrm{S}y\rangle \tag{12}$$

by the tautology theorem, i.e., $\vdash_P x = \mathrm{S}x' \to \forall x\mathbf{B} \to \mathbf{B}[y|\mathrm{S}y]$. By the $\exists$-introduction rule and $\vdash_P x \neq \dot{0} \to \exists x'(x = \mathrm{S}x')$, we obtain (9).

Suppose that $\mathbf{A}$ is $\neg(\mathbf{x} < \mathbf{y})$, and let $\mathbf{B}$ be $\neg(x < y) \to \mathrm{Thm}_{\mathrm{N}}\,\mathrm{Neg}\langle\dot{\sigma}(<), \mathrm{Num}\,x, \mathrm{Num}\,y\rangle$, the formula to be derived. By N7, $\vdash_P \mathrm{Thm}_{\mathrm{N}}\ulcorner\neg(x < \dot{0})\urcorner$, whence $\vdash_P \mathrm{Thm}_{\mathrm{N}}\,\mathrm{Neg}\langle\dot{\sigma}(<), \mathrm{Num}\,x, \mathrm{Num}\,\dot{0}\rangle$ by the formalized substitution rule and the definition of Num. By the tautology theorem, $\vdash_P \mathbf{B}[y|\dot{0}]$. Now by N8, $\vdash_{\mathrm{N}} \neg(x < y) \to x \neq y \to \neg(x < \mathrm{S}y)$, so $\vdash_P \mathrm{Thm}_{\mathrm{N}}\ulcorner\neg(x < y) \to x \neq y \to \neg(x < \mathrm{S}y)\urcorner$. Hence

$$\begin{aligned}\vdash_P \mathrm{Thm}_{\mathrm{N}}\,\mathrm{Neg}\langle\dot{\sigma}(<), \mathrm{Num}\,x, \mathrm{Num}\,y\rangle &\to \mathrm{Thm}_{\mathrm{N}}\,\mathrm{Neg}\langle\dot{=}, \mathrm{Num}\,x, \mathrm{Num}\,y\rangle \\ &\to \mathrm{Thm}_{\mathrm{N}}\,\mathrm{Neg}\langle\dot{\sigma}(<), \mathrm{Num}\,x, \mathrm{Num}\,\mathrm{S}y\rangle\end{aligned} \tag{13}$$

by the formalized substitution rule, the definition of Num, and the formalized detachment rule. By the previous case,

$$\vdash_P x \neq y \to \mathrm{Thm}_{\mathrm{N}}\,\mathrm{Neg}\langle\dot{=}, \mathrm{Num}\,x, \mathrm{Num}\,y\rangle \tag{14}$$

By (13), (14), and an instance of N8, we obtain $\vdash_P \mathbf{B} \to \mathbf{B}[y|\mathrm{S}y]$ by the tautology theorem. So $\vdash_P \mathbf{B}$ by the induction axioms.

Suppose that $\mathbf{A}$ is $\mathbf{B} \vee \mathbf{C}$ where $\mathbf{B}$ and $\mathbf{C}$ are strict RE-formulae. By the induction hypothesis and the tautology theorem,

$$\begin{aligned}\vdash_P \mathbf{A} \to \mathrm{Thm}_{\mathrm{N}}\,&\mathrm{Sub}\ulcorner\mathbf{B}\urcorner\langle\ulcorner\mathbf{x}_1\urcorner, \ldots, \ulcorner\mathbf{x}_n\urcorner\rangle\langle\mathrm{Num}\,\mathbf{x}_1, \ldots, \mathrm{Num}\,\mathbf{x}_n\rangle \\ &\vee \mathrm{Thm}_{\mathrm{N}}\,\mathrm{Sub}\ulcorner\mathbf{C}\urcorner\langle\ulcorner\mathbf{x}_1\urcorner, \ldots, \ulcorner\mathbf{x}_n\urcorner\rangle\langle\mathrm{Num}\,\mathbf{x}_1, \ldots, \mathrm{Num}\,\mathbf{x}_n\rangle,\end{aligned}$$

whence the desired result by (v) of §2.2. If $\mathbf{A}$ is $\mathbf{B} \wedge \mathbf{C}$, the proof is similar using (vi) of §2.2 instead.

Suppose that $\mathbf{A}$ is $\forall\mathbf{x}(\mathbf{x} < \mathbf{y} \to \mathbf{B})$, where $\mathbf{B}$ is a strict RE-formula. By (1), we may suppose that $\mathbf{y}$ is $\mathbf{x}_1$ and that $\mathbf{x}$ is not among $\mathbf{x}_1, \ldots, \mathbf{x}_n$. We have $\vdash_{\mathrm{N}} \mathbf{A}[\mathbf{x}_1|\dot{0}]$ by N7, and so $\vdash_P \mathrm{Thm}_{\mathrm{N}}\,\mathrm{Sub}\ulcorner\mathbf{A}[\mathbf{x}_1|\dot{0}]\urcorner\langle\ulcorner\mathbf{x}_2\urcorner, \ldots, \ulcorner\mathbf{x}_n\urcorner\rangle\langle\mathrm{Num}\,\mathbf{x}_2, \ldots, \mathrm{Num}\,\mathbf{x}_n\rangle$ by the formalized substitution rule. By (1),

$$\begin{aligned}\vdash_P \mathrm{Sub}\ulcorner\mathbf{A}\urcorner&\langle\ulcorner\mathbf{x}_1\urcorner, \ldots, \ulcorner\mathbf{x}_n\urcorner\rangle\langle\mathrm{Num}\,\dot{0}, \mathrm{Num}\,\mathbf{x}_2, \ldots, \mathrm{Num}\,\mathbf{x}_n\rangle \\ &= \mathrm{Sub}\,\mathrm{Sub}\ulcorner\mathbf{A}\urcorner\langle\ulcorner\mathbf{x}_1\urcorner\rangle\langle\mathrm{Num}\,\dot{0}\rangle\langle\ulcorner\mathbf{x}_2\urcorner, \ldots, \ulcorner\mathbf{x}_n\urcorner\rangle\langle\mathrm{Num}\,\mathbf{x}_2, \ldots, \mathrm{Num}\,\mathbf{x}_n\rangle,\end{aligned}$$

and by the definition of Num, $\vdash_P \ulcorner \mathbf{A}[\mathbf{x}_1|\dot{0}]\urcorner = \mathrm{Sub}\ulcorner \mathbf{A}\urcorner\langle\ulcorner \mathbf{x}_1\urcorner\rangle\langle \mathrm{Num}\,\dot{0}\rangle$, whence

$$\vdash_P \mathrm{Thm}_{\mathrm{N}}\,\mathrm{Sub}\ulcorner \mathbf{A}\urcorner\langle\ulcorner \mathbf{x}_1\urcorner,\dots,\ulcorner \mathbf{x}_n\urcorner\rangle\langle \mathrm{Num}\,\dot{0},\mathrm{Num}\,\mathbf{x}_2,\dots,\mathrm{Num}\,\mathbf{x}_n\rangle$$

by the equality theorem. Hence

$$\vdash_P \mathbf{A}[\mathbf{x}_1|\dot{0}] \to \mathrm{Thm}_{\mathrm{N}}\,\mathrm{Sub}\ulcorner \mathbf{A}\urcorner\langle\ulcorner \mathbf{x}_1\urcorner,\dots,\ulcorner \mathbf{x}_n\urcorner\rangle\langle \mathrm{Num}\,\dot{0},\mathrm{Num}\,\mathbf{x}_2,\dots,\mathrm{Num}\,\mathbf{x}_n\rangle \qquad (15)$$

by the tautology theorem. Let $\mathbf{B}'$ be a variant of $\mathbf{B}$ in which $\mathbf{x}_1$ is substitutible for $\mathbf{x}$. From N8 we obtain easily $\vdash_{\mathrm{N}} \mathbf{A} \to \mathbf{B}'[\mathbf{x}|\mathbf{x}_1] \to \mathbf{A}[\mathbf{x}_1|\mathrm{S}\mathbf{x}_1]$, whence $\vdash_P \mathrm{Thm}_{\mathrm{N}}\ulcorner \mathbf{A} \to \mathbf{B}'[\mathbf{x}|\mathbf{x}_1] \to \mathbf{A}[\mathbf{x}_1|\mathrm{S}\mathbf{x}_1]\urcorner$. By the formalized substitution rule, (1), the definition of Num, and the formalized detachment rule, we obtain

$$\begin{aligned}\vdash_P &\mathrm{Thm}_{\mathrm{N}}\,\mathrm{Sub}\ulcorner \mathbf{A}\urcorner\langle\ulcorner \mathbf{x}_1\urcorner,\dots,\ulcorner \mathbf{x}_n\urcorner\rangle\langle \mathrm{Num}\,\mathbf{x}_1,\dots,\mathrm{Num}\,\mathbf{x}_n\rangle\\ &\to \mathrm{Thm}_{\mathrm{N}}\,\mathrm{Sub}\ulcorner \mathbf{B}'\urcorner\langle\ulcorner \mathbf{x}\urcorner,\ulcorner \mathbf{x}_1\urcorner,\dots,\ulcorner \mathbf{x}_n\urcorner\rangle\langle \mathrm{Num}\,\mathbf{x}_1,\mathrm{Num}\,\mathbf{x}_1,\dots,\mathrm{Num}\,\mathbf{x}_n\rangle\\ &\to \mathrm{Thm}_{\mathrm{N}}\,\mathrm{Sub}\ulcorner \mathbf{A}\urcorner\langle\ulcorner \mathbf{x}_1\urcorner,\dots,\ulcorner \mathbf{x}_n\urcorner\rangle\langle \mathrm{Num}\,\mathrm{S}\mathbf{x}_1,\mathrm{Num}\,\mathbf{x}_2,\dots,\mathrm{Num}\,\mathbf{x}_n\rangle. \qquad (16)\end{aligned}$$

By induction hypothesis, the variant theorem, the substitution rule, and the remarks at the beginning of the proof,

$$\vdash_P \mathbf{B}'[\mathbf{x}|\mathbf{x}_1] \to \mathrm{Thm}_{\mathrm{N}}\,\mathrm{Sub}\ulcorner \mathbf{B}'\urcorner\langle\ulcorner \mathbf{x}\urcorner,\ulcorner \mathbf{x}_1\urcorner,\dots,\ulcorner \mathbf{x}_n\urcorner\rangle\langle \mathrm{Num}\,\mathbf{x}_1,\mathrm{Num}\,\mathbf{x}_1,\dots,\mathrm{Num}\,\mathbf{x}_n\rangle. \qquad (17)$$

As a tautological consequence of (16), (17), and $\vdash_P \mathbf{A}[\mathbf{x}_1|\mathrm{S}\mathbf{x}_1] \to \mathbf{A} \wedge \mathbf{B}'[\mathbf{x}|\mathbf{x}_1]$, we obtain

$$\begin{aligned}\vdash_P &(\mathbf{A} \to \mathrm{Thm}_{\mathrm{N}}\,\mathrm{Sub}\ulcorner \mathbf{A}\urcorner\langle\ulcorner \mathbf{x}_1\urcorner,\dots,\ulcorner \mathbf{x}_n\urcorner\rangle\langle \mathrm{Num}\,\mathbf{x}_1,\dots,\mathrm{Num}\,\mathbf{x}_n\rangle)\\ &\to \mathbf{A}[\mathbf{x}_1|\mathrm{S}\mathbf{x}_1] \to \mathrm{Thm}_{\mathrm{N}}\,\mathrm{Sub}\ulcorner \mathbf{A}\urcorner\langle\ulcorner \mathbf{x}_1\urcorner,\dots,\ulcorner \mathbf{x}_n\urcorner\rangle\langle \mathrm{Num}\,\mathrm{S}\mathbf{x}_1,\mathrm{Num}\,\mathbf{x}_2,\dots,\mathrm{Num}\,\mathbf{x}_n\rangle. \qquad (18)\end{aligned}$$

From (15) and (18) by the induction axioms, we obtain the desired result.

Finally, suppose that $\mathbf{A}$ is $\exists\mathbf{x}\mathbf{B}$ where $\mathbf{B}$ is a strict RE-formula. Here we assume that $\mathbf{x}$ is not among $\mathbf{x}_1, \dots, \mathbf{x}_n$. The formula $\mathbf{B} \to \mathbf{A}$ is a substitution axiom, and hence $\vdash_P \mathrm{Thm}_{\mathrm{N}}\ulcorner \mathbf{B} \to \mathbf{A}\urcorner$. By the formalized substitution rule and the formalized detachment rule,

$$\begin{aligned}\vdash_P &\mathrm{Thm}_{\mathrm{N}}\,\mathrm{Sub}\ulcorner \mathbf{B}\urcorner\langle\ulcorner \mathbf{x}\urcorner,\ulcorner \mathbf{x}_1\urcorner,\dots,\ulcorner \mathbf{x}_n\urcorner\rangle\langle \mathrm{Num}\,\mathbf{x},\mathrm{Num}\,\mathbf{x}_1,\dots,\mathrm{Num}\,\mathbf{x}_n\rangle\\ &\to \mathrm{Thm}_{\mathrm{N}}\,\mathrm{Sub}\ulcorner \mathbf{A}\urcorner\langle\ulcorner \mathbf{x}_1\urcorner,\dots,\ulcorner \mathbf{x}_n\urcorner\rangle\langle \mathrm{Num}\,\mathbf{x}_1,\dots,\mathrm{Num}\,\mathbf{x}_n\rangle.\end{aligned}$$

and so by induction hypothesis,

$$\vdash_P \mathbf{B} \to \mathrm{Thm}_{\mathrm{N}}\,\mathrm{Sub}\ulcorner \mathbf{A}\urcorner\langle\ulcorner \mathbf{x}_1\urcorner,\dots,\ulcorner \mathbf{x}_n\urcorner\rangle\langle \mathrm{Num}\,\mathbf{x}_1,\dots,\mathrm{Num}\,\mathbf{x}_n\rangle.$$

Then by the $\exists$-introduction rule, we obtain the desired result. □

Let L be an extension of $L(\mathrm{N})$ and let $\mathfrak{L}$ describe L in P. We define an arithmetical language $\mathfrak{L}_{\mathrm{N}}$ in P as follows. We set $\mathrm{Vr}_{\mathrm{N}}\,x = \mathrm{Vr}\,x$, $\mathrm{Func}_{\mathrm{N}}\,xy \leftrightarrow (y = \dot{0} \wedge x = \dot{\sigma}(\dot{0})) \vee (y = \dot{1} \wedge x = \dot{\sigma}(\mathrm{S})) \vee (y = \dot{2} \wedge x = \dot{\sigma}(+) \vee x = \dot{\sigma}(\cdot))$, $\mathrm{Pred}_{\mathrm{N}}\,xy \leftrightarrow (y = 2 \wedge x = \dot{\sigma}(=) \vee x = \dot{\sigma}(<))$, $\dot{\vee}_{\mathrm{N}} = \dot{\sigma}(\vee)$, $\dot{\neg}_{\mathrm{N}} = \dot{\sigma}(\neg)$, $\dot{\exists}_{\mathrm{N}} = \dot{\sigma}(\exists)$, $\dot{=}_{\mathrm{N}} = \dot{\sigma}(=)$, $\ddot{0}_{\mathrm{N}} = \dot{\sigma}(\dot{0})$, and $\dot{\mathrm{S}}_{\mathrm{N}} = \dot{\sigma}(\mathrm{S})$. We let σ_{N} be the restriction of the numerotation σ to $L(\mathrm{N})$, and we let $L(\mathrm{N})$ be arithmetized from σ_{N}. It is then obvious that $\mathfrak{L}_{\mathrm{N}}$ describes $L(\mathrm{N})$. Since $\mathfrak{L}$ describes L and since L is an extension of $L(\mathrm{N})$, we have $\vdash_P \mathrm{Func}\,\dot{\sigma}(\mathbf{f})\dot{n}$ and $\vdash_P \mathrm{Pred}\,\dot{\sigma}(\mathbf{p})\dot{n}$ for all n-ary function and predicate symbols of $L(\mathrm{N})$, and so $\mathfrak{L}$ is an extension of $\mathfrak{L}_{\mathrm{N}}$ by definition of $\mathrm{Func}_{\mathrm{N}}$ and $\mathrm{Pred}_{\mathrm{N}}$.

Suppose now that T is an extension of N and that $\mathfrak{T}$ describes T in P. We can construct $\mathfrak{L}_{\mathrm{N}}$ and σ_{N} as above so that $\mathfrak{L}_{\mathrm{N}}$ describes $L(\mathrm{N})$ and $\mathfrak{L}(\mathfrak{T})$ is an extension of $\mathfrak{L}_{\mathrm{N}}$. We can then define as at the beginning of this paragraph the arithmetical theory $\mathfrak{T}_{\mathrm{N}}$ with language $\mathfrak{L}_{\mathrm{N}}$ which describes N, and we claim that $\mathfrak{T}$ is an extension of $\mathfrak{T}_{\mathrm{N}}$. It suffices to prove $\vdash_P \mathrm{Thm}\,\dot{n}_i$ for $1 \le i \le 9$. This follows from the facts that $\mathfrak{T}$ describes T and that N1–N9 are theorems of T. Then by lemma 2 of §2.6, we have $\vdash_P \mathrm{Thm}_{\mathrm{N}}\,x \to \mathrm{Thm}\,x$, and so

Corollary. Let $\mathbf{A}$ be an RE-formula of $L(\mathrm{N})$ and let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be distinct variables including the variables free in $\mathbf{A}$. If $\mathfrak{T}$ describes an extension T of N in P, then

$$\vdash_P \mathbf{A} \to \mathrm{Thm}\,\mathrm{Sub}\ulcorner \mathbf{A}\urcorner\langle\ulcorner \mathbf{x}_1\urcorner,\dots,\ulcorner \mathbf{x}_n\urcorner\rangle\langle \mathrm{Num}\,\mathbf{x}_1,\dots,\mathrm{Num}\,\mathbf{x}_n\rangle.$$

Note that the corollary uses the fact that N has finitely many nonlogical symbols and nonlogical axioms in an essential way.

4.2 The theorems on consistency proofs. In this paragraph we shall prove several general versions of the result known as Gödel's second incompleteness theorem. We now assume that our fixed coding function symbol B is recursive on $L(\mathrm{N})$ in P. An arithmetical theory $\mathfrak{T}$ in P will be called *recursively enumerable* if Ω, Vr, Func, Pred, $\dot{\vee}$, $\dot{\neg}$, $\dot{\exists}$, $\dot{=}$, $\ddot{0}$, $\dot{\mathrm{S}}$ are recursive on $L(\mathrm{N})$ and if Nlax is recursively enumerable on $L(\mathrm{N})$. In this case, Thm is recursively enumerable on $L(\mathrm{N})$.

Theorem. Let P_0 be an extension of PA, P a good conservative extension of P_0, T an extension of P_0 arithmetized from a numerotation, and $\mathfrak{T}$ a recursively enumerable arithmetical theory in P which describes T. Let $\mathbf{D}$ be a formula of P_0 in which only $\mathbf{x}$ is free. If $\vdash_P \mathrm{Thm}\,\mathbf{x} \leftrightarrow \mathbf{D}$, then $\mathbf{D}$ with $\mathbf{x}$ satisfies in P_0 the derivability conditions for T.

Proof. The first derivability condition is a special case of the positive representability of thm_T by Thm in P, and the second derivability condition follows from $\vdash_P \ulcorner\mathbf{A} \to \mathbf{B}\urcorner = \mathrm{Imp}\ulcorner\mathbf{A}\urcorner\ulcorner\mathbf{B}\urcorner$ and an instance of (ii) of §2.6.

Since Thm is recursively enumerable on $L(\mathrm{N})$, there is an RE-formula $\mathbf{A}$ of $L(\mathrm{N})$ in which only $\mathbf{x}$ is free such that $\vdash_P \mathrm{Thm}\,\mathbf{x} \leftrightarrow \mathbf{A}$. By the hypothesis on $\mathbf{D}$, $\vdash_P \mathbf{D} \leftrightarrow \mathbf{A}$. Since P_0 and hence T is an extension of N, the corollary of §4.1 yields $\vdash_P \mathbf{A} \to \mathrm{Thm}\,\mathrm{Sub}\ulcorner\mathbf{A}\urcorner\langle\ulcorner\mathbf{x}\urcorner\rangle\langle\mathrm{Num}\,\mathbf{x}\rangle$, whence

$$\vdash_P \mathbf{D} \to \mathbf{D}[\mathbf{x}|\,\mathrm{Sub}\ulcorner\mathbf{A}\urcorner\langle\ulcorner\mathbf{x}\urcorner\rangle\langle\mathrm{Num}\,\mathbf{x}\rangle]. \tag{19}$$

By conservativity, we also have $\vdash_{P_0} \mathbf{D} \leftrightarrow \mathbf{A}$, and since T is an extension of P_0, $\vdash_T \mathbf{D} \leftrightarrow \mathbf{A}$, whence $\vdash_P \mathrm{Thm}\ulcorner\mathbf{D} \leftrightarrow \mathbf{A}\urcorner$ by positive representability. By the formalized substitution rule, $\vdash_P \mathrm{Thm}\,\mathrm{Sub}\ulcorner\mathbf{D} \leftrightarrow \mathbf{A}\urcorner\langle\ulcorner\mathbf{x}\urcorner\rangle\langle\mathrm{Num}\,\mathbf{x}\rangle$. By properties of Sub and (viii) of §2.6,

$$\vdash_P \mathrm{Thm}\,\mathrm{Sub}\ulcorner\mathbf{D}\urcorner\langle\ulcorner\mathbf{x}\urcorner\rangle\langle\mathrm{Num}\,\mathbf{x}\rangle \leftrightarrow \mathrm{Thm}\,\mathrm{Sub}\ulcorner\mathbf{A}\urcorner\langle\ulcorner\mathbf{x}\urcorner\rangle\langle\mathrm{Num}\,\mathbf{x}\rangle,$$

and hence, by the hypothesis on $\mathbf{D}$,

$$\vdash_P \mathbf{D}[\mathbf{x}|\,\mathrm{Sub}\ulcorner\mathbf{D}\urcorner\langle\ulcorner\mathbf{x}\urcorner\rangle\langle\mathrm{Num}\,\mathbf{x}\rangle] \leftrightarrow \mathbf{D}[\mathbf{x}|\,\mathrm{Sub}\ulcorner\mathbf{A}\urcorner\langle\ulcorner\mathbf{x}\urcorner\rangle\langle\mathrm{Num}\,\mathbf{x}\rangle].$$

From this and (19), we obtain

$$\vdash_P \mathbf{D} \to \mathbf{D}[\mathbf{x}|\,\mathrm{Sub}\ulcorner\mathbf{D}\urcorner\langle\ulcorner\mathbf{x}\urcorner\rangle\langle\mathrm{Num}\,\mathbf{x}\rangle]. \tag{20}$$

Let $\mathbf{A}$ be a closed formula of T. From (20) by the substitution rule, we obtain

$$\vdash_P \mathbf{D}[\mathbf{x}|\ulcorner\mathbf{A}\urcorner] \to \mathbf{D}[\mathbf{x}|\,\mathrm{Sub}\ulcorner\mathbf{D}\urcorner\langle\ulcorner\mathbf{x}\urcorner\rangle\langle\mathrm{Num}\ulcorner\mathbf{A}\urcorner\rangle].$$

But $\vdash_P \mathrm{Sub}\ulcorner\mathbf{D}\urcorner\langle\ulcorner\mathbf{x}\urcorner\rangle\langle\mathrm{Num}\ulcorner\mathbf{A}\urcorner\rangle = \ulcorner\mathbf{D}[\mathbf{x}|\ulcorner\mathbf{A}\urcorner]\urcorner$ because $\mathfrak{L}(\mathfrak{T})$ describes $L(T)$, and hence

$$\vdash_P \mathbf{D}[\mathbf{x}|\ulcorner\mathbf{A}\urcorner] \to \mathbf{D}[\mathbf{x}|\ulcorner\mathbf{D}[\mathbf{x}|\ulcorner\mathbf{A}\urcorner]\urcorner].$$

Since P is a conservative extension of P_0, this proves that $\mathbf{D}$ with $\mathbf{x}$ satisfies in P_0 the third derivability condition for T. □

Note that (ii) of §2.6 (the formalized detachment rule) and (20) above are much stronger than the second and third derivability conditions: the latter consist of infinitely many numerical instances of the former.

Theorem on Consistency Proofs 1. Let P_0 be an extension of PA, P a good conservative extension of P_0, T an extension of P_0 arithmetized from a numerotation, and $\mathfrak{T}$ a recursively enumerable arithmetical theory in P which describes T. If $\mathbf{C}$ is a formula of P_0 such that $\vdash_P \mathbf{C} \to \mathrm{Con}$ and if $\vdash_T \mathbf{C}$, then T is inconsistent.

Proof. By (xi) of §2.6, $\vdash_P \mathrm{Con} \leftrightarrow \forall x(\mathrm{Fm}\,x \to \neg\,\mathrm{Thm}\,x \vee \neg\,\mathrm{Thm}\,\mathrm{Neg}\,x)$, whence by the substitution theorem, $\vdash_P \mathrm{Con} \to \mathrm{Fm}\ulcorner\dot{0} = \dot{0}\urcorner \to \neg\,\mathrm{Thm}\ulcorner\dot{0} = \dot{0}\urcorner \vee \neg\,\mathrm{Thm}\,\mathrm{Neg}\ulcorner\dot{0} = \dot{0}\urcorner$. But $\vdash_P \mathrm{Fm}\ulcorner\dot{0} = \dot{0}\urcorner$, $\vdash_P \mathrm{Thm}\ulcorner\dot{0} = \dot{0}\urcorner$, and $\vdash_P \mathrm{Neg}\ulcorner\dot{0} = \dot{0}\urcorner = \ulcorner\dot{0} \neq \dot{0}\urcorner$ because $\mathfrak{T}$ describes T, and so by the tautology theorem and the equality theorem,

$$\vdash_P \mathbf{C} \to \neg\,\mathrm{Thm}\ulcorner\dot{0} \neq \dot{0}\urcorner. \tag{21}$$

Since $\mathfrak{T}$ is recursively enumerable, there is a formula $\mathbf{D}$ of $L(\mathrm{N})$ in which only x is free such that $\vdash_P \mathrm{Thm}\, x \leftrightarrow \mathbf{D}$. Then by the previous theorem, $\mathbf{D}$ with x satisfies in P_0 (and therefore also in T) the derivability conditions for T. By (21), $\vdash_P \mathbf{C} \to \neg\mathbf{D}[x|\ulcorner \dot{0} \neq \dot{0}\urcorner]$, whence by conservativity, $\vdash_{P_0} \mathbf{C} \to \neg\mathbf{D}[x|\ulcorner \dot{0} \neq \dot{0}\urcorner]$. By the hypothesis and the detachment rule, $\vdash_T \neg\mathbf{D}[x|\ulcorner \dot{0} \neq \dot{0}\urcorner]$. By the corollary of §1.2, T is inconsistent. □

Note that a formula $\mathbf{C}$ as in the statement of the theorem on consistency proofs always exists. This is because there are formulae $\mathbf{A}$ and $\mathbf{B}$ of $L(\mathrm{N})$ satisfying $\vdash_P \mathrm{Fm}\, x \leftrightarrow \mathbf{A}$ and $\vdash_P \mathrm{Thm}\, x \leftrightarrow \mathbf{B}$, and hence we can take $\mathbf{C}$ to be the formula $\exists x(\mathbf{A} \wedge \neg\mathbf{B})$ of $L(\mathrm{N})$.

Theorem on Consistency Proofs 2. Let P_0, P, T, and $\mathfrak{T}$ be as in the first theorem on consistency proofs. Let T' be a first-order theory arithmetized from a numerotation, I an interpretation of T in T', $\mathfrak{T}'$ an arithmetical theory in P which describes T', and $\mathfrak{I}$ an arithmetical interpretation in P of $\mathfrak{T}$ in $\mathfrak{T}'$ which describes I. If $\mathbf{C}$ is a formula of P_0 such that $\vdash_P \mathbf{C} \to \mathrm{Con}'$ and if $\vdash_{T'} \mathbf{C}^I$, then $T[\mathbf{C}]$ and T' are inconsistent.

Proof. We define an arithmetical theory $\mathfrak{T}^*$ in an extension by definitions P' of P as follows. Its language is $\mathfrak{L}(\mathfrak{T})$ and $\mathrm{Nlax}^*\, x \leftrightarrow \mathrm{Nlax}\, x \vee x = \ulcorner\mathbf{C}\urcorner$. Clearly $\mathfrak{T}^*$ describes $T[\mathbf{C}]$. Assume $\vdash_{T'} \mathbf{C}^I$. Now since $\vdash_P \mathrm{Int}\ulcorner\mathbf{C}\urcorner = \ulcorner\mathbf{C}^I\urcorner$ and $\vdash_P \mathrm{Thm}'\ulcorner\mathbf{C}^I\urcorner$, we find that $\mathfrak{I}$ is an arithmetical interpretation of $\mathfrak{T}^*$ in $\mathfrak{T}'$. By the corollary to the arithmetical interpretation theorem, $\vdash_P \mathrm{Con}' \to \mathrm{Con}^*$. As $\mathfrak{T}$ is recursively enumerable, $\mathfrak{T}^*$ is also recursively enumerable. Therefore we can find a formula $\mathbf{A}$ of $L(\mathrm{N})$ such that $\vdash_P \mathrm{Con}^* \leftrightarrow \mathbf{A}$, and we have $\vdash_P \mathbf{C} \to \mathbf{A}$. Since P is a conservative extension of P_0 and T is an extension of P_0, $\vdash_{T[\mathbf{C}]} \mathbf{A}$. By the first theorem on consistency proofs applied to P_0, P', $T[\mathbf{C}]$, and $\mathfrak{T}^*$, $T[\mathbf{C}]$ is inconsistent. But I is an interpretation of $T[\mathbf{C}]$ in T', so T' is inconsistent by the interpretation theorem. □

In practice, verifying that $\mathfrak{I}$ is an arithmetical interpretation of $\mathfrak{T}$ in $\mathfrak{T}'$ can be quite technical. It turns out that, to obtain the conlusion of the second theorem on consistency proofs, it is enough that $\mathfrak{I}$ be an arithmetical interpretation of $\mathfrak{L}(\mathfrak{T})$ in $\mathfrak{T}'$, *provided that* $\mathfrak{T}_{\mathfrak{I}}$ *is recursively enumerable*. For then, as we know, $\mathfrak{I}$ is an interpretation of $\mathfrak{T}_{\mathfrak{I}}$ in $\mathfrak{T}'$ and $\mathfrak{T}_{\mathfrak{I}}$ describes T, so the hypotheses of the theorem remain valid if we replace $\mathfrak{T}$ by $\mathfrak{T}_{\mathfrak{I}}$. For $\mathfrak{T}_{\mathfrak{I}}$ to be recursively enumerable, it suffices that $\mathfrak{U}_{\mathfrak{I}}$ and $()_{\mathfrak{I}}$ be recursive on $L(\mathrm{N})$ and that $\mathfrak{T}'$ be recursively enumerable. We shall discuss a concrete example of this in §6.3.

4.3 A counterexample. In this paragraph we consider the following specialization of the hypotheses of the theorems on consistency proofs: P_0 is a good extension of PA and P is an extension by definitions of P_0. Let T be an extension of P_0, and let $\mathfrak{T}$ be an arithmetical theory in P. We say that the pair $(T, \mathfrak{T})$ is *reflexive* if $\vdash_T \mathrm{Con}_{\restriction \dot{n}}$ for all n (by this we really mean that a translation of $\mathrm{Con}_{\restriction \dot{n}}$ into P_0 is a theorem of T). More generally, let I be an interpretation of P_0 in T, with P_0, P, and $\mathfrak{T}$ as before. Assume that $=_I$ is $=$ and that I satisfies the other nonrestrictive assumptions of §3.6. We say that the pair $(T, \mathfrak{T})$ is *reflexive with respect to* I if $\vdash_T (\mathrm{Con}_{\restriction \dot{n}})^I$ for all n. This concept is only interesting when $\mathfrak{T}$ describes T with respect to I, in which case its meaning is close to "T proves the consistency of all its subtheories with finitely many nonlogical axioms". We shall see that this implies "T can prove its own consistency". However, we will not be able to deduce that T is inconsistent, because one hypothesis of the theorem on consistency proofs will not be fulfilled, namely, that of recursive enumerability.

The result of this paragraph was discovered by Feferman [3]. Its relevancy will only appear once we know of interesting theories having the reflexivity property. We do not discuss the question of reflexivity further here (but see §6.3).

We shall admit the following result without proof: if P is a good extension of PA and $\mathfrak{T}$ is an arithmetical theory in P, then $\vdash_P \mathrm{Con}_{\restriction \dot{0}}$. This is of course a formalization of the fact that a first-order theory with no nonlogical axioms is consistent (cf. ch. II §1.2). The reader should convince himself that the methods used in the proof of this fact are all amenable to formalization within PA.

Lemma. Let P be a good extension of PA and $\mathfrak{T}$ an arithmetical theory in P. Let $\mathfrak{T}'$ be the arithmetical theory with the same language as $\mathfrak{T}$ and with $\mathrm{Nlax}'\, x \leftrightarrow \mathrm{Nlax}\, x \wedge \mathrm{Con}_{\restriction \mathrm{S}x}$. Then $\vdash_P \mathrm{Con}'$.

Proof. Note that $\mathfrak{T}$ is an extension of $\mathfrak{T}'$, and hence

$$\vdash_P \mathrm{Con} \to \mathrm{Con}' . \tag{22}$$

By (iii) of §2.4, $\vdash_P \neg\,\mathrm{Con} \to \exists z \neg\,\mathrm{Con}_{\restriction z}$. By the least number principle,

$$\vdash_P \neg\,\mathrm{Con} \to \exists z(\neg\,\mathrm{Con}_{\restriction z} \wedge \forall x(x < z \to \mathrm{Con}_{\restriction x})). \tag{23}$$

Now $\vdash_P \mathrm{Con}_{\restriction \dot{0}}$, and so $\vdash_P \neg\,\mathrm{Con}_{\restriction z} \to \exists y(z = \mathrm{S}y)$. From this and (23), we obtain

$$\vdash_P \neg\,\mathrm{Con} \to \exists y(\neg\,\mathrm{Con}_{\restriction \mathrm{S}y} \wedge \forall x(x < \mathrm{S}y \to \mathrm{Con}_{\restriction x})). \tag{24}$$

Using the obvious theorems $\vdash_P x < y \to \mathrm{S}x < \mathrm{S}y$, $\vdash_P \forall x(x < \mathrm{S}y \to \mathrm{Con}_{\restriction x}) \to \mathrm{Con}_{\restriction y}$, and $\vdash_P \neg\,\mathrm{Con}_{\restriction \mathrm{S}z} \to \mathrm{Con}_{\restriction \mathrm{S}x} \to x < z$ together with (24), we infer $\vdash_P \neg\,\mathrm{Con} \to \exists z(\mathrm{Con}_{\restriction z} \wedge \forall x(x < z \leftrightarrow \mathrm{Con}_{\restriction \mathrm{S}w}))$. From the definition of Nlax′ comes $\vdash_P \neg\,\mathrm{Con} \to \exists z(\mathrm{Con}_{\restriction z} \wedge \forall x(\mathrm{Nlax}_{\restriction z}\, x \leftrightarrow \mathrm{Nlax}'\, x))$. By lemma 2 of §2.6 and the remark following it, $\vdash_P \forall x(\mathrm{Nlax}_{\restriction z}\, x \leftrightarrow \mathrm{Nlax}'\, x) \to \mathrm{Con}_{\restriction z} \leftrightarrow \mathrm{Con}'$, and therefore $\vdash_P \neg\,\mathrm{Con} \to \exists z(\mathrm{Con}_{\restriction z} \wedge (\mathrm{Con}_{\restriction z} \leftrightarrow \mathrm{Con}'))$, whence

$$\vdash_P \neg\,\mathrm{Con} \to \mathrm{Con}'. \tag{25}$$

From (22) and (25) by the tautology theorem, $\vdash_P \mathrm{Con}'$. □

Theorem. Let P_0, P, T, and I be as above. Let $\mathfrak{T}$ be an arithmetical theory that describes (resp. represents) T with respect to I. If $(T, \mathfrak{T})$ is reflexive with respect to I, there is an arithmetical theory $\mathfrak{T}'$ in an extension by definitions of P which describes (resp. represents) T with respect to I and such that $\vdash_P \mathrm{Con}'$. In particular, $\vdash_T (\mathrm{Con}')^I$.

Proof. Define $\mathfrak{T}'$ as in the lemma, so that $\vdash_P \mathrm{Con}'$. Suppose that $\mathfrak{T}$ describes T with respect to I. Let $\mathbf{A}$ be a nonlogical axiom of T. Then $\vdash_T (\mathrm{Nlax}\ulcorner\mathbf{A}\urcorner)^I$ by hypothesis and $\vdash_T (\mathrm{Con}_{\restriction \mathrm{S}\ulcorner\mathbf{A}\urcorner})^I$ by reflexivity, so $\vdash_T (\mathrm{Nlax}'\ulcorner\mathbf{A}\urcorner)^I$ by definition of Nlax′ and the interpretation theorem. If moreover $\mathfrak{T}$ represents T and a is not the expression number of a nonlogical axiom of T, then $\vdash_T (\neg\,\mathrm{Nlax}\,\dot{a})^I$ and so $\vdash_T (\neg\,\mathrm{Nlax}'\,\dot{a})^I$ by the tautology theorem. □

§5 Arithmetical completeness

5.1 The interpretation in PA. In this section we shall prove that any reasonable first-order theory T has an interpretation in the first-order theory obtained from PA by adding a suitable axiom expressing the consistency of T. The proof is a direct formalization of the result of model theory known as the completeness theorem. We begin with the definition of the interpretation.

Let T be a first-order theory arithmetized from a numerotation, and let $\mathfrak{T}$ be an arithmetical theory in P which describes T. We define an interpretation I of $L(T)$ in $L(P')$, where P' is an extension by definitions of P, by

(i) $\mathrm{U}_I x \leftrightarrow \mathrm{Tm}\, x \wedge \mathrm{Cl}\, x$;

(ii) $\mathbf{f}_I x_1 \ldots x_n = \langle \dot{\sigma}(\mathbf{f}), x_1, \ldots, x_n \rangle$;

(iii) $\mathbf{p}_I x_1 \ldots x_n \leftrightarrow \mathrm{Thm}\langle \dot{\sigma}(\mathbf{p}), x_1, \ldots, x_n \rangle$.

Lemma 1. Let $\mathbf{u}$ be a designator of $L(T)$. If $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are distinct variables including the variables free in $\mathbf{u}$, then $\vdash_P \mathrm{U}_I \mathbf{x}_1 \to \cdots \to \mathrm{U}_I \mathbf{x}_n \to \mathrm{Cl}\,\mathrm{Sub}\ulcorner\mathbf{u}\urcorner\langle\ulcorner\mathbf{x}_1\urcorner, \ldots, \ulcorner\mathbf{x}_n\urcorner\rangle\langle\mathbf{x}_1, \ldots, \mathbf{x}_n\rangle$.

Proof. This follows from §2.2 (xv′) and the fact that $\mathfrak{L}(\mathfrak{T})$ describes $L(T)$. □

Lemma 2. If $\vdash_P \exists x\, \mathrm{Func}\, x\dot{0}$, then I is an interpretation of $L(T)$ in P'.

Proof. We must prove

$$\vdash_P \exists x \mathrm{U}_I x \tag{1}$$

and for every function symbol $\mathbf{f}$ of T,

$$\vdash_P \mathrm{U}_I x_1 \to \cdots \to \mathrm{U}_I x_n \to \mathrm{U}_I \mathbf{f}_I x_1 \ldots x_n. \tag{2}$$

Now (1) follows from the hypothesis and $\vdash_P \mathrm{Func}\, x\dot{0} \to \mathrm{U}_I\langle x\rangle$ by the $\exists$-introduction rule and the substitution axioms. Let $\mathbf{f}$ be an n-ary function symbol of T. Then $\vdash_P \mathrm{Tm}\ulcorner\mathbf{f}x_1 \ldots x_n\urcorner$, and hence $\vdash_P \mathrm{Tm}\,\mathbf{f}_I x_1 \ldots x_n$ by (xi′) of §2.2. Then (2) follows from this and lemma 1. □

LEMMA 3. For every term $\mathbf{a}$ of T, if $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are distinct variables including the variables occurring in $\mathbf{a}$,

$$\vdash_P \mathbf{a}_I = \mathrm{Sub}\ulcorner\mathbf{a}\urcorner\langle\ulcorner\mathbf{x}_1\urcorner, \ldots, \ulcorner\mathbf{x}_n\urcorner\rangle\langle\mathbf{x}_1, \ldots, \mathbf{x}_n\rangle.$$

Proof. By induction on the length of $\mathbf{a}$. If $\mathbf{a}$ is a variable, then, since $\mathbf{a}$ is one of the $\mathbf{x}_i$, $\vdash_P \mathbf{a} = \mathrm{Sub}\ulcorner\mathbf{a}\urcorner\langle\ulcorner\mathbf{x}_1\urcorner, \ldots, \ulcorner\mathbf{x}_n\urcorner\rangle\langle\mathbf{x}_1, \ldots, \mathbf{x}_n\rangle$. Suppose that $\mathbf{a}$ is $\mathbf{f}\mathbf{a}_1 \ldots \mathbf{a}_n$. By the definition of $\mathbf{f}_I$ and the substitution rule, $\vdash_P \mathbf{a}_I = \langle\dot{\sigma}(\mathbf{f}), (\mathbf{a}_1)_I, \ldots, (\mathbf{a}_n)_I\rangle$. By induction hypothesis,

$$\vdash_P \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to (\mathbf{a}_i)_I = \mathrm{Sub}\ulcorner\mathbf{a}_i\urcorner\langle\ulcorner\mathbf{x}_1\urcorner, \ldots, \ulcorner\mathbf{x}_n\urcorner\rangle\langle\mathbf{x}_1, \ldots, \mathbf{x}_n\rangle$$

for each i, and so

$$\begin{aligned}&\vdash_P \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \\ &\qquad \mathbf{a}_I = \langle\dot{\sigma}(\mathbf{f}), \mathrm{Sub}\ulcorner\mathbf{a}_i\urcorner\langle\ulcorner\mathbf{x}_1\urcorner, \ldots, \ulcorner\mathbf{x}_n\urcorner\rangle\langle\mathbf{x}_1, \ldots, \mathbf{x}_n\rangle, \ldots, \mathrm{Sub}\ulcorner\mathbf{a}_i\urcorner\langle\ulcorner\mathbf{x}_1\urcorner, \ldots, \ulcorner\mathbf{x}_n\urcorner\rangle\langle\mathbf{x}_1, \ldots, \mathbf{x}_n\rangle\rangle\end{aligned}$$

by the tautology theorem and the equality theorem. By the definition of Sub, we obtain $\vdash_P \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathbf{a}_I = \mathrm{Sub}\ulcorner\mathbf{a}\urcorner\langle\ulcorner\mathbf{x}_1\urcorner, \ldots, \ulcorner\mathbf{x}_n\urcorner\rangle\langle\mathbf{x}_1, \ldots, \mathbf{x}_n\rangle$. □

LEMMA 4. Suppose that $\vdash_P \mathrm{Con} \wedge \mathrm{Cm} \wedge \mathrm{Hk}$. For every formula $\mathbf{A}$ of T, if $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are distinct variables including the variables free in $\mathbf{A}$,

$$\vdash_P \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathbf{A}_I \leftrightarrow \mathrm{Thm}\,\mathrm{Sub}\ulcorner\mathbf{A}\urcorner\langle\ulcorner\mathbf{x}_1\urcorner, \ldots, \ulcorner\mathbf{x}_n\urcorner\rangle\langle\mathbf{x}_1, \ldots, \mathbf{x}_n\rangle.$$

Proof. We use induction on the length of $\mathbf{A}$. We shall denote by $\mathbf{u}$ the two-term expression $\langle\ulcorner\mathbf{x}_1\urcorner, \ldots, \ulcorner\mathbf{x}_n\urcorner\rangle\langle\mathbf{x}_1, \ldots, \mathbf{x}_n\rangle$. Suppose that $\mathbf{A}$ is $\mathbf{p}\mathbf{a}_1 \ldots \mathbf{a}_k$. By lemma 3, $\vdash_P (\mathbf{a}_i)_I = \mathrm{Sub}\ulcorner\mathbf{a}_i\urcorner\mathbf{u}$ for each i. By definition of $\mathbf{p}_I$ and the equality theorem, we obtain

$$\vdash_P \mathbf{A}_I \leftrightarrow \mathrm{Thm}\langle\dot{\sigma}(\mathbf{p}), \mathrm{Sub}\ulcorner\mathbf{a}_1\urcorner\mathbf{u}, \ldots, \mathrm{Sub}\ulcorner\mathbf{a}_k\urcorner\mathbf{u}\rangle,$$

whence by properties of Sub, $\vdash_P \mathbf{A}_I \leftrightarrow \mathrm{Thm}\,\mathrm{Sub}\ulcorner\mathbf{A}\urcorner\mathbf{u}$, from which the desired result follows by the tautology theorem.

Recall that, by (xiii) of §2.6, $\vdash_P \mathrm{Con} \wedge \mathrm{Cm}$ implies

$$\vdash_P \mathrm{Fm}\,x \to \mathrm{Cl}\,x \to \mathrm{Thm}\,\mathrm{Neg}\,x \leftrightarrow \neg\,\mathrm{Thm}\,x. \tag{3}$$

Suppose that $\mathbf{A}$ is $\mathbf{B} \vee \mathbf{C}$. By lemma 1, $\vdash_P \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathrm{Cl}\,\mathrm{Sub}\ulcorner\mathbf{A}\urcorner\mathbf{u}$, $\vdash_P \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathrm{Cl}\,\mathrm{Sub}\ulcorner\mathbf{B}\urcorner\mathbf{u}$, and $\vdash_P \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathrm{Cl}\,\mathrm{Sub}\ulcorner\mathbf{C}\urcorner\mathbf{u}$, so by (3),

$$\vdash_P \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathrm{Thm}\,\mathrm{Sub}\ulcorner\neg\mathbf{A}\urcorner\mathbf{u} \leftrightarrow \neg\,\mathrm{Thm}\,\mathrm{Sub}\ulcorner\mathbf{A}\urcorner\mathbf{u}, \tag{4}$$

$\vdash_P \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathrm{Thm}\,\mathrm{Sub}\ulcorner\neg\mathbf{B}\urcorner\mathbf{u} \leftrightarrow \neg\,\mathrm{Thm}\,\mathrm{Sub}\ulcorner\mathbf{B}\urcorner\mathbf{u}$ and $\mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathrm{Thm}\,\mathrm{Sub}\ulcorner\neg\mathbf{C}\urcorner\mathbf{u} \leftrightarrow \neg\,\mathrm{Thm}\,\mathrm{Sub}\ulcorner\mathbf{C}\urcorner\mathbf{u}$. Thus by (vi) of §2.6,

$$\vdash_P \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \neg\,\mathrm{Thm}\,\mathrm{Sub}\ulcorner\mathbf{B}\urcorner\mathbf{u} \wedge \neg\,\mathrm{Thm}\,\mathrm{Sub}\ulcorner\mathbf{C}\urcorner\mathbf{u} \leftrightarrow \mathrm{Thm}\,\mathrm{Sub}\ulcorner\neg\mathbf{B} \wedge \neg\mathbf{C}\urcorner\mathbf{u}. \tag{5}$$

By (vii) of §2.6,

$$\vdash_P \mathrm{Thm}\,\mathrm{Sub}\ulcorner\neg\mathbf{B} \wedge \neg\mathbf{C}\urcorner\mathbf{u} \leftrightarrow \mathrm{Thm}\,\mathrm{Sub}\ulcorner\neg\mathbf{A}\urcorner\mathbf{u}. \tag{6}$$

From (4), (5), and (6) by the tautology theorem, $\vdash_P \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathrm{Thm}\,\mathrm{Sub}\ulcorner\mathbf{B}\urcorner\mathbf{u} \vee \mathrm{Thm}\,\mathrm{Sub}\ulcorner\mathbf{C}\urcorner\mathbf{u} \leftrightarrow \mathrm{Thm}\,\mathrm{Sub}\ulcorner\mathbf{A}\urcorner\mathbf{u}$. By induction hypothesis and the tautology theorem, we obtain $\vdash_P \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathbf{A}_I \leftrightarrow \mathrm{Thm}\,\mathrm{Sub}\ulcorner\mathbf{A}\urcorner\mathbf{u}$, as required.

If $\mathbf{A}$ is $\neg\mathbf{B}$, the result follows from (4), with $\mathbf{B}$ instead of $\mathbf{A}$, the induction hypothesis, and the tautology theorem.

Finally, suppose that $\mathbf{A}$ is $\exists\mathbf{x}\mathbf{B}$. By properties of Sub, we may assume that $\mathbf{x}$ is not among $\mathbf{x}_1, \ldots, \mathbf{x}_n$, and we have

$$\vdash_P \mathrm{U}_I\mathbf{x} \to \mathrm{Sub}\,\mathrm{Sub}\ulcorner\mathbf{B}\urcorner\mathbf{u}\langle\ulcorner\mathbf{x}\urcorner\rangle\langle\mathbf{x}\rangle = \mathrm{Sub}\ulcorner\mathbf{B}\urcorner\langle\ulcorner\mathbf{x}\urcorner, \ulcorner\mathbf{x}_1\urcorner, \ldots, \ulcorner\mathbf{x}_n\urcorner\rangle\langle\mathbf{x}, \mathbf{x}_1, \ldots, \mathbf{x}_n\rangle,$$

so by induction hypothesis, the tautology theorem, and the equality theorem

$$\vdash_P \mathrm{U}_I\mathbf{x} \to \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathrm{Thm}\,\mathrm{Sub}\,\mathrm{Sub}\ulcorner\mathbf{B}\urcorner\mathbf{u}\langle\ulcorner\mathbf{x}\urcorner\rangle\langle\mathbf{x}\rangle \leftrightarrow \mathbf{B}_I. \tag{7}$$

Since $\vdash_P \mathrm{Hk}$, $\vdash_P \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathrm{Cl}\,\mathrm{Sub}\ulcorner\mathbf{A}\urcorner\mathbf{u}$, and $\vdash_P \mathrm{Func}\,x\dot{0} \to \mathrm{U}_I\langle x\rangle$, we obtain from (7) that $\vdash_P \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathrm{Thm}\,\mathrm{Sub}\ulcorner\mathbf{A}\urcorner\mathbf{u} \to \exists\mathbf{x}(\mathrm{U}_I\mathbf{x} \wedge \mathbf{B}_I)$, i.e.,

$$\vdash_P \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathrm{Thm}\,\mathrm{Sub}\ulcorner\mathbf{A}\urcorner\mathbf{u} \to \mathbf{A}_I. \tag{8}$$

The formula $\mathbf{B} \to \mathbf{A}$ is a substitution axiom of T, and so by the formalized substitution rule and the formalized detachment rule, $\vdash_P \mathrm{Thm}\,\mathrm{Sub}\ulcorner\mathbf{B}\urcorner\langle\ulcorner\mathbf{x}\urcorner, \ulcorner\mathbf{x}_1\urcorner, \ldots, \ulcorner\mathbf{x}_n\urcorner\rangle\langle\mathbf{x}, \mathbf{x}_1, \ldots, \mathbf{x}_n\rangle \to \mathrm{Thm}\,\mathrm{Sub}\ulcorner\mathbf{A}\urcorner\mathbf{u}$, whence by the induction hypothesis, the tautology theorem, and the $\exists$-introduction rule, $\vdash_P \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathbf{A}_I \to \mathrm{Thm}\,\mathrm{Sub}\ulcorner\mathbf{A}\urcorner\mathbf{u}$. Together with (8), this completes the proof. □

In the statement of the following theorem, "an extension by definitions of $P[\mathrm{Con}, \mathrm{Cm}, \mathrm{Hk}]$" means of course an extension by definitions of $P'[\mathrm{Con}, \mathrm{Cm}, \mathrm{Hk}]$ for some extension by definitions P' of P in which Con, Cm, and Hk are defined. We shall commit this abuse of terminology several times before the end of this chapter.

Theorem. If $\vdash_P \exists x\,\mathrm{Func}\,x\dot{0}$, then I is an interpretation of T in an extension by definitions of $P[\mathrm{Con}, \mathrm{Cm}, \mathrm{Hk}]$.

Proof. We prove directly that $\vdash_{P[\mathrm{Con},\mathrm{Cm},\mathrm{Hk}]} \mathbf{A}^I$ for every theorem $\mathbf{A}$ of T. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be the variables free in $\mathbf{A}$ in reverse alphabetical order. Since $P[\mathrm{Con}, \mathrm{Cm}, \mathrm{Hk}]$ is a good extension of PA satisfying the hypothesis of lemma 4, we have

$$\vdash_{P[\mathrm{Con},\mathrm{Cm},\mathrm{Hk}]} \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathbf{A}_I \leftrightarrow \mathrm{Thm}\,\mathrm{Sub}\ulcorner\mathbf{A}\urcorner\langle\ulcorner\mathbf{x}_1\urcorner, \ldots, \ulcorner\mathbf{x}_n\urcorner\rangle\langle\mathbf{x}_1, \ldots, \mathbf{x}_n\rangle.$$

Since $\mathfrak{T}$ describes T, $\vdash_P \mathrm{U}_I\mathbf{x}_1 \to \cdots \to \mathrm{U}_I\mathbf{x}_n \to \mathrm{Thm}\,\mathrm{Sub}\ulcorner\mathbf{A}\urcorner\langle\ulcorner\mathbf{x}_1\urcorner, \ldots, \ulcorner\mathbf{x}_n\urcorner\rangle\langle\mathbf{x}_1, \ldots, \mathbf{x}_n\rangle$ by the formalized substitution rule. By the tautology theorem, $\vdash_{P[\mathrm{Con},\mathrm{Cm},\mathrm{Hk}]} \mathbf{A}^I$. □

5.2 The arithmetical completeness theorem. As in the proof of the completeness theorem, it remains to build an extension $\mathfrak{T}'$ of $\mathfrak{T}$ for which $\vdash_P \mathrm{Con} \to \mathrm{Con}' \wedge \mathrm{Cm}' \wedge \mathrm{Hk}'$. Here we only state the relevant results without complete proofs.

Arithmetical Henkin's Lemma. Let P be a good extension of PA in which there is an arithmetical theory $\mathfrak{T}$. Then in some recursive extension by definitions P' of P there are arithmetical theories $\mathfrak{T}^*$ and $\mathfrak{T}_c$ such that $\mathfrak{T}^*$ differs from $\mathfrak{T}$ by a change of numerotation, $\mathfrak{T}_c$ is an extension of $\mathfrak{T}^*$, $\vdash_{P'} \mathrm{Hk}_c$, and $\vdash_{P'} \mathrm{Fm}^*\,x \to \mathrm{Thm}_c\,x \to \mathrm{Thm}^*\,x$.

Define a unary function symbol $\mathbf{f}$ by $\mathbf{f}x = x\cdot\dot{2}$. Then $\vdash_P \mathbf{f}x = \mathbf{f}y \to x = y$. We let $\mathfrak{T}^*$ be the arithmetical theory $\mathfrak{T}_{\mathbf{f}}$, so that $\mathbf{f}$ is a change of numerotation from $\mathfrak{T}$ to $\mathfrak{T}^*$. Define $\mathfrak{L}(\mathfrak{T}_c)$ as follows: $\mathrm{Vr}_c\,x = \mathrm{Vr}^*\,x$; $\mathrm{Func}_c\,xy \leftrightarrow \mathrm{Func}^*\,xy \vee y = \dot{0} \wedge \exists z(x = \mathrm{S}(z\cdot\dot{2}))$; $\mathrm{Pred}_c\,xy \leftrightarrow \mathrm{Pred}^*\,xy$; for $\mathbf{e}$ among $\dot{\vee}$, $\dot{\neg}$, $\dot{\exists}$, $\dot{=}$, $\dot{0}$, and $\dot{\mathrm{S}}$, $\mathbf{e}_c = \mathbf{e}^*$. Set

$$\mathrm{Nlax}_c\,x \leftrightarrow \mathrm{Nlax}^*\,x \vee \exists y(y \le x \wedge x = \mathrm{Imp}_c\,\mathbf{g}y\,\mathrm{Sub}_c(\mathbf{g}y)_2\langle(\mathbf{g}y)_1\rangle\langle\langle\mathrm{S}(y\cdot\dot{2})\rangle\rangle)$$

where $\mathbf{g}$ is defined by recursion so that $\vdash_P \mathbf{g}x = \mu y(\mathrm{Cl}_c\,y \wedge y = \mathrm{Inst}_c(y)_1(y)_2 \wedge \forall z(z < y \to y \ne \mathbf{g}z))$. It is then easy to prove that $\vdash_P \mathrm{Hk}_c$. The proof of the last assertion, however, is not as easy. It mainly requires to derive in P a formalized version of the theorem on constants. Obviously the last assertion implies $\vdash_P \mathrm{Con}^* \to \mathrm{Con}_c$, and by the proposition of §2.5, we have in fact $\vdash_P \mathrm{Con} \to \mathrm{Con}_c$.

Of course, $\mathfrak{T}_c$ also describes a suitably arithmetized Henkin extension T_c of T, but this is of no interest here.

Remark. The construction of the arithmetical Henkin extension $\mathfrak{T}_c$ and the derivation of its conservativity over $\mathfrak{T}$ are the first step towards an arithmetical Herbrand's theorem. As Herbrand's theorem is a basis for many finitary proofs of consistency, this can be used to formalize those proofs within PA. We can prove, for instance, that $\vdash_{\mathrm{PA}} \mathrm{Con}_{\mathrm{N}}$ for an appropriate arithmetical theory $\mathfrak{T}_{\mathrm{N}}$ describing N in PA. This is an argument in favour of the informal thesis that all finitary reasonings can be formalized in PA, or at least those used in consistency proofs; if one believes this thesis, then the theorem on consistency proofs acquires a new meaning, namely: we cannot hope to prove the consistency of a first-order theory satisfying the hypotheses of the theorem by finitary methods.

ARITHMETICAL LINDENBAUM'S LEMMA. Let P be a good extension of PA in which there is an arithmetical theory $\mathfrak{T}$. Then in some extension by definitions P' of P there is an extension $\mathfrak{T}'$ of $\mathfrak{T}$ with the same language as $\mathfrak{T}$ such that $\vdash_{P'}\mathrm{Cm}'$ and $\vdash_{P'}\mathrm{Con} \to \mathrm{Con}'$.

Define $\mathbf{g}$ by

$$y = \mathbf{g}x \leftrightarrow (\mathrm{Con} \wedge \Omega[x] \wedge y = \mu z(\mathrm{Fm}\, z \wedge \mathrm{Cl}\, z \wedge \neg(z \in x) \wedge \neg\, \mathrm{Thm}[x]\, \mathrm{Neg}\, z)) \\ \vee (\neg\, \mathrm{Con} \wedge \Omega[x] \wedge y = \mu z(\mathrm{Fm}\, z \wedge \mathrm{Cl}\, z \wedge \neg(z \in x))) \vee (\neg\Omega[x] \wedge y = \dot{0}).$$

The necessary existence conditions for this definition follow from $\vdash_P \mathrm{Fm}\,\mathbf{a} \wedge \mathrm{Cl}\,\mathbf{a} \wedge \neg(\mathbf{a} \in x) \wedge \neg(\mathrm{Neg}\,\mathbf{a} \in x)$ where $\mathbf{a}$ is $\mathrm{Inst}\langle \mathrm{Vr}\, x\rangle\langle \dot{=}, \langle \mathrm{Vr}\, x\rangle, \langle \mathrm{Vr}\, x\rangle\rangle$. Then define $\mathbf{h}$ by recursion so that $\vdash_P \mathbf{h}\dot{0} = \langle \mathbf{g}\dot{0}\rangle$ and $\vdash_P x \neq \dot{0} \to \mathbf{h}x = \langle \mathbf{gh}(x - \dot{1})\rangle * \mathbf{h}(x - \dot{1})$, and finally let $\mathrm{Nlax}'\, x \leftrightarrow \exists y(x = (\mathbf{h}y)_0)$. The proof that $\vdash_P \mathrm{Cm}'$ is a straightforward inspection of the definitions. The proof that $\vdash_P \mathrm{Con} \to \mathrm{Con}'$ uses (iii) of §2.4. That we cannot expect P' to be a recursive extension of P reflects the nonconstructivity of the classical lemma.

It is clear from the definition of Hk that if $\mathfrak{T}'$ is an extension of $\mathfrak{T}$ with the same language as $\mathfrak{T}$, then $\vdash_P \mathrm{Hk} \to \mathrm{Hk}'$.

ARITHMETICAL COMPLETENESS THEOREM. Let T be a first-order theory arithmetized from a numerotation. Let P be a good extension of PA and $\mathfrak{T}$ an arithmetical theory in P which describes T. Then there is an interpretation of T in an extension by definitions of $P[\mathrm{Con}]$.

Proof. Let $\mathfrak{T}_c$ be constructed from $\mathfrak{T}$ as in Henkin's lemma, so that $\vdash_P \mathrm{Con} \to \mathrm{Con}_c$. Let $\mathfrak{T}'$ be constructed from $\mathfrak{T}_c$ as in Lindenbaum's lemma, so that $\vdash_P \mathrm{Cm}'$ and $\vdash_P \mathrm{Con}_c \to \mathrm{Con}'$. Then $\vdash_{P[\mathrm{Con}]} \mathrm{Con}' \wedge \mathrm{Cm}' \wedge \mathrm{Hk}'$ and $\vdash_P \exists x\, \mathrm{Func}'\, x\dot{0}$. Since $\mathfrak{T}'$ is an extension of $\mathfrak{T}$, $\mathfrak{T}'$ describes T. By the theorem of §5.1, we find an interpretation I of T in an extension by definitions of $P[\mathrm{Con}]$. ☐

§6 Application to the first-order theory ZF

6.1 Describing PA. In this section we choose our fixed coding function symbol B in a recursive extension by definitions of PA. We begin by defining an arithmetical theory $\mathfrak{P}$ in a recursive extension by definitions of PA. The symbols associated with $\mathfrak{P}$ will be written with the index "PA". The arithmetical language $\mathfrak{L}(\mathfrak{P})$ is given by the defining axioms: $\mathrm{Vr}_{\mathrm{PA}}\, x = x + \dot{9}$, $\mathrm{Func}_{\mathrm{PA}}\, xy \leftrightarrow (y = \dot{0} \wedge x = \dot{4}) \vee (y = \dot{1} \wedge x = \dot{5}) \vee (y = \dot{2} \wedge x = \dot{6} \vee x = \dot{7})$, $\mathrm{Pred}_{\mathrm{PA}}\, xy \leftrightarrow (y = \dot{2} \wedge x = \dot{3} \vee x = \dot{8})$, $\dot{\vee}_{\mathrm{PA}} = \dot{0}$, $\dot{\neg}_{\mathrm{PA}} = \dot{1}$, $\dot{\exists}_{\mathrm{PA}} = \dot{2}$, $\dot{=}_{\mathrm{PA}} = \dot{3}$, $\ddot{0}_{\mathrm{PA}} = \dot{4}$, $\dot{\mathrm{S}}_{\mathrm{PA}} = \dot{5}$. It is obvious that $\mathfrak{L}(\mathfrak{P})$ is an arithmetical language in (an extension by definitions of) PA. We let σ_{PA} be the numerotation $L(\mathrm{N})$ defined by: $\sigma_{\mathrm{PA}}(\vee) = 0$, $\sigma_{\mathrm{PA}}(\neg) = 1$, $\sigma_{\mathrm{PA}}(\exists) = 2$, $\sigma_{\mathrm{PA}}(=) = 3$, $\sigma_{\mathrm{PA}}(\dot{0}) = 4$, $\sigma_{\mathrm{PA}}(\mathrm{S}) = 5$, $\sigma_{\mathrm{PA}}(+) = 6$, $\sigma_{\mathrm{PA}}(\cdot) = 7$, $\sigma_{\mathrm{PA}}(<) = 8$, and if $\mathbf{x}$ is the $(n+1)$th variable in the alphabetical order, set $\sigma_{\mathrm{PA}}(\mathbf{x}) = n + 9$; we endow $L(\mathrm{N})$ with the arithmetization obtained from σ_{PA} by β. It is then equally obvious that $\mathfrak{L}(\mathfrak{P})$ represents $L(\mathrm{N})$. We let $n_1, \ldots, n_8$ be the expression numbers of the axioms N1–N8 and we define

$$\mathrm{Nlax}_{\mathrm{PA}}\, x \leftrightarrow x = \dot{n}_1 \vee \cdots \vee x = \dot{n}_8 \vee \exists y \exists z(y < x \wedge z < x \wedge \mathrm{Fm}_{\mathrm{PA}}\, y \wedge \mathrm{Vble}_{\mathrm{PA}}\, z \\ \wedge x = \mathrm{Imp}_{\mathrm{PA}}\, \mathrm{Sub}_{\mathrm{PA}}\, y\langle z\rangle\langle \ddot{0}_{\mathrm{PA}}\rangle\, \mathrm{Imp}_{\mathrm{PA}}\, \mathrm{Gen}_{\mathrm{PA}}\, z\, \mathrm{Imp}_{\mathrm{PA}}\, y\, \mathrm{Sub}_{\mathrm{PA}}\, y\langle z\rangle\langle\langle \dot{\mathrm{S}}_{\mathrm{PA}}, z\rangle\rangle y).$$

Observe that $\mathrm{Nlax}_{\mathrm{PA}}$ is recursive on $L(\mathrm{N})$. By the results of this chapter, it is clear that $\vdash_{\mathrm{PA}} \mathrm{Nlax}_{\mathrm{PA}}\, x \to \mathrm{Fm}_{\mathrm{PA}}\, x$, so that $\mathfrak{P}$ is an arithmetical theory in PA, and that $\mathfrak{P}$ represents PA. Since all the hypotheses of the first theorem on consistency proofs are satisfied when P_0 and T are PA and P is the extension by definitions of PA just defined, it follows that *if* $\vdash_{\mathrm{PA}} \mathrm{Con}_{\mathrm{PA}}$, *then* PA *is inconsistent*.

In this case the arithmetical completeness theorem yields an interpretation of PA in an extension by definitions of $\mathrm{PA}[\mathrm{Con}_{\mathrm{PA}}]$.

6.2 Describing ZF. The language of ZF has a single nonlogical symbol $\in$ which is a binary predicate symbol. The axioms of ZF are

(i) $\forall z(z \in x \leftrightarrow z \in y) \to x = y$ (*extensionality axiom*);
(ii) $\exists y(y \in x) \to \exists y(y \in x \wedge \neg\exists z(z \in x \wedge z \in y))$ (*regularity axiom*);
(iii) $\exists w \forall y(\forall z(z \in y \to z \in x) \to y \in w)$ (*power set axiom*);
(iv) $\exists x(\exists y(y \in x \wedge \forall z \neg(z \in y)) \wedge \forall y(y \in x \to \exists z(z \in x \wedge \forall w(w \in z \leftrightarrow w \in y \vee w = y))))$ (*infinity axiom*);

(v) $\exists \mathbf{z} \forall \mathbf{x}(\mathbf{x} \in \mathbf{z} \leftrightarrow \mathbf{x} \in \mathbf{y} \wedge \mathbf{A})$ where $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ are distinct and $\mathbf{y}$ and $\mathbf{z}$ do not occur in $\mathbf{A}$ (*subset axioms*);

(vi) $\forall \mathbf{x} \exists \mathbf{z} \forall \mathbf{y}(\mathbf{A} \leftrightarrow \mathbf{y} \in \mathbf{z}) \rightarrow \exists \mathbf{z} \forall \mathbf{y}(\exists \mathbf{x}(\mathbf{x} \in \mathbf{w} \wedge \mathbf{A}) \rightarrow \mathbf{y} \in \mathbf{z})$ where $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$, and $\mathbf{w}$ are distinct and $\mathbf{z}$ and $\mathbf{w}$ do not occur in $\mathbf{A}$ (*replacement axioms*).

In elementary developments of ZF, one proves that there is an extension by definitions ZF$'$ obtained from ZF by the adjunction of the five symbols $\dot{0}$, S, Nn, $\oplus$, and $\otimes$ such that the following interpretation I is an interpretation of PA in ZF$'$: U_I is Nn, $=_I$ is $=$, $\dot{0}_I$ is $\dot{0}$, S_I is S, $+_I$ is $\oplus$, $\cdot_I$ is $\otimes$, and $<_I$ is $\in$. We define an arithmetical language $\mathfrak{L}$ as follows: $\mathrm{Vr}\, x = x + \mathrm{S}\dot{9}$, $\mathrm{Func}\, xy \leftrightarrow (y = \dot{0} \wedge x = \dot{5}) \vee (y = \dot{1} \wedge x = \dot{6}) \vee (y = \dot{2} \wedge x = \dot{7} \vee x = \dot{8})$, $\mathrm{Pred}\, xy \leftrightarrow (y = \dot{1} \wedge x = \dot{0}) \vee (y = \dot{2} \wedge x = \dot{4} \vee x = \dot{9})$, $\dot{\vee} = \dot{1}$, $\dot{\neg} = \dot{2}$, $\dot{\exists} = \dot{3}$, $\dot{=} = \dot{4}$, $\ddot{0} = \dot{5}$, $\dot{\mathrm{S}} = \dot{6}$. Set $\sigma_{\mathrm{ZF}}(\mathrm{Nn}) = 0$, $\sigma_{\mathrm{ZF}}(\vee) = 1$, $\sigma_{\mathrm{ZF}}(\neg) = 2$, $\sigma_{\mathrm{ZF}}(\exists) = 3$, $\sigma_{\mathrm{ZF}}(=) = 4$, $\sigma_{\mathrm{ZF}}(\dot{0}) = 5$, $\sigma_{\mathrm{ZF}}(\mathrm{S}) = 6$, $\sigma_{\mathrm{ZF}}(\oplus) = 7$, $\sigma_{\mathrm{ZF}}(\otimes) = 8$, $\sigma_{\mathrm{ZF}}(\in) = 9$, and if $\mathbf{x}$ is the $(n+1)$th variable in the alphabetical order, set $\sigma_{\mathrm{ZF}}(\mathbf{x}) = n + 10$. We now arithmetize $L(\mathrm{ZF}')$ from the numerotation σ_{ZF}, so that $\mathfrak{L}$ represents $L(\mathrm{ZF}')$. Let $m_1, \ldots, m_4$ be the expression numbers of the formulae (i)–(iv), and $m_5, \ldots, m_9$ the expression numbers of the defining axioms of $\dot{0}$, S, Nn, $\oplus$, and $\otimes$. We now define an arithmetical theory $\mathfrak{Z}$ in a recursive extension by definitions of PA. Its language is $\mathfrak{L}$ and the other symbols associated with $\mathfrak{Z}$ will be written with the index "ZF". We define $\mathrm{Nlax}_{\mathrm{ZF}}$ by

$$\mathrm{Nlax}_{\mathrm{ZF}}\, x \leftrightarrow x = \dot{m}_1 \vee \cdots \vee x = \dot{m}_9 \vee \mathbf{A} \vee \mathbf{B}$$

where $\mathbf{A}$ is

$$\begin{aligned}&\exists y \exists z \exists w \exists x'(y < x \wedge z < x \wedge w < x \wedge x' < x \\ &\qquad \wedge \mathrm{Vble}\, y \wedge \mathrm{Vble}\, z \wedge \mathrm{Vble}\, w \wedge \mathrm{Fm}\, x' \wedge y \neq z \wedge z \neq w \wedge y \neq w \wedge \neg \mathrm{Occ}\, x'z \wedge \neg \mathrm{Occ}\, x'w \\ &\qquad\qquad \wedge x = \mathrm{Inst}\, w\, \mathrm{Gen}\, y\, \mathrm{Eqv}\langle \dot{\sigma}_{\mathrm{ZF}}(\in), y, w\rangle\, \mathrm{Cnj}\langle \dot{\sigma}_{\mathrm{ZF}}(\in), y, z\rangle x')\end{aligned}$$

and $\mathbf{B}$ is

$$\begin{aligned}&\exists y \exists z \exists w \exists x' \exists y'(y < x \wedge z < x \wedge w < x \wedge x' < x \wedge y' < x \wedge \mathrm{Vble}\, y \wedge \mathrm{Vble}\, z \\ &\qquad \wedge \mathrm{Vble}\, w \wedge \mathrm{Vble}\, x' \wedge \mathrm{Fm}\, y' \wedge y \neq z \wedge y \neq w \wedge y \neq x' \wedge z \neq w \wedge z \neq x' \wedge w \neq x' \\ &\qquad \wedge \neg \mathrm{Occ}\, y'w \wedge \neg \mathrm{Occ}\, y'x' \wedge x = \mathrm{Imp}\, \mathrm{Gen}\, y\, \mathrm{Inst}\, w\, \mathrm{Gen}\, z\, \mathrm{Eqv}\, y'\langle \dot{\sigma}_{\mathrm{ZF}}(\in), z, w\rangle \\ &\qquad\qquad \mathrm{Inst}\, w\, \mathrm{Gen}\, z\, \mathrm{Imp}\, \mathrm{Inst}\, y\, \mathrm{Cnj}\langle \dot{\sigma}_{\mathrm{ZF}}(\in), y, x'\rangle y'\langle \dot{\sigma}_{\mathrm{ZF}}(\in), z, w\rangle).\end{aligned}$$

It is clear that $\mathfrak{Z}$ represents ZF$'$. By the arithmetical completeness theorem, *there is an interpretation of* ZF *in an extension by definitions of* $\mathrm{PA}[\mathrm{Con}_{\mathrm{ZF}}]$.

6.3 Conclusion. We let ZF$'$, I, $\mathfrak{P}$, and $\mathfrak{Z}$ be as in the previous paragraphs. We introduce an arithmetical interpretation $\mathfrak{I}$ of $\mathfrak{L}(\mathfrak{P})$ in $\mathfrak{L}(\mathfrak{Z})$ by $\mathfrak{U}_{\mathfrak{I}} = \dot{\sigma}_{\mathrm{ZF}}(\mathrm{Nn})$ and $(x)_{\mathfrak{I}} = \mathrm{S}x$ so that $\mathfrak{I}$ describes I. Since I is an interpretation of $L(\mathrm{N})$ in ZF$'$, we have $\vdash_{\mathrm{ZF}'} \exists x\, \mathrm{Nn}\, x$, and for every n-ary function symbol $\mathbf{f}$ of $L(\mathrm{N})$, $\vdash_{\mathrm{ZF}'} \mathrm{Nn}\, \mathbf{x}_1 \rightarrow \cdots \rightarrow \mathrm{Nn}\, \mathbf{x}_n \rightarrow \mathrm{Nn}\, \mathbf{f}_I \mathbf{x}_1 \ldots \mathbf{x}_n$, where $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are the n first variables in the reverse alphabetical order. By the definition of $\mathrm{Func}_{\mathrm{PA}}$ and the fact that $\mathfrak{Z}$ describes ZF$'$, we obtain that $\mathfrak{I}$ is an arithmetical interpretation of $\mathfrak{L}(\mathfrak{P})$ in $\mathfrak{Z}$, and hence is an arithmetical interpretation of $\mathfrak{P}_{\mathfrak{I}}$ in $\mathfrak{Z}$. All the defined symbols of PA introduced so far are recursive on $L(\mathrm{N})$, so $\mathfrak{P}_{\mathfrak{I}}$ is certainly recursively enumerable. Moreover, $\mathfrak{P}_{\mathfrak{I}}$ describes (in fact represents) PA by the proposition of §3.5. Thus the hypotheses of the second theorem on consistency proofs are satisfied when P_0 and T are PA, T' is ZF$'$, I is I, $\mathfrak{T}$ is $\mathfrak{P}_{\mathfrak{I}}$, $\mathfrak{T}'$ is $\mathfrak{Z}$, $\mathfrak{I}$ is $\mathfrak{I}$, P is an extension by definitions of PA in which $\mathfrak{P}_{\mathfrak{I}}$, $\mathfrak{Z}$, and $\mathfrak{I}$ are defined, and $\mathbf{C}$ is a translation of $\mathrm{Con}_{\mathrm{ZF}}$ into PA. Instead of translating $\mathrm{Con}_{\mathrm{ZF}}$, it is also possible, by the interpretation extension theorem, to extend I to an interpretation of P in an extension by definitions of ZF. Thus by the second theorem on consistency proofs, *if* $\vdash_{\mathrm{ZF}} (\mathrm{Con}_{\mathrm{ZF}})^I$, *then* ZF *is inconsistent.*

Finally, we mention, without proof and without having seen one, that $(\mathrm{PA}, \mathfrak{P})$ is reflexive and that $(\mathrm{ZF}, \mathfrak{Z})$ is reflexive with respect to I. By the theorem of §4.3, there exist arithmetical theories $\mathfrak{P}'$ and $\mathfrak{Z}'$ in an extension by definitions of PA, representing respectively PA and ZF$'$, the latter with respect to I, and satisfying $\vdash_{\mathrm{PA}} \mathrm{Con}'_{\mathrm{PA}}$ and $\vdash_{\mathrm{ZF}} (\mathrm{Con}'_{\mathrm{ZF}})^I$.

Chapter Six
First-Order Set Theory

§1 The first-order theory ZF

1.1 The language and the axioms. We define a first-order theory called *Zermelo–Fraenkel set theory* and denoted by ZF whose only nonlogical symbol is the binary predicate symbol $\in$ (recall that, according to ch. I §2.5 (viii), we abbreviate $\in\mathbf{ab}$ by $(\mathbf{a} \in \mathbf{b})$, dropping parentheses when possible) and whose nonlogical axioms are the following:

(i) $\forall z(z \in x \leftrightarrow z \in y) \to x = y$ (*extensionality axiom*);

(ii) $\exists y(y \in x) \to \exists y(y \in x \wedge \neg\exists z(z \in x \wedge z \in y))$ (*regularity axiom*);

(iii) $\exists w \forall y(\forall z(z \in y \to z \in x) \to y \in w)$ (*power set axiom*);

(iv) $\exists x(\exists y(y \in x \wedge \forall z \neg(z \in y)) \wedge \forall y(y \in x \to \exists z(z \in x \wedge \forall w(w \in z \leftrightarrow w \in y \vee w = y))))$ (*infinity axiom*);

(v) $\exists\mathbf{z}\forall\mathbf{x}(\mathbf{x} \in \mathbf{z} \leftrightarrow \mathbf{x} \in \mathbf{y} \wedge \mathbf{A})$ where $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ are distinct and $\mathbf{y}$ and $\mathbf{z}$ do not occur in $\mathbf{A}$ (*subset axioms*);

(vi) $\forall\mathbf{x}\exists\mathbf{z}\forall\mathbf{y}(\mathbf{A} \leftrightarrow \mathbf{y} \in \mathbf{z}) \to \exists\mathbf{z}\forall\mathbf{y}(\exists\mathbf{x}(\mathbf{x} \in \mathbf{w} \wedge \mathbf{A}) \to \mathbf{y} \in \mathbf{z})$ where $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$, and $\mathbf{w}$ are distinct and $\mathbf{z}$ and $\mathbf{w}$ do not occur in $\mathbf{A}$ (*replacement axioms*).

The first-order theory obtained by omitting (ii) (resp. (iv)) is denoted by ZF_- (resp. ZF_ω). We abbreviate $\neg(\mathbf{a} \in \mathbf{b})$ to $(\mathbf{a} \notin \mathbf{b})$.

We should note that the first-order theory ZF is often defined with the following axioms in place of (v) and (vi):

(v′) $\exists w \forall x(\exists z(x \in z \wedge z \in y) \to x \in w)$ (*union axiom*);

(vi′) $\forall\mathbf{x}\exists\mathbf{z}\forall\mathbf{y}(\mathbf{A} \leftrightarrow \mathbf{y} = \mathbf{z}) \to \exists\mathbf{z}\forall\mathbf{y}(\mathbf{y} \in \mathbf{z} \leftrightarrow \exists\mathbf{x}(\mathbf{x} \in \mathbf{w} \wedge \mathbf{A}))$ where $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$, and $\mathbf{w}$ are distinct and $\mathbf{z}$ and $\mathbf{w}$ do not occur in $\mathbf{A}$ (*replacement axioms*).

It turns out that the theory defined in this way is equivalent to ZF. We shall only use the axioms (i)–(vi) above.

1.2 Good extensions. An extension T of ZF is called a *good extension* if (v) and (vi) are theorems of T for any formula $\mathbf{A}$ of T. Those theorems are then also called subset axioms and replacement axioms of T. This is certainly the case if T is obtained from ZF by the adjunction of new axioms and new constants (by the substitution rule). Note also that if T' is an extension by definitions of a good extension T, then T' is a good extension as well. For a translation of (v) or (vi) into T is obtained by replacing $\mathbf{A}$ by a translation $\mathbf{A}^*$ of $\mathbf{A}$ into T, so it is a subset or replacement axiom of T.

The individuals whose behaviour ZF is meant to formalize are called *sets*. The formula $\mathbf{a} \in \mathbf{b}$ means that $\mathbf{a}$ is a *member* of $\mathbf{b}$, or an *element* of $\mathbf{b}$, or that $\mathbf{a}$ *belongs* to $\mathbf{b}$. A set $\mathbf{a}$ is then viewed as the collection of all the sets which belong to $\mathbf{a}$.

§2 Definitions in ZF

2.1 Separation 1. In this section we shall give general methods to build extensions by definitions of ZF, as well as introduce such extensions. We let T be a good extension of ZF, $\mathbf{D}$ a formula of T, and $\mathbf{x}, \mathbf{y}_1, \dots, \mathbf{y}_n$, $\mathbf{y}$, and $\mathbf{y}'$ distinct variables such that $\mathbf{x}, \mathbf{y}_1, \dots, \mathbf{y}_n$ include the variables free in $\mathbf{D}$. Denote by $\mathbf{D}'$ the formula $\forall\mathbf{x}(\mathbf{x} \in \mathbf{y} \leftrightarrow \mathbf{D})$.

LEMMA 1. $\vdash_T \mathbf{D}' \to \mathbf{D}'[\mathbf{y}|\mathbf{y}'] \to \mathbf{y} = \mathbf{y}'$.

Proof. Note that $\mathbf{D}'[\mathbf{y}|\mathbf{y}']$ is $\forall\mathbf{x}(\mathbf{x} \in \mathbf{y}' \leftrightarrow \mathbf{D})$. The formula $(\mathbf{x} \in \mathbf{y} \leftrightarrow \mathbf{D}) \wedge (\mathbf{x} \in \mathbf{y}' \leftrightarrow \mathbf{D}) \to (\mathbf{x} \in \mathbf{y} \leftrightarrow \mathbf{x} \in \mathbf{y}')$ is a tautology. By the distribution rule, ch. I §4.1 (vii), and the equivalence theorem, $\vdash_T \forall\mathbf{x}(\mathbf{x} \in \mathbf{y} \leftrightarrow \mathbf{D}) \wedge \forall\mathbf{x}(\mathbf{x} \in \mathbf{y}' \leftrightarrow \mathbf{D}) \to \forall\mathbf{x}(\mathbf{x} \in \mathbf{y} \leftrightarrow \mathbf{x} \in \mathbf{y}')$, that is, $\vdash_T \mathbf{D}' \wedge \mathbf{D}'[\mathbf{y}|\mathbf{y}'] \to \forall\mathbf{x}(\mathbf{x} \in \mathbf{y} \leftrightarrow \mathbf{x} \in \mathbf{y}')$. Finally, we get $\vdash_T \mathbf{D}' \to \mathbf{D}'[\mathbf{y}|\mathbf{y}'] \to \mathbf{y} = \mathbf{y}'$ as a tautological consequence of the latter and a version of the extensionality axiom, as was to be shown. □

LEMMA 2. Suppose that

$$\vdash_T \exists \mathbf{z} \forall \mathbf{x}(\mathbf{D} \to \mathbf{x} \in \mathbf{z}) \tag{1}$$

for some $\mathbf{z}$ distinct from $\mathbf{x}$ and $\mathbf{y}$ and not occurring in $\mathbf{D}$. Then $\vdash_T \exists \mathbf{y}\mathbf{D}'$.

Proof. By the subset axioms, $\vdash_T \exists \mathbf{y}\forall \mathbf{x}(\mathbf{x} \in \mathbf{y} \leftrightarrow \mathbf{x} \in \mathbf{z} \wedge \mathbf{D})$, whence $\vdash_T \forall \mathbf{x}(\mathbf{D} \to \mathbf{x} \in \mathbf{z}) \to \forall \mathbf{x}(\mathbf{D} \to \mathbf{x} \in \mathbf{z}) \wedge \exists \mathbf{y}\forall \mathbf{x}(\mathbf{x} \in \mathbf{y} \leftrightarrow \mathbf{x} \in \mathbf{z} \wedge \mathbf{D})$ by the tautology theorem. From this by the distribution rule, the hypothesis, and the detachment rule, we obtain $\vdash_T \exists \mathbf{z}(\forall \mathbf{x}(\mathbf{D} \to \mathbf{x} \in \mathbf{z}) \wedge \exists \mathbf{y}\forall \mathbf{x}(\mathbf{x} \in \mathbf{y} \leftrightarrow \mathbf{x} \in \mathbf{z} \wedge \mathbf{D}))$, whence $\vdash_T \exists \mathbf{z}\exists \mathbf{y}\forall \mathbf{x}((\mathbf{D} \to \mathbf{x} \in \mathbf{z}) \wedge (\mathbf{x} \in \mathbf{y} \leftrightarrow \mathbf{x} \in \mathbf{z} \wedge \mathbf{D}))$ by prenex operations, ch. I §4.1 (vii), and the equivalence theorem. From this and the tautology $(\mathbf{D} \to \mathbf{x} \in \mathbf{z}) \wedge (\mathbf{x} \in \mathbf{y} \leftrightarrow \mathbf{x} \in \mathbf{z} \wedge \mathbf{D}) \to (\mathbf{x} \in \mathbf{y} \leftrightarrow \mathbf{D})$, we infer $\vdash_T \exists \mathbf{y}\forall \mathbf{x}(\mathbf{x} \in \mathbf{y} \leftrightarrow \mathbf{D})$ by the distribution rule, the $\exists$-introduction rule, and the detachment rule, as was to be shown. □

THEOREM ON SET DEFINITIONS 1. If $\vdash_T \exists \mathbf{z}\forall \mathbf{x}(\mathbf{D} \to \mathbf{x} \in \mathbf{z})$ is verified for some $\mathbf{z}$ distinct from $\mathbf{x}$ and not occurring in $\mathbf{D}$, then existence and uniqueness conditions for $\mathbf{y}$ in $\mathbf{D}'$ are theorems of T.

Proof. Note that if $\mathbf{z}$ is $\mathbf{y}$, we may replace $\mathbf{z}$ by a suitable variable by the variant theorem. Hence, the hypothesis of Lemma 2 is satisfied. □

A formula of the form of (1) with $\mathbf{z}$ distinct from $\mathbf{x}$ and not occurring in $\mathbf{D}$ will be called an *existence condition for the set of all* $\mathbf{x}$ *such that* $\mathbf{D}$.

By the theorem on functional definitions, if some existence condition for the set of all $\mathbf{x}$ such that $\mathbf{D}$ is a theorem of T, the first-order theory obtained from T by the adjunction of a new n-ary function symbol $\mathbf{f}$ and the new axiom $\mathbf{y} = \mathbf{f}\mathbf{y}_1 \dots \mathbf{y}_n \leftrightarrow \forall \mathbf{x}(\mathbf{x} \in \mathbf{y} \leftrightarrow \mathbf{D})$ is an extension by definitions of T. We often abbreviate by $\mathbf{f}\mathbf{y}_1 \dots \mathbf{y}_n = \{\mathbf{x} \mid \mathbf{D}\}$ the defining axiom of $\mathbf{f}$. This is not a strictly legit abbreviation since the variable $\mathbf{y}$ appearing in the defining axiom cannot be recovered from it, but different choice of the variable (as long as it is distinct from $\mathbf{x}$, $\mathbf{y}_1, \dots, \mathbf{y}_n$) yield equivalent theories by the substitution rule. Sometimes we also abbreviate the term $\mathbf{f}\mathbf{a}_1 \dots \mathbf{a}_n$ by $\{\mathbf{x} \mid \mathbf{D}[\mathbf{y}_1, \dots \mathbf{y}_n | \mathbf{a}_1, \dots, \mathbf{a}_n]\}$ if the $\mathbf{a}_i$ are substitutible for the $\mathbf{y}_i$. Again, this abbreviation is not really legit, for it does not contain enough information to recover what it abbreviates. However, this will not lead to any confusion; for if the symbols $\mathbf{f}$ and $\mathbf{f}'$ happen to be defined so that $\mathbf{f}\mathbf{a}_1 \dots \mathbf{a}_n$ and $\mathbf{f}'\mathbf{b}_1 \dots \mathbf{b}_m$ yield the same abbreviation as above, then it is easily seen using the extensionality axiom that $\vdash_{T'} \mathbf{f}\mathbf{a}_1 \dots \mathbf{a}_n = \mathbf{f}'\mathbf{b}_1 \dots \mathbf{b}_m$; so by the equality theorem, the explicit definition of $\mathbf{f}$ does not matter.

The following criterion is useful to prove existence conditions; it follows at once from the generalization rule and the substitution theorem.

PROPOSITION. If $\mathbf{D}$ as in this paragraph is such that $\vdash_T \mathbf{D} \to \mathbf{x} \in \mathbf{a}$ for some $\mathbf{a}$ (in particular if $\mathbf{D}$ is of the form $\mathbf{x} \in \mathbf{a} \wedge \mathbf{C}$), then an existence condition for the set of all $\mathbf{x}$ such that $\mathbf{D}$ is a theorem of T.

2.2 Separation 2. Let T be a good extension of ZF, $\mathbf{D}$ a formula of T, $\mathbf{x}$, $\mathbf{y}_1, \dots, \mathbf{y}_n$, $\mathbf{y}$, and $\mathbf{w}$ distinct variables such that $\mathbf{x}$, $\mathbf{y}_1, \dots, \mathbf{y}_n$ include the variables free in $\mathbf{D}$, and $\mathbf{y}'_1, \dots, \mathbf{y}'_k$ variables among $\mathbf{y}_1, \dots, \mathbf{y}_n$. Let $\mathbf{f}$ be a $(k+1)$-ary function symbol of T. Denote by $\mathbf{D}'$ the formula $\exists \mathbf{x}(\mathbf{D} \wedge \mathbf{y} = \mathbf{f}\mathbf{x}\mathbf{y}'_1 \dots \mathbf{y}'_k)$. An existence condition for the set of all $\mathbf{y}$ such that $\mathbf{D}'$ is given by $\exists \mathbf{z}\forall \mathbf{y}(\exists \mathbf{x}(\mathbf{D} \wedge \mathbf{y} = \mathbf{f}\mathbf{x}\mathbf{y}'_1 \dots \mathbf{y}'_k) \to \mathbf{y} \in \mathbf{z})$ for some $\mathbf{z}$ distinct from $\mathbf{y}$ and not occurring in $\mathbf{D}'$.

THEOREM ON SET DEFINITIONS 2. Suppose that some existence condition for the set of all $\mathbf{x}$ such that $\mathbf{D}$ is a theorem of T. Then an existence condition for the set of all $\mathbf{y}$ such that $\mathbf{D}'$ is a theorem of T.

Proof. Choose $\mathbf{z}$ distinct from $\mathbf{y}$, $\mathbf{x}$, $\mathbf{x}_1, \dots, \mathbf{x}_n$ and not occurring in $\mathbf{D}'$. By the substitution theorem, $\vdash_T \forall y(\forall z(z \in y \to z \in x) \to y \in w) \to \forall z(z \in x \to z \in x) \to x \in w$. From this using $\vdash_T \forall z(z \in x \to z \in x)$, the tautology theorem, and the distribution rule, we obtain $\vdash_T \exists w \forall y(\forall z(z \in y \to z \in x) \to y \in w) \to \exists w(x \in w)$. By the power set axiom and the detachment rule, we find $\vdash_T \exists w(x \in w)$, whence $\vdash_T \exists \mathbf{z}\forall \mathbf{y}(\mathbf{y} = \mathbf{f}\mathbf{x}\mathbf{y}'_1 \dots \mathbf{y}'_k \to \mathbf{y} \in \mathbf{z})$ by the version theorem, the replacement theorem, and the equivalence theorem. This last formula is an existence condition for the set of all $\mathbf{y}$ such that $\mathbf{y} = \mathbf{f}\mathbf{x}\mathbf{y}'_1 \dots \mathbf{y}'_k$. Hence by Lemma 2 of §2.1, $\vdash_T \exists \mathbf{z}\forall \mathbf{y}(\mathbf{y} \in \mathbf{z} \leftrightarrow \mathbf{y} = \mathbf{f}\mathbf{x}\mathbf{y}'_1 \dots \mathbf{y}'_k)$. From this by the generalization rule, the replacement axioms, and the detachment rule we get $\vdash_T \exists \mathbf{z}\forall \mathbf{y}(\exists \mathbf{x}(\mathbf{x} \in \mathbf{w} \wedge \mathbf{y} = \mathbf{f}\mathbf{x}\mathbf{y}'_1 \dots \mathbf{y}'_k) \to \mathbf{y} \in \mathbf{z})$. Let U be the first-order theory obtained from T by the adjunction of a new n-ary predicate symbol $\mathbf{g}$ and the new nonlogical axiom $\mathbf{y} = \mathbf{g}\mathbf{y}_1 \dots \mathbf{y}_n \leftrightarrow \forall \mathbf{x}(\mathbf{x} \in \mathbf{y} \leftrightarrow \mathbf{D})$; we know that U is an extension by definitions of T. Then $\vdash_U \exists \mathbf{z}\forall \mathbf{y}(\exists \mathbf{x}(\mathbf{x} \in \mathbf{g}\mathbf{y}_1 \dots \mathbf{y}_n \wedge \mathbf{y} = \mathbf{f}\mathbf{x}\mathbf{y}'_1 \dots \mathbf{y}'_k) \to \mathbf{y} \in \mathbf{z})$ by the substitution rule, and since $\vdash_U \mathbf{x} \in \mathbf{g}\mathbf{y}_1 \dots \mathbf{y}_n \leftrightarrow \mathbf{D}$, we obtain $\vdash_T \exists \mathbf{z}\forall \mathbf{y}(\exists \mathbf{x}(\mathbf{D} \wedge \mathbf{y} = \mathbf{f}\mathbf{x}\mathbf{y}'_1 \dots \mathbf{y}'_k) \to \mathbf{y} \in \mathbf{z})$ by the equivalence theorem. This is a desired existence condition, for $\mathbf{z}$ is distinct from $\mathbf{y}$ and does not occur in $\mathbf{D}'$. □

Thus if some existence condition for the set of all $\mathbf{x}$ such that $\mathbf{D}$ is a theorem of T, then the first-order theory obtained from T by the adjunction of a new n-ary function symbol $\mathbf{g}$ and the new nonlogical axiom $\mathbf{w} = \mathbf{g}\mathbf{y}_1 \dots \mathbf{y}_n \leftrightarrow \forall\mathbf{y}(\mathbf{y} \in \mathbf{w} \leftrightarrow \exists\mathbf{x}(\mathbf{D} \wedge \mathbf{y} = \mathbf{f}\mathbf{x}\mathbf{y}'_1 \dots \mathbf{y}'_k))$ is an extension by definitions of T. The defining axiom for $\mathbf{g}$ is often abbreviated to $\mathbf{g}\mathbf{y}_1 \dots \mathbf{y}_n = \{\mathbf{f}\mathbf{x}\mathbf{y}'_1 \dots \mathbf{y}'_k \mid \mathbf{D}\}$. We sometimes abbreviate $\mathbf{g}\mathbf{y}_1 \dots \mathbf{y}_n$ by $\{\mathbf{f}\mathbf{x}\mathbf{y}'_1 \dots \mathbf{y}'_k \mid \mathbf{D}\}$, and similarly if the variables $\mathbf{y}_i$ are replaced by terms $\mathbf{a}_i$ substitutible for the $\mathbf{y}_i$ in $\mathbf{D}$. Even though those abbreviations are not legit, they are harmless for the same reason as in §2.1.

2.3 Defined symbols. Here is a list of defining axioms for new symbols.

(i) $x = \dot{0} \leftrightarrow \forall y(y \in x \leftrightarrow y \neq y)$;

(ii) $x \subseteq y \leftrightarrow \forall z(z \in x \rightarrow z \in y)$;

(iii) $x = \mathrm{P}y \leftrightarrow \forall w(w \in x \leftrightarrow \forall z(z \in w \rightarrow z \in y))$;

(iv) $z = \{\}_2 xy \leftrightarrow \forall w(w \in z \leftrightarrow w = x \vee w = y)$;

(v) $\{\}_1 x = \{\}_2 xx$;

(vi) $\langle\rangle_2 xy = \{\}_2\{\}_1 x\{\}_2 xy$;

(vii) $\langle\rangle_1 x = x$;

(viii) for $n \geq 3$, $\langle\rangle_n x_1 \dots x_n = \langle\rangle_2 x_1 \langle\rangle_{n-1} x_2 \dots x_n$;

(ix) $z = \mathrm{Un}\, w \leftrightarrow \forall y(y \in z \leftrightarrow \exists x(x \in w \wedge y \in x))$;

(x) for $n \geq 3$, $\{\}_n x_1 \dots x_n = \mathrm{Un}\{\}_2\{\}_1 x_1 \{\}_{n-1} x_2 \dots x_n$;

(xi) $\cup xy = \mathrm{Un}\{\}_2 xy$;

(xii) $z = \cap xy \leftrightarrow \forall w(w \in z \leftrightarrow w \in x \wedge w \in y)$;

(xiii) $z = -xy \leftrightarrow \forall w(w \in z \leftrightarrow w \in x \wedge \neg w \in y)$;

(xiv) $\mathrm{S}x = \cup x\{\}_1 x$;

(xv) $\dot{1} = \mathrm{S}\dot{0}$, $\dot{2} = \mathrm{S}\dot{1}$, $\dot{3} = \mathrm{S}\dot{2}$, $\dot{4} = \mathrm{S}\dot{3}$, $\dot{5} = \mathrm{S}\dot{4}$, $\dot{6} = \mathrm{S}\dot{5}$, $\dot{7} = \mathrm{S}\dot{6}$, $\dot{8} = \mathrm{S}\dot{7}$, $\dot{9} = \mathrm{S}\dot{8}$;

(xvi) $z = \times_2 xy \leftrightarrow \forall w(w \in z \leftrightarrow \exists x'\exists y'(w = \langle\rangle_2 x'y' \wedge x' \in x \wedge y' \in y))$;

(xvii) for $n \geq 3$, $\times_n x_1 \dots x_n = \times_2 x_1 \times_{n-1} x_2 \dots x_n$;

(xviii) for $n \geq 1$ and for $1 \leq i \leq n$, $y = \pi^n_i x \leftrightarrow (\exists x_1 \dots \exists x_n(x = \langle\rangle_n x_1 \dots x_n \wedge y = x_i)) \vee (\neg\exists x_1 \dots \exists x_n(x = \langle\rangle_n x_1 \dots x_n) \wedge y = \dot{0})$;

(i′) $\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$ abbreviates $\{\}_n \mathbf{a}_1 \dots \mathbf{a}_n$;

(ii′) $\langle\mathbf{a}_1, \dots, \mathbf{a}_n\rangle$ abbreviates $\langle\rangle_n \mathbf{a}_1 \dots \mathbf{a}_n$;

(iii′) $(\mathbf{a}_1 \times \dots \times \mathbf{a}_n)$ abbreviates $\times_n \mathbf{a}_1 \dots \mathbf{a}_n$;

(iv′) $(\mathbf{a} \cup \mathbf{b})$, $(\mathbf{a} \cap \mathbf{b})$, $(\mathbf{a} - \mathbf{b})$ abbreviate respectively $\cup\mathbf{ab}$, $\cap\mathbf{ab}$, $-\mathbf{ab}$.

We now prove that all of them are defined symbols. For (v)–(viii), (x)–(xi), (xiv)–(xv), and (xvii), this follows from the Proposition 1 of ch. II §2.2; for (xii)–(xiii), this follows from the proposition of §2.1 and the first theorem on set definitions. We settle the remaining cases, namely (i), (iii), (iv), (ix), (xvi), and (xviii).

As a tautological consequence of the identity axiom $y = y$ and by the generalization rule and the substitution theorem, we have $\vdash_{\mathrm{ZF}} \exists x \forall y(y \neq y \rightarrow y \in x)$. This is an existence condition for the set of all y such that $y \neq y$. So (i) is a valid definition by the first theorem on set definitions.

An existence condition for the set of all w such that $\forall z(z \in w \rightarrow z \in y)$ is just a version of the power set axiom, and hence is a theorem of ZF. This proves that (iii) is a valid definition.

For (iv), we must prove $\vdash_{\mathrm{ZF}} \exists x' \forall w(w = x \vee w = y \rightarrow w \in x')$. We define two new function symbols $\mathbf{f}$ and $\mathbf{g}$ by $w = \mathbf{f}zxy \leftrightarrow \mathbf{A}$, where $\mathbf{A}$ is $(z = \dot{0} \wedge x = w) \vee (\neg z = \dot{0} \wedge y = w)$, and $z = \mathbf{g}xy \leftrightarrow \forall x'(x' \in z \leftrightarrow \exists w(w \in \mathrm{PP}\dot{0} \wedge x' = \mathbf{f}wxy))$. It is easy to derive uniqueness and existence conditions for w in $\mathbf{A}$ using the proposition 2 of ch. II §2.2, so by the second theorem on set definitions $\mathbf{f}$ and $\mathbf{g}$ are defined symbols. By the substitution axioms, it remains to prove that $\vdash_{\mathrm{ZF}} \forall w(w = x \vee w = y \rightarrow w \in \mathbf{g}xy)$, which is inferrable from $x \in \mathbf{g}xy$ and $y \in \mathbf{g}xy$. Now $\vdash_{\mathrm{ZF}} \dot{0} \in \mathrm{PP}\dot{0} \wedge x = \mathbf{f}\dot{0}xy$, so $\vdash_{\mathrm{ZF}} \exists w(w \in \mathrm{PP}\dot{0} \wedge x = \mathbf{f}wxy)$ whence $\vdash_{\mathrm{ZF}} x \in \mathbf{g}xy$. Using $\vdash_{\mathrm{ZF}} \mathrm{P}\dot{0} \in \mathrm{PP}\dot{0}$ and $\vdash_{\mathrm{ZF}} \neg\mathrm{P}\dot{0} = \dot{0}$, we find similarly $\vdash_{\mathrm{ZF}} y \in \mathbf{g}xy$.

The formula $\forall x \exists z \forall y(y \in x \leftrightarrow y \in z) \rightarrow \exists z \forall y(\exists x(x \in w \wedge y \in x) \rightarrow y \in z)$ is a replacement axiom of ZF. Since $\forall x \exists z \forall y(y \in x \leftrightarrow y \in z)$ is inferred from the tautology $y \in x \leftrightarrow y \in x$ by the substitution theorem and the generalization rule, we have $\vdash_{\mathrm{ZF}} \exists z \forall y(\exists x(x \in w \wedge y \in x) \rightarrow y \in z)$ by the detachment rule. This is an existence condition for the set of all y such that $\exists x(x \in w \wedge y \in x)$. Thus the definition (ix) is legit by the first theorem on set definitions.

To prove that (xvi) is a valid definition, it will suffice to prove $\vdash_{\mathrm{ZF}} \exists x' \exists y'(w = \langle x', y' \rangle \wedge x' \in x \wedge y' \in y) \to w \in \mathrm{PP}(x \cup y)$, for then a desired existence condition is obtained by the substitution axioms and the detachment rule. We shall use the following easily derived results:

$$\vdash_{\mathrm{ZF}} x \cup y = y \cup x,$$
$$\vdash_{\mathrm{ZF}} x \in y \to x \in y \cup z, \text{ and}$$
$$\vdash_{\mathrm{ZF}} x \in z \wedge y \in z \to \{x, y\} \in \mathrm{P}z.$$

From these we find $\vdash_{\mathrm{ZF}} x' \in x \wedge y' \in y \to x' \in x \cup y \wedge y' \in x \cup y$, whence $\vdash_{\mathrm{ZF}} x' \in x \wedge y' \in y \to \{x'\} \in \mathrm{P}(x \cup y) \wedge \{x', y'\} \in \mathrm{P}(x \cup y)$, and finally $\vdash_{\mathrm{ZF}} x' \in x \wedge y' \in y \to \langle x', y' \rangle \in \mathrm{PP}(x \cup y)$. The desired result follows from the equality theorem and the $\exists$-introduction rule.

The existence condition for (xviii) is easily derived, and the uniqueness condition will follow from $\vdash_{\mathrm{ZF}} \langle x_1, \dots, x_n \rangle = \langle y_1, \dots, y_n \rangle \to x_1 = y_1 \wedge \dots \wedge x_n = y_n$, for $n \geq 1$. This is obvious when $n = 1$. Suppose that $n = 2$, and form T by the adjunction of four new constants $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}'_1, \mathbf{e}'_2$ and the axiom $\langle\rangle_2 \mathbf{e}_1 \mathbf{e}_2 = \langle\rangle_2 \mathbf{e}'_1 \mathbf{e}'_2$. In view of the deduction theorem, it will suffice to prove $\vdash_T \mathbf{e}_1 = \mathbf{e}'_1 \wedge \mathbf{e}_2 = \mathbf{e}'_2$. From the definition of $\{\}_2$ and the extensionality axiom, we find

$$\vdash_{\mathrm{ZF}} \{\mathbf{a}, \mathbf{b}\} = \{\mathbf{c}, \mathbf{d}\} \leftrightarrow (\mathbf{a} = \mathbf{c} \wedge \mathbf{b} = \mathbf{d}) \vee \mathbf{a} = \mathbf{d} \wedge \mathbf{b} = \mathbf{c}.$$

Using this, the definition of $\langle\rangle_2$, and the tautology theorem, we find

$$\vdash_T \mathbf{e}_1 = \mathbf{e}'_1 \vee (\mathbf{e}_1 = \mathbf{e}'_1 \wedge \mathbf{e}_1 = \mathbf{e}'_2),$$
$$\vdash_T (\mathbf{e}_1 = \mathbf{e}'_1 \wedge \mathbf{e}_2 = \mathbf{e}'_1) \vee (\mathbf{e}_1 = \mathbf{e}'_1 \wedge \mathbf{e}_2 = \mathbf{e}'_2) \vee (\mathbf{e}_1 = \mathbf{e}'_2 \wedge \mathbf{e}_2 = \mathbf{e}'_1), \text{ and}$$
$$\vdash_T (\mathbf{e}'_1 = \mathbf{e}_1 \wedge \mathbf{e}'_2 = \mathbf{e}_1) \vee (\mathbf{e}'_1 = \mathbf{e}_1 \wedge \mathbf{e}'_2 = \mathbf{e}_2) \vee (\mathbf{e}'_1 = \mathbf{e}_2 \wedge \mathbf{e}'_2 = \mathbf{e}_1),$$

whence respectively $\vdash_T \mathbf{e}_1 = \mathbf{e}'_1$, $\vdash_T \mathbf{e}_2 \neq \mathbf{e}'_2 \to \mathbf{e}_2 = \mathbf{e}'_1$, and $\vdash_T \mathbf{e}'_2 \neq \mathbf{e}_2 \to \mathbf{e}'_2 = \mathbf{e}_1$ by the tautology theorem. From these, using the equality axioms, the symmetry theorem, and the tautology theorem, we obtain $\vdash_T \mathbf{e}_2 \neq \mathbf{e}'_2 \to \mathbf{e}_2 = \mathbf{e}'_2$, whence $\vdash_T \mathbf{e}_2 = \mathbf{e}'_2$ by the tautology theorem. The general case follows easily by induction using the definition of $\langle\rangle_n$.

We give some English terminology which will be used in the informal exposition. A set is *empty* if it is equal to $\dot{0}$. We say that $\mathbf{a}$ is a *subset* of $\mathbf{b}$, or is *included* in $\mathbf{b}$, if $\mathbf{a} \subseteq \mathbf{b}$. A set of the form $\langle\rangle_n \mathbf{a}_1 \dots \mathbf{a}_n$ is called an *n-tuple*, or an *ordered pair* if $n = 2$.

2.4 Separation 3. Let T be a good extension of ZF, $\mathbf{D}$ a formula of T, $\mathbf{x}_1, \dots, \mathbf{x}_m, \mathbf{y}_1, \dots, \mathbf{y}_n, \mathbf{y}$, and $\mathbf{w}$ distinct variables such that $\mathbf{x}_1, \dots, \mathbf{x}_m, \mathbf{y}_1, \dots, \mathbf{y}_n$ include the variables free in $\mathbf{D}$, and $\mathbf{y}'_1, \dots, \mathbf{y}'_k$ variables among $\mathbf{y}_1, \dots, \mathbf{y}_n$. Let $\mathbf{f}$ be a $(k+m)$-ary function symbol of T. If $\mathbf{z}_1, \dots, \mathbf{z}_m$ are m variables distinct from $\mathbf{x}_1, \dots, \mathbf{x}_m$ and not occurring in $\mathbf{D}$, the formula

$$\exists \mathbf{z}_1 \dots \exists \mathbf{z}_m \forall \mathbf{x}_1 \dots \forall \mathbf{x}_m (\mathbf{D} \to \mathbf{x}_1 \in \mathbf{z}_1 \wedge \dots \wedge \mathbf{x}_m \in \mathbf{z}_m)$$

is called a *joint existence condition for the set of all* $\mathbf{x}_1, \dots, \mathbf{x}_m$ *such that* $\mathbf{D}$.

Theorem on set definitions 3. Suppose that some joint existence condition for the set of all $\mathbf{x}_1, \dots, \mathbf{x}_m$ such that $\mathbf{D}$ is a theorem of T. Then an existence condition for the set of all $\mathbf{y}$ such that $\exists \mathbf{x}_1 \dots \exists \mathbf{x}_m (\mathbf{D} \wedge \mathbf{y} = \mathbf{f}\mathbf{x}_1 \dots \mathbf{x}_m \mathbf{y}'_1 \dots \mathbf{y}'_k)$ is a theorem of T.

Proof. By the first theorem on set definitions, it will suffice to prove

$$\vdash_T \exists \mathbf{z} \forall \mathbf{y} (\exists \mathbf{x}_1 \dots \exists \mathbf{x}_m (\mathbf{D} \wedge \mathbf{y} = \mathbf{f}\mathbf{x}_1 \dots \mathbf{x}_m \mathbf{y}'_1 \dots \mathbf{y}'_k) \to \mathbf{y} \in \mathbf{z}) \tag{2}$$

for a suitable variable $\mathbf{z}$. We may suppose that sufficiently many symbols are defined in T, in particular symbols $\mathbf{f}'$ and $\mathbf{h}$ defined by $\mathbf{f}' x y_1 \dots y_n = \mathbf{f} \pi^m_1 x \dots \pi^m_m x y_1 \dots y_n$ and $y = \mathbf{h} x_1 \dots x_m y_1 \dots y_k \leftrightarrow \forall z(z \in y \leftrightarrow \exists w(w \in (x_1 \times \dots \times x_m) \wedge z = \mathbf{f}' w y_1 \dots y_n))$, the latter being valid by the second theorem on set definitions and the Proposition of §2.1. We then derive $\vdash_T \mathbf{f} x_1 \dots x_m y_1 \dots y_n = \mathbf{f}' \langle x_1, \dots, x_m \rangle y_1 \dots y_n$. Let T' be obtained from T by the adjunction of m new constants $\mathbf{e}_1, \dots, \mathbf{e}_m$, and let $\mathbf{A}$ be $\forall \mathbf{x}_1 \dots \forall \mathbf{x}_m (\mathbf{D} \to \mathbf{x}_1 \in \mathbf{e}_1 \wedge \dots \wedge \mathbf{x}_m \in \mathbf{e}_m)$. Using the definitions, we find $\vdash_{T'[\mathbf{A}]} \exists \mathbf{x}_1 \dots \exists \mathbf{x}_m (\mathbf{D} \wedge \mathbf{y} = \mathbf{f}\mathbf{x}_1 \dots \mathbf{x}_m \mathbf{y}'_1 \dots \mathbf{y}'_k) \to \mathbf{y} \in \mathbf{h}\mathbf{e}_1 \dots \mathbf{e}_m \mathbf{y}'_1 \dots \mathbf{y}'_k$, whence $\vdash_{T'[\mathbf{A}]} \exists\mathbf{z}\forall\mathbf{y}(\exists\mathbf{x}_1 \ldots \exists\mathbf{x}_m(\mathbf{D} \wedge \mathbf{y} = \mathbf{f}\mathbf{x}_1 \ldots \mathbf{x}_m\mathbf{y}'_1 \ldots \mathbf{y}'_k) \to \mathbf{y} \in \mathbf{z})$ by the generalization rule and the substitution axioms. By the deduction theorem,

$$\vdash_T \forall\mathbf{x}_1 \ldots \forall\mathbf{x}_m(\mathbf{D} \to \mathbf{x}_1 \in \mathbf{z}_1 \wedge \cdots \wedge \mathbf{x}_m \in \mathbf{z}_m) \to \exists\mathbf{z}\forall\mathbf{y}(\exists\mathbf{x}_1 \ldots \exists\mathbf{x}_m(\mathbf{D} \wedge \mathbf{y} = \mathbf{f}\mathbf{x}_1 \ldots \mathbf{x}_m\mathbf{y}'_1 \ldots \mathbf{y}'_k) \to \mathbf{y} \in \mathbf{z}),$$

whence (2) by the $\exists$-introduction rule and the joint existence condition. □

Thus if some joint existence condition for the set of all $\mathbf{x}_1, \ldots, \mathbf{x}_m$ such that $\mathbf{D}$ is a theorem of T, then the first-order theory obtained from T by the adjunction of a new n-ary function symbol $\mathbf{g}$ and the new nonlogical axiom $\mathbf{w} = \mathbf{g}\mathbf{y}_1 \ldots \mathbf{y}_n \leftrightarrow \forall\mathbf{y}(\mathbf{y} \in \mathbf{w} \leftrightarrow \exists\mathbf{x}_1 \ldots \exists\mathbf{x}_m(\mathbf{D} \wedge \mathbf{y} = \mathbf{f}\mathbf{x}_1 \ldots \mathbf{x}_m\mathbf{y}'_1 \ldots \mathbf{y}'_k))$ is an extension by definitions of T. As for the first two theorems on set definitions (see the remarks following those theorems), we usually write $\mathbf{g}\mathbf{y}_1 \ldots \mathbf{y}_n = \{\mathbf{f}\mathbf{x}_1 \ldots \mathbf{x}_m\mathbf{y}'_1 \ldots \mathbf{y}'_k \mid \mathbf{D}\}$ for the defining axiom of $\mathbf{g}$, and we sometimes use the abbreviation $\{\mathbf{f}\mathbf{x}_1 \ldots \mathbf{x}_m\mathbf{y}'_1 \ldots \mathbf{y}'_k \mid \mathbf{D}\}$ for $\mathbf{g}\mathbf{y}_1 \ldots \mathbf{y}_n$, and similarly if the variables $\mathbf{y}_i$ are replaced by terms $\mathbf{a}_i$ substitutible for them in $\mathbf{D}$.

2.5 More defined symbols. We introduce some more definitions.

(i) $z = \times^1 xy \leftrightarrow \forall w(w \in z \leftrightarrow \exists x'\exists y'\exists z'(y' \in x \wedge \langle x', z'\rangle \in y \wedge w = \langle x', y', z'\rangle))$;
(ii) $z = \times^2 xy \leftrightarrow \forall w(w \in z \leftrightarrow \exists x'\exists y'\exists z'(z' \in x \wedge \langle x', y'\rangle \in y \wedge w = \langle x', y', z'\rangle))$;
(iii) $y = \operatorname{Dom} x \leftrightarrow \forall z(z \in y \leftrightarrow \exists w(w \in x \wedge z = \pi^2_2 w))$;
(iv) $y = \operatorname{Im} x \leftrightarrow \forall z(z \in y \leftrightarrow \exists w(w \in x \wedge z = \pi^2_1 w))$;
(v) $y = \operatorname{Cnv} x \leftrightarrow \forall z(z \in y \leftrightarrow \exists x'\exists y'(\langle y', x'\rangle \in x \wedge z = \langle x', y'\rangle))$;
(vi) $\operatorname{Func} x \leftrightarrow x \subseteq \operatorname{Im} x \times \operatorname{Dom} x \wedge \forall y\forall y'\forall z(\langle y, z\rangle \in x \to \langle y', z\rangle \in x \to y = y')$;
(vii) $\operatorname{IFunc} x \leftrightarrow \operatorname{Func} x \wedge \forall y\forall z\forall z'(\langle y, z\rangle \in x \to \langle y, z'\rangle \in x \to z = z')$;
(viii) $z = {\restriction} xy \leftrightarrow \forall w(w \in z \leftrightarrow \exists x'\exists x''(w = \langle x', x''\rangle \wedge x'' \in y) \wedge w \in x)$;
(ix) $z = {`}xy \leftrightarrow ((\operatorname{Func} x \wedge y \in \operatorname{Dom} x) \wedge \langle z, y\rangle \in x) \vee (\neg(\operatorname{Func} x \wedge \neg y \in \operatorname{Dom} x) \wedge z = \dot{0})$;
(x) $z = \circ xy \leftrightarrow \forall w(w \in z \leftrightarrow \exists x'\exists y'(\exists z'(\langle x', z'\rangle \in x \wedge \langle z', y'\rangle \in y) \wedge w = \langle x', y'\rangle))$;
(i′) $(\mathbf{a} \times^i \mathbf{b})$ abbreviates $\times^i\mathbf{ab}$;
(ii′) $(\mathbf{a} \restriction \mathbf{b})$ abbreviates ${\restriction}\mathbf{ab}$;
(iii′) $(\mathbf{a}`\mathbf{b})$ abbreviates ${`}\mathbf{ab}$;
(iv′) $(\mathbf{a} \circ \mathbf{b})$ abbreviates $\circ\mathbf{ab}$.

We have $\vdash_{\mathrm{ZF}} y' \in x \wedge \langle x', z'\rangle \in y \to x' \in \pi^2_1 y \wedge y' \in x \wedge z' \in \pi^2_2 y$, so (i) is a valid definition by the third theorem on set definitions. The validity of (ii), (iii), (iv), (v), (viii), and (x) is proved in a similar way using the theorems on set definitions. For (ix), it suffices to check that $\vdash_{\mathrm{ZF}} \operatorname{Func} x \wedge y \in \operatorname{Dom} x \to \exists z(\langle z, y\rangle \in x)$ and $\vdash_{\mathrm{ZF}} \operatorname{Func} x \to \langle z, y\rangle \in x \to \langle z', y\rangle \in x \to z = z'$. Both are derived at once from the definitions.

In the informal exposition, $\operatorname{Dom}\mathbf{a}$ is called the *domain* of $\mathbf{a}$, $\operatorname{Im}\mathbf{a}$ the *image* or *range* of $\mathbf{a}$; $\operatorname{Func}\mathbf{a}$ means that $\mathbf{a}$ is a *function*, and $\operatorname{IFunc}\mathbf{a}$ that $\mathbf{a}$ is an *injective function*. A function with domain $\mathbf{a}$ is also called a function *on* $\mathbf{a}$. The set $\mathbf{a}`\mathbf{b}$ is called the *value* of $\mathbf{a}$ at $\mathbf{b}$.

§3 Ordinals and cardinals

3.1 Results on ordinals 1. We define the unary predicate symbols Tr and Ord by $\operatorname{Tr} x \leftrightarrow \forall y\forall z(y \in x \to z \in y \to z \in x)$ and $\operatorname{Ord} x \leftrightarrow \operatorname{Tr} x \wedge \forall y(y \in x \to \operatorname{Tr} y)$. In English, $\operatorname{Tr}\mathbf{a}$ means that $\mathbf{a}$ is *transitive*, and $\operatorname{Ord}\mathbf{a}$ that $\mathbf{a}$ is an *ordinal*. An n-ary function symbol $\mathbf{f}$ of an extension T of ZF is an *ordinal function symbol* if $\vdash_T \operatorname{Ord}\mathbf{f}x_1 \ldots x_n$. We often abbreviate $\in$ by $<$ and $\subseteq$ by $\leq$ when concerned with ordinals, for reasons that will appear shortly. We now derive some theorems involving Ord:

(i) $\vdash_{\mathrm{ZF}} \operatorname{Ord} x \to y \in x \to \operatorname{Ord} y$;
(ii) $\vdash_{\mathrm{ZF}} \operatorname{Ord} x \to \operatorname{Ord} y \to \operatorname{Ord} z \to x \in y \to y \in z \to x \in z$;
(iii) $\vdash_{\mathrm{ZF}} x \notin x$;
(iv) $\vdash_{\mathrm{ZF}} \neg(x \in y \wedge y \in x)$;
(v) if $\mathbf{A}$ is a formula of a good extension T of ZF and if $\mathbf{y}$ is distinct from $\mathbf{x}$ and not free in $\mathbf{A}$, $\vdash_T \exists\mathbf{x}(\operatorname{Ord}\mathbf{x} \wedge \mathbf{A}) \to \exists\mathbf{x}(\operatorname{Ord}\mathbf{x} \wedge \mathbf{A} \wedge \forall\mathbf{y}(\mathbf{y} \in \mathbf{x} \to \neg\mathbf{A}[\mathbf{x}|\mathbf{y}]))$;

(vi) $\vdash_{ZF} \mathrm{Ord}\, x \to \mathrm{Ord}\, y \to x \in y \lor x = y \lor y \in x$;

(vii) $\vdash_{ZF} \mathrm{Ord}\, x \to \mathrm{Ord}\, y \to x \subseteq y \leftrightarrow x \in y \lor x = y$.

Let x be an ordinal, y a member of x. Then by definition y is transitive, so any member of y is a member of x, and hence is transitive. Thus y is an ordinal, and (i) holds; (ii) is obvious from the transitivity of z. By the regularity axiom, $\{x\}$ has a member y such that $\neg\exists z(z \in \{x\} \land z \in y)$; but $y = x$, so $\neg\exists z(z = x \land z \in x)$, and hence $x \notin x$. This proves (iii). By the regularity axiom, either x or y has no member in common with $\{x, y\}$. In particular, either $y \notin x$ or $x \notin y$, which proves (iv).

We derive (v) when $\mathbf{x}$ is x and $\mathbf{y}$ is y. Assume $\mathbf{A}$ for some ordinal x. If $\forall y(y \in x \to \neg\mathbf{A}[x|y])$, then x satisfies the conclusion. Otherwise, $\{y \mid y \in x \land \mathbf{A}[x|y]\}$ has a member, so by the regularity axiom it has a member z which has no member in common with itself. Then $z \in x$, so z is an ordinal by (i), and $\mathbf{A}[x|z]$. Suppose that $y \in z$; then $y \notin \{y \mid y \in x \land \mathbf{A}[x|y]\}$ and by (ii), $y \in x$. So $\neg\mathbf{A}[x|y]$. Thus z is a desired ordinal. In the informal exposition, such an ordinal is called a *minimal ordinal x such that* $\mathbf{A}$ (we shall prove in §3.2 that such an ordinal is unique).

To derive (vi), assume that $\exists x\exists y(\mathrm{Ord}\, x \land \mathrm{Ord}\, y \land \neg(x \in y \lor x = y \lor y \in x))$. By (v), there is a minimal ordinal x such that $\exists y(\mathrm{Ord}\, y \land \neg(x \in y \lor x = y \lor y \in x))$, and a minimal ordinal y such that $\neg(x \in y \lor x = y \lor y \in x)$. Let $z \in y$, and let us prove that $z \in x$. By (i), z is an ordinal, so $x \in z \lor x = z \lor z \in x$ by minimality of y. But if $x \in z \lor x = z$, then $x \in y$ by (ii), which contradicts $\neg(x \in y \lor x = y \lor y \in x)$. So $z \in x$, and hence $y \subseteq x$. Since $x \neq y$, there exists w in $x - y$, and w is an ordinal by (i). By minimality of x, $w \in y \lor w = y \lor y \in w$, and by choice of w, $w = y \lor y \in w$. Hence by (ii), $y \in x$, which contradicts $\exists x\exists y(\mathrm{Ord}\, x \land \mathrm{Ord}\, y \land \neg(x \in y \lor x = y \lor y \in x))$.

The implication from right to left in (vii) is a consequence of the transitivity of ordinals. To prove the converse, let x and y be ordinals, and assume that $\neg(x \in y \lor x = y)$. Then by (vi), $y \in x$. Since $y \notin y$ by (iii), we have $y \in x - y$, so $\neg(x \subseteq y)$.

Principle of Transfinite Induction. Let T be a good extension of ZF and $\mathbf{A}$ a formula of T. If $\mathbf{y}$ is not free in $\mathbf{A}$, then

$$\vdash_T \forall\mathbf{x}(\mathrm{Ord}\,\mathbf{x} \to \forall\mathbf{y}(\mathbf{y} \in \mathbf{x} \to \mathbf{A}[\mathbf{x}|\mathbf{y}]) \to \mathbf{A}) \to \forall\mathbf{x}(\mathrm{Ord}\,\mathbf{x} \to \mathbf{A}).$$

Proof. This follows by the tautology theorem and the equivalence theorem from (v) where $\mathbf{A}$ is replaced by $\neg\mathbf{A}$. ☐

Corollary. Let T be a good extension of ZF, $\mathbf{f}$ an n-ary ordinal function symbol of T, and $\mathbf{A}$ a formula of T. If $\mathbf{y}_1, \dots, \mathbf{y}_n$ are not free in $\mathbf{A}$, then

$$\vdash_T \forall\mathbf{x}_1 \dots \forall\mathbf{x}_n(\forall\mathbf{y}_1 \dots \forall\mathbf{y}_n(\mathbf{f}\mathbf{y}_1 \dots \mathbf{y}_n \in \mathbf{f}\mathbf{x}_1 \dots \mathbf{x}_n \to \mathbf{A}[\mathbf{x}_1, \dots, \mathbf{x}_n|\mathbf{y}_1, \dots, \mathbf{y}_n]) \to \mathbf{A}) \to \forall\mathbf{x}_1 \dots \forall\mathbf{x}_n\mathbf{A}.$$

Proof. Let $\mathbf{z}$ and $\mathbf{w}$ be distinct from $\mathbf{x}_1, \dots, \mathbf{x}_n, \mathbf{y}_1, \dots, \mathbf{y}_n$ and not free in $\mathbf{A}$, and let $\mathbf{B}$ be $\forall\mathbf{x}_1 \dots \forall\mathbf{x}_n(\mathbf{z} = \mathbf{f}\mathbf{x}_1 \dots \mathbf{x}_n \to \mathbf{A})$. Using prenex operations, the replacement theorem, and the fact that $\mathbf{f}$ is an ordinal function symbol, we find

$$\begin{aligned}\vdash_T \forall\mathbf{x}_1 \dots \forall\mathbf{x}_n(\forall\mathbf{y}_1 \dots \forall\mathbf{y}_n(\mathbf{f}\mathbf{y}_1 \dots \mathbf{y}_n \in \mathbf{f}\mathbf{x}_1 \dots \mathbf{x}_n \to \mathbf{A}[\mathbf{x}_1, \dots, \mathbf{x}_n|\mathbf{y}_1, \dots, \mathbf{y}_n]) \to \mathbf{A})\\ \leftrightarrow \forall\mathbf{z}(\mathrm{Ord}\,\mathbf{z} \to \forall\mathbf{w}(\mathbf{w} \in \mathbf{z} \to \mathbf{B}[\mathbf{z}|\mathbf{w}]) \to \mathbf{B}),\end{aligned}$$

and similarly $\vdash_T \forall\mathbf{z}(\mathrm{Ord}\,\mathbf{z} \to \mathbf{B}) \to \forall\mathbf{x}_1 \dots \forall\mathbf{x}_n\mathbf{A}$. By the principle of transfinite induction, $\vdash_T \forall\mathbf{z}(\mathrm{Ord}\,\mathbf{z} \to \forall\mathbf{w}(\mathbf{w} \in \mathbf{z} \to \mathbf{B}[\mathbf{z}|\mathbf{w}]) \to \mathbf{B}) \to \forall\mathbf{z}(\mathrm{Ord}\,\mathbf{z} \to \mathbf{B})$. Combining those three formulae with the tautology theorem, we obtain the desired result. ☐

Informally, the principle of transfinite induction means that in order to prove that $\mathbf{A}$ holds for any ordinal $\mathbf{x}$, it suffices to prove it under the hypothesis that $\forall\mathbf{y}(\mathbf{y} \in \mathbf{x} \to \mathbf{A}[\mathbf{x}|\mathbf{y}])$. Such a proof is called a *proof by transfinite induction* on $\mathbf{x}$ (or by transfinite induction on $\mathbf{f}\mathbf{x}_1 \dots \mathbf{x}_n$ if we use intead the corollary). The formula $\forall\mathbf{y}(\mathbf{y} \in \mathbf{x} \to \mathbf{A}[\mathbf{x}|\mathbf{y}])$ is called the *induction hypothesis*. Even though there is a clash in terminology, the reader should realize that proofs by transfinite induction are completely unrelated to usual proofs by induction that we have already used and shall yet use. The context will always prevent any possible confusion.

Remark. The derivations of the above theorems show our first use of the regularity axiom. In general, many theorems that do not involve ordinals are derivable in ZF_-. We did not insist on this earlier, because the regularity axiom is true for the meaning we have in mind for sets. However, omitting the regularity axiom yields the theory ZF_- which may be viewed as formalizing a more complex notion of set. Since ZF_- is weaker than ZF but sufficiently strong for most developments of set theory, it is common to investigate first ZF_- and mention the regularity axiom as a possible new axiom that puts a restriction on the kind of sets we wish to study. In that setting, it is possible to refine the definition of Ord so that the above theorems remain derivable in ZF_-. To do this first define Reg in ZF_- by $\mathrm{Reg}\, x \leftrightarrow \forall y(y \subseteq x \to \exists z(z \in y) \to \exists z(z \in y \wedge \neg\exists w(w \in y \wedge w \in z)))$; note that the closure of the regularity axiom is equivalent in ZF_- to $\forall x\, \mathrm{Reg}\, x$. Then the definition of Ord is $\mathrm{Ord}\, x \leftrightarrow \mathrm{Tr}\, x \wedge \forall y(y \in x \to \mathrm{Tr}\, y) \wedge \mathrm{Reg}\, x$. As can be seen, the regularity axiom is just "included" in the definition, so that $\vdash_{\mathrm{ZF}_-} \mathrm{Ord}\, x \to x \notin x$ and $\vdash_{\mathrm{ZF}_-} \mathrm{Ord}\, x \to \mathrm{Ord}\, y \to \neg(x \in y \wedge y \in x)$. These theorems can then replace (iii) and (iv) in most of our future applications.

3.2. Let $\mathbf{D}$ be a formula of some good extension T of ZF, and let $\mathbf{x}_1, \dots, \mathbf{x}_n, \mathbf{y}, \mathbf{y}'$, and $\mathbf{z}$ be distinct variables such that $\mathbf{x}_1, \dots, \mathbf{x}_n$, and $\mathbf{y}$ include the variables free in $\mathbf{D}$ and such that $\mathbf{z}$ is substitutible for $\mathbf{y}$ in $\mathbf{D}$. Denote by $\mathbf{D}'$ the formula $\mathrm{Ord}\,\mathbf{y} \wedge \mathbf{D} \wedge \forall\mathbf{z}(\mathbf{z} \in \mathbf{y} \to \neg\mathbf{D}[\mathbf{y}|\mathbf{z}])$. This formula means that $\mathbf{y}$ is the *first ordinal* such that $\mathbf{D}$.

Theorem on ordinal definitions. Suppose that $\vdash_T \exists\mathbf{y}(\mathrm{Ord}\,\mathbf{y} \wedge \mathbf{D})$. Then existence and uniqueness conditions for $\mathbf{y}$ in $\mathbf{D}'$ are theorems of T.

Proof. Clearly $\vdash_T \mathbf{y} \in \mathbf{y}' \to \mathbf{D} \to \neg\mathbf{D}'[\mathbf{y}|\mathbf{y}']$ by the tautology theorem and the substitution theorem, and symmetrically $\vdash_T \mathbf{y}' \in \mathbf{y} \to \mathbf{D}[\mathbf{y}|\mathbf{y}'] \to \neg\mathbf{D}'$. By (vi) of §3.1, $\vdash_T \mathrm{Ord}\,\mathbf{y} \to \mathrm{Ord}\,\mathbf{y}' \to \mathbf{y} \neq \mathbf{y}' \to \mathbf{y} \in \mathbf{y}' \vee \mathbf{y}' \in \mathbf{y}$. From these we obtain $\vdash_T \mathbf{D}' \to \mathbf{D}'[\mathbf{y}|\mathbf{y}'] \to \mathbf{y} = \mathbf{y}'$, which is a uniqueness condition for $\mathbf{y}$ in $\mathbf{D}'$. The existence condition follows directly from the hypothesis and (v) §3.1. □

Thus if $\vdash_T \exists\mathbf{y}(\mathrm{Ord}\,\mathbf{y} \wedge \mathbf{D})$, the first-order theory obtained from T by the adjunction of a new n-ary function symbol $\mathbf{f}$ and the axiom $\mathbf{y} = \mathbf{f}\mathbf{x}_1 \dots \mathbf{x}_n \leftrightarrow \mathrm{Ord}\,\mathbf{y} \wedge \mathbf{D} \wedge \forall\mathbf{z}(\mathbf{z} \in \mathbf{y} \to \neg\mathbf{D}[\mathbf{y}|\mathbf{z}])$ is an extension by definitions of T. We often abbreviate the defining axiom of $\mathbf{f}$ by $\mathbf{f}\mathbf{x}_1 \dots \mathbf{x}_n = \mu\mathbf{y}\mathbf{D}$. As always, some information is lost in this abbreviation (namely $\mathbf{z}$), but different choices yield equivalent theories. We may even use $\mu\mathbf{y}\mathbf{D}$ as abbreviating a term; as before, this abuse is possible by the equality theorem. Note that we may always define $\mathbf{f}$ by $\mathbf{y} = \mathbf{f}\mathbf{x}_1 \dots \mathbf{x}_n \leftrightarrow (\mathrm{Ord}\,\mathbf{y} \wedge \mathbf{D} \wedge \forall\mathbf{z}(\mathbf{z} \in \mathbf{y} \to \neg\mathbf{D}[\mathbf{y}|\mathbf{z}])) \vee (\neg\exists\mathbf{y}(\mathrm{Ord}\,\mathbf{y} \wedge \mathbf{D}) \wedge \mathbf{y} = \dot{0})$. In both cases, the symbol so defined is an ordinal function symbol.

3.3 Results on ordinals 2. We derive some further results on ordinals.

(i) $\vdash_{\mathrm{ZF}} \mathrm{S}x = \mathrm{S}y \leftrightarrow x = y$;

(ii) $\vdash_{\mathrm{ZF}} \mathrm{Ord}\, x \to \mathrm{Ord}\,\mathrm{S}x$;

(iii) $\vdash_{\mathrm{ZF}} \forall y(y \in x \to \mathrm{Ord}\, y) \to \mathrm{Ord}\,\mathrm{Un}\, x$;

(iv) $\vdash_{\mathrm{ZF}} \exists y(\mathrm{Ord}\, y \wedge \forall z(z \in x \to \mathrm{Ord}\, z \to z \in y))$;

(v) $\vdash_{\mathrm{ZF}} \exists y(\mathrm{Ord}\, y \wedge y \notin x)$;

(vi) for any $(n+1)$-ary function symbol $\mathbf{f}$ in a good extension T of ZF, $\vdash_T \exists z\forall y(\exists x(\mathrm{Ord}\, x \wedge y = \mathbf{f}xy^n) \to y \in z) \to \exists x\exists y(\mathrm{Ord}\, x \wedge \mathrm{Ord}\, y \wedge x < y \wedge \mathbf{f}xy^n = \mathbf{f}yy^n)$.

To prove (i), assume that $x \cup \{x\} = y \cup \{y\}$. Then $(x \in y \vee x = y) \wedge (y \in x \vee x = y)$, but since $\neg(x \in y \wedge y \in x)$, $x \notin x$, and $y \notin y$, we are left with $x = y \wedge x = y$. The other implication is an equality axiom.

Let x be an ordinal. If $y \in \mathrm{S}x$, then $y \in x$ or $y = x$. In both cases, y is transitive and $y \subseteq \mathrm{S}x$, so $\mathrm{S}x$ is an ordinal, proving (ii). If all the member of x are ordinals and if $y \in \mathrm{Un}\, x$, then $y \in z$ for some ordinal $z \in x$. So y is transitive and $y \subseteq z$, whence $y \subseteq \mathrm{Un}\, x$. This proves (iv). Let x be a set and let $y \in x$ be an ordinal. Then $y \subseteq \mathrm{Un}\{z \mid z \in x \wedge \mathrm{Ord}\, z\}$. But $\mathrm{Un}\{z \mid z \in x \wedge \mathrm{Ord}\, z\}$ is an ordinal by (iv), so $y \in \mathrm{S}\,\mathrm{Un}\{z \mid z \in x \wedge \mathrm{Ord}\, z\}$. This proves (v), and (vi) follows from (v) and $y \notin y$.

To prove (vii), we define an $(n+1)$-ary function symbol $\mathbf{g}$ by $z = \mathbf{g}yy^n \leftrightarrow (\mathrm{Ord}\, z \wedge y = \mathbf{f}zy^n \wedge \forall w(w \in z \to y \neq \mathbf{f}wy^n)) \vee (\neg\exists x(\mathrm{Ord}\, x \wedge y = \mathbf{f}xy^n) \wedge z = \dot{0})$, i.e., $\mathbf{g}yy^n$ is the first ordinal x such that $y = \mathbf{f}xy^n$ if such an ordinal exists and $\dot{0}$ otherwise. Assume that there is a set z such that $\forall y(\exists x(\mathrm{Ord}\, x \wedge y = \mathbf{f}xy^n) \to y \in z)$. By (v), there exist an ordinal w such that every ordinal in $\{\mathbf{g}yy^n \mid y \in z\}$ is a member of w. Then since

$\mathbf{f}wy^n \in z$, $\mathbf{a} \in \{\mathbf{g}yy^n \mid y \in z\}$ where $\mathbf{a}$ is $\mathbf{gf}wy^n y^n$. Hence $\mathbf{a} < w$ by choice of w. But $\mathbf{fa}y^n = \mathbf{f}wy^n$ by definition of $\mathbf{g}$, whence (vii).

3.4 Transfinite recursion. In this paragraph we consider how the principle of transfinite induction may be used to define function symbols and predicate symbols. We let T be a good extension of ZF.

Principle of Transfinite Recursion 1. Let $\mathbf{g}$ be an $(n+2)$-ary function symbol of T. There is a defined $(n+1)$-ary function symbol $\mathbf{f}$ such that $\vdash_T \mathrm{Ord}\, x \to \mathbf{f}xy^n = \mathbf{g}\{\langle \mathbf{f}wy^n, w\rangle \mid w < x\}xy^n$.

Proof. Define a new function symbol $\mathbf{h}$ by $y = \mathbf{h}zxy^n \leftrightarrow y = \mathbf{g}(z \restriction x)xy^n$ and a new predicate symbol $\mathbf{r}$ by $\mathbf{r}zxy^n \leftrightarrow \mathrm{Func}\, z \wedge x \subseteq \mathrm{Dom}\, z \wedge \forall w(w \in x \to z`w = \mathbf{h}zwy^n)$. The actual definition of $\mathbf{f}$ is then

$$y = \mathbf{f}xy^n \leftrightarrow (\mathrm{Ord}\, x \wedge \exists z(\mathbf{r}zxy^n \wedge y = \mathbf{h}zxy^n)) \vee (\neg(\mathrm{Ord}\, x \wedge \exists z\mathbf{r}zxy^n) \wedge y = \dot{0}).$$

We have to prove first that this definition is valid. The existence condition for y in the above is obvious. From the definition of $\mathbf{r}$,

$$\vdash_T \mathrm{Tr}\, x \to \mathbf{r}zxy^n \to w \in x \to \mathbf{r}zwy^n. \tag{1}$$

We now derive

$$\mathrm{Ord}\, x \to \mathbf{r}zxy^n \to \mathbf{r}z'xy^n \to \mathbf{h}zxy^n = \mathbf{h}z'xy^n \tag{2}$$

using the principle of transfinite induction on the ordinal x. If $w < x$, then by (1), $\mathbf{r}zwy^n$ and $\mathbf{r}z'wy^n$. By induction hypothesis, $\mathbf{h}zwy^n = \mathbf{h}z'wy^n$; since $z`w = \mathbf{h}zwy^n$ and $z'`w = \mathbf{h}z'wy^n$, we find $z`w = z'`w$. From this it follows that $z \restriction x = z' \restriction x$, whence $\mathbf{h}zxy^n = \mathbf{h}z'xy^n$, which proves (2). This gives the required uniqueness condition.

We now prove that $\mathbf{f}$ has the desired properties. We claim that

$$\vdash_T \mathrm{Ord}\, x \to \exists z(\mathbf{r}zxy^n) \to \mathbf{f}xy^n = \mathbf{g}\{\langle \mathbf{f}wy^n, w\rangle \mid w < x\}xy^n. \tag{3}$$

To prove this, let x be an ordinal, and let z be such that $\mathbf{r}zxy^n$. For all $w < x$, $\mathbf{r}zwy^n$ by (1). So $\mathbf{f}wy^n = \mathbf{h}zwy^n = z`w$. Hence $z \restriction x = \{\langle \mathbf{f}wy^n, w\rangle \mid w < x\}$, so $\mathbf{f}xy^n = \mathbf{h}zxy^n = \mathbf{g}(z \restriction x)xy^n = \mathbf{g}\{\langle \mathbf{f}wy^n, w\rangle \mid w < x\}xy^n$, as claimed.

In view of (3), it will suffice to prove $\vdash_T \mathrm{Ord}\, x \to \exists z(\mathbf{r}zxy^n)$ to conclude the proof. We prove this by transfinite induction on x. Let $\mathbf{a}$ be $\{\langle \mathbf{f}wy^n, w\rangle \mid w < x\}$, and let us verify that $\mathbf{ra}xy^n$. Note that for any $w < x$, $\mathbf{a}`w = \mathbf{f}wy^n$. We have $\mathrm{Func}\, \mathbf{a}$ and $x = \mathrm{Dom}\, \mathbf{a}$, so in particular $x \subseteq \mathrm{Dom}\, \mathbf{a}$. Let $w < x$. Applying the induction hypothesis to w and using (3), we have $\mathbf{f}wy^n = \mathbf{g}\{\langle \mathbf{f}w'y^n, w'\rangle \mid w' < w\}wy^n$. So $\mathbf{a}`w = \mathbf{f}wy^n = \mathbf{g}\{\langle \mathbf{f}w'y^n, w'\rangle \mid w' < w\}wy^n = \mathbf{g}(\mathbf{a} \restriction w)wy^n = \mathbf{ha}wy^n$. This shows that $\mathbf{ra}xy^n$, whence $\exists z(\mathbf{r}zxy^n)$ by the substitution axioms. □

Corollary. Let $\mathbf{g}$ be an $(m+n+1)$-ary function symbol of T, and let $\mathbf{h}$ be an m-ary ordinal function symbol of T such that $\vdash_T \mathrm{Ord}\, x \to \exists z_1 \ldots \exists z_m \forall x_1 \ldots \forall x_m(\mathbf{h}x^m \le x \to x_1 \in z_1 \wedge \cdots \wedge x_m \in z_m)$. Then there is a defined $(m+n)$-ary function symbol $\mathbf{f}$ such that

$$\vdash_T \mathbf{f}x^m y^n = \mathbf{g}\{\lozenge_{m+1}\mathbf{f}w^m y^n w^m \mid \mathbf{h}w^m < \mathbf{h}x^m\}x^m y^n.$$

Proof. We shall first use the principle of transfinite recursion to define an $(n+1)$-ary function symbol $\mathbf{f}'$. For this we must define an $(n+2)$-ary function symbol $\mathbf{g}'$. Its definition is $\mathbf{g}'zxy^n = \{\lozenge_{m+1}\mathbf{g}\, \mathrm{Un}\, \mathrm{Im}\, zw^m y^n w^m \mid \mathbf{h}w^m = x\}$, which is valid by the hypothesis on $\mathbf{h}$, the fact that $\vdash_T \mathbf{h}w^m = x \to \mathbf{h}w^m \le x$, and the third theorem on set definitions. We then let $\mathbf{f}'$ be defined using $\mathbf{g}'$ as in the principle of transfinite recursion, and we define $\mathbf{f}$ by $\mathbf{f}x^m y^n = (\mathbf{f}'\mathbf{h}x^m y^n)`\lozenge_m x^m$. It remains to prove that $\mathbf{f}$ is as claimed. We know that $\vdash_T \mathrm{Ord}\, x \to \mathbf{f}'xy^n = \mathbf{g}'\{\langle \mathbf{f}'wy^n, w\rangle \mid w < x\}xy^n$, so that $\vdash_T \mathbf{f}x^m y^n = (\mathbf{g}'\{\langle \mathbf{f}'wy^n, w\rangle \mid w < \mathbf{h}x^m\}\mathbf{h}x^m y^n)`\lozenge_m x^m$. Thus by definition of $\mathbf{g}'$, $\vdash_T \mathbf{f}x^m y^n = \mathbf{a}$ where $\mathbf{a}$ is $(\{\lozenge_{m+1}\mathbf{g}\, \mathrm{Un}\, \mathrm{Im}\{\langle \mathbf{f}'wy^n, w\rangle \mid w < \mathbf{h}x^m\}w^m y^n w^m \mid \mathbf{h}w^m = \mathbf{h}x^m\})`\lozenge_m x^m$. But clearly $\vdash_T \mathbf{a} = \mathbf{g}\, \mathrm{Un}\{\mathbf{f}'wy^n \mid w < \mathbf{h}x^m\}x^m y^n$, so it will suffice to prove

$$\vdash_T \{\lozenge_{m+1}\mathbf{f}w^m y^n w^m \mid \mathbf{h}w^m < \mathbf{h}x^m\} = \mathrm{Un}\{\mathbf{f}'wy^n \mid w < \mathbf{h}x^m\}.$$

We proceed in English. Since $\mathbf{f}'wy^n$ is a function with domain $\{\lozenge_m w^m \mid \mathbf{h}w^m = w\}$ if w is an ordinal, the function on the right has domain $\{\lozenge_m w^m \mid \mathbf{h}w^m < \mathbf{h}x^m\}$; this is also the domain of the function on the left. Moreover, the two functions have the same values on their domain by definition of $\mathbf{f}$, so they are equal. □

A definition in the form of the corollary is called a definition by transfinite recursion on $\mathbf{h}x^m$. In practice, definitions by transfinite recursion may have various forms, and we now review some of the most common and justify them.

(i) If $\mathbf{g}$ is $(n+1)$-ary, then we may define an $(n+1)$-ary $\mathbf{f}$ such that $\vdash_T \mathrm{Ord}\, x \to \mathbf{f}xy^n = \mathbf{g}\{\mathbf{f}wy^n | w < x\}y^n$.

To prove this, define $\mathbf{g}'$ by $\mathbf{g}'zxy^n = \mathbf{g}\,\mathrm{Im}\, zy^n$, and let $\mathbf{f}$ be obtained using $\mathbf{g}'$ by the principle of transfinite recursion. Then $\vdash_T \mathrm{Ord}\, x \to \mathbf{f}xy^n = \mathbf{g}'\{\langle \mathbf{f}wy^n, w\rangle \,|\, w < x\}xy^n$, whence (i) by definition of $\mathbf{g}'$. In the more general context of the corollary, this becomes:

(ii) If $\mathbf{h}$ is m-ary as in the corollary and if $\mathbf{g}$ is $(n+1)$-ary, then we may define an $(m+n)$-ary $\mathbf{f}$ such that $\vdash_T \mathbf{f}x^m y^n = \mathbf{g}\{\mathbf{f}w^m y^n \,|\, \mathbf{h}w^m < \mathbf{h}x^m\}y^n$.

It suffices to define $\mathbf{g}'$ by $\mathbf{g}'zx^m y^n = \mathbf{g}\,\mathrm{Im}\, zy^n$.

(iii) Let $\mathbf{g}$ and $\mathbf{h}$ be as in the corollary and let $\mathbf{A}$ be a formula with free variables among $x_1, \dots, x_m, w_1, \dots, w_m$ such that $\vdash_T \mathbf{A} \to \mathbf{h}w^m < \mathbf{h}x^m$. Then we may define an $(m+n)$-ary $\mathbf{f}$ such that $\vdash_T \mathbf{f}x^m y^n = \mathbf{g}\{\Diamond_{m+1}\mathbf{f}w^m y^n w^m \,|\, \mathbf{A}\}x^m y^n$.

This is a generalization of the corollary, but it can be seen to be a particular case by setting $\mathbf{g}'zx^m y^n = \mathbf{g}(z \upharpoonright \{\Diamond_m w^m \,|\, \mathbf{A}\})x^m y^n$ and letting $\mathbf{f}$ be defined by $\mathbf{g}'$ as in the corollary. Clearly $\mathbf{f}$ is as required. Another frequently encountered form of the principle of transfinite recursion is the following:

(iv) If $\mathbf{g}$ is $(n+2)$-ary and $\mathbf{h}$ is n-ary, we may define an $(n+1)$-ary $\mathbf{f}$ such that $\vdash_T \mathbf{f}\dot{0}y^n = \mathbf{h}y^n$, $\vdash_T \mathrm{Ord}\, x \to x = \mathrm{S}x' \to \mathbf{f}xy^n = \mathbf{g}\mathbf{f}x'y^n x'y^n$, and $\vdash_T \mathrm{Lim}\, x \to \mathbf{f}xy^n = \mathrm{Un}\{\mathbf{f}x'y^n \,|\, x' < x\}$.

This is done by defining $y = \mathbf{g}'zxy^n \leftrightarrow (x = \dot{0} \wedge y = \mathbf{h}y^n) \vee (\mathrm{Ord}\, x \wedge \exists x'(x = \mathrm{S}x' \wedge y = \mathbf{g}(z`x')x'y^n)) \vee (\mathrm{Lim}\, x \wedge y = \mathrm{Un}\{z`x' \,|\, x' < x\}) \vee (\neg\, \mathrm{Ord}\, x \wedge y = \dot{0})$ and obtaining $\mathbf{f}$ using $\mathbf{g}'$ as in the principle of transfinite recursion.

We now turn to the problem of defining predicate symbols by transfinite induction. This will be a simple application of the principle of transfinite recursion once we note that a predicate symbol may be characterized by a function symbol of same index taking two distinct constants as values.

Principle of Transfinite Recursion 2. Let $\mathbf{q}$ be an $(m+n+1)$-ary predicate symbol of T, and let $\mathbf{h}$ be an m-ary ordinal function symbol of T such that $\vdash_T \mathrm{Ord}\, x \to \exists z_1 \dots \exists z_m \forall x_1 \dots \forall x_m(\mathbf{h}x^m \leq x \to x_1 \in z_1 \wedge \dots \wedge x_m \in z_m)$. Then there is a defined $(m+n)$-ary predicate symbol $\mathbf{p}$ such that $\vdash_T \mathbf{p}x^m y^n \leftrightarrow \mathbf{q}\{\Diamond_m w^m \,|\, \mathbf{h}w^m < \mathbf{h}x^m \wedge \mathbf{p}w^m y^n\}x^m y^n$.

Proof. Define by cases $w = \mathbf{f}'zx^m y^n \leftrightarrow (\mathbf{q}zx^m y^n \wedge w = \dot{1}) \vee (\neg\mathbf{q}zx^m y^n \wedge w = \dot{0})$. Then define an $(m+n+1)$-ary function symbol $\mathbf{g}$ by $\mathbf{g}zx^m y^n = \mathbf{f}'\{\Diamond_m w^m \,|\, \mathbf{h}w^m < \mathbf{h}x^m \wedge z`\Diamond_m w^m = \dot{1}\}x^m y^n$. Using the principle of transfinite recursion, there is a defined $(m+n)$-ary function symbol $\mathbf{f}$ such that $\vdash_T \mathbf{f}x^m y^n = \mathbf{g}\{\Diamond_{m+1}\mathbf{f}w^m y^n w^m \,|\, \mathbf{h}w^m < \mathbf{h}x^m\}x^m y^n$. We then define $\mathbf{p}$ by $\mathbf{p}x^m y^n \leftrightarrow \mathbf{f}x^m y^n = \dot{1}$. Going backwards through the definitions, we see that $\vdash_T \mathbf{p}x^m y^n \leftrightarrow \mathbf{q}\mathbf{a}x^m y^n$ where $\mathbf{a}$ is $\{\Diamond_m w^m \,|\, \mathbf{h}w^m < \mathbf{h}x^m \wedge \mathbf{f}w^m y^n = \dot{1}\}$, but by definition of $\mathbf{p}$, $\vdash_T \mathbf{a} = \{\Diamond_m w^m \,|\, \mathbf{h}w^m < \mathbf{h}x^m \wedge \mathbf{p}w^m y^n\}$. □

As in (iii) above, we can also replace $\mathbf{h}w^m < \mathbf{h}x^m$ by any formula $\mathbf{A}$ with free variables among $x_1, \dots, x_m, w_1, \dots, w_m$ such that $\vdash_T \mathbf{A} \to \mathbf{h}w^m < \mathbf{h}x^m$.

3.5. We define $\mathrm{Max}\, x = \pi_1^2 x \cup \pi_2^2 x$, so that $\vdash_{\mathrm{ZF}} \mathrm{Ord}\, x \to \mathrm{Ord}\, y \to x \subseteq y \leftrightarrow \mathrm{Max}\langle x, y\rangle = y$. Our goal is to define a one to one correspondence O^n between ordinals and n-tuples of ordinals, for $n \geq 2$. We shall do this by listing ordered pairs of ordinals in this way: $\langle \dot{0}, \dot{0}\rangle$, $\langle \dot{0}, \dot{1}\rangle$, $\langle \dot{1}, \dot{0}\rangle$, $\langle \dot{1}, \dot{1}\rangle$, $\langle \dot{0}, \dot{2}\rangle$, $\langle \dot{1}, \dot{2}\rangle$, $\langle \dot{2}, \dot{0}\rangle$, $\langle \dot{2}, \dot{1}\rangle$, $\langle \dot{2}, \dot{2}\rangle$, $\langle \dot{0}, \dot{3}\rangle$, $\langle \dot{1}, \dot{3}\rangle$, etc. We now show how this can be accomplished formally. From $\vdash_{\mathrm{ZF}} \exists y(\mathrm{Ord}\, y \wedge y \notin \mathrm{P}\,\mathrm{Im}\, x)$ we derive $\vdash_{\mathrm{ZF}} \exists y(\mathrm{Ord}\, y \wedge \neg(y \times y \subseteq x))$. Thus we may define $\mathrm{MP}_0 x = \mu y \neg(y \times y \subseteq x)$. We then have $\vdash_{\mathrm{ZF}} \exists y(\mathrm{Ord}\, y \wedge \exists z(z \in \mathrm{MP}_0 x \wedge \langle y, z\rangle \notin x))$, so we may define $\mathrm{MP}_1 x = \mu y \exists z(z \in \mathrm{MP}_0 x \wedge \langle y, z\rangle \notin x)$, and then $\mathrm{MP}_2 x = \mu y(\langle \mathrm{MP}_1 x, y\rangle \notin x)$. Finally, we set $\mathrm{MP}\, x = \langle \mathrm{MP}_1 x, \mathrm{MP}_2 x\rangle$, and we define a unary function symbol O^2 by transfinite recursion so that $\vdash_{\mathrm{ZF}} \mathrm{Ord}\, x \to \mathrm{O}^2 x = \mathrm{MP}\{\mathrm{O}^2 y \,|\, y < x\}$. From the definitions, we derive $\vdash_{\mathrm{ZF}} \exists y \exists z(\mathrm{Ord}\, y \wedge \mathrm{Ord}\, z \wedge \mathrm{MP}\, x = \langle y, z\rangle)$ and $\vdash_{\mathrm{ZF}} \mathrm{MP}\, x \notin x$. Intuitively, $\mathrm{MP}\, x$ is the first ordered pair of ordinals in the list above that does not belong to x. For $n \geq 3$, we define $\mathrm{O}^n x = \langle \pi_1^2 \mathrm{O}^2 x, \mathrm{O}^{n-1}\pi_2^2 \mathrm{O}^2 x\rangle$. To prove that O^n is a one to one correspondence between ordinals and n-tuples of ordinals, it will obviously suffice to treat the case $n = 2$; so we shall prove

(i) $\vdash_{\mathrm{ZF}} \mathrm{Ord}\,x \to \exists y \exists z(\mathrm{Ord}\,y \wedge \mathrm{Ord}\,z \wedge \mathrm{O}^2 x = \langle y, z\rangle)$;

(ii) $\vdash_{\mathrm{ZF}} \mathrm{Ord}\,x \to \mathrm{Ord}\,y \to x \neq y \to \mathrm{O}^2 x \neq \mathrm{O}^2 y$;

(iii) $\vdash_{\mathrm{ZF}} \mathrm{Ord}\,x \to \mathrm{Ord}\,y \to \exists z(\mathrm{Ord}\,z \wedge \mathrm{O}^2 z = \langle x, y\rangle)$.

The first assertion is obvious. Assume that x and y are ordinals such that $x \neq y$, and say $x < y$. Then since $\mathrm{O}^2 y \notin \{\mathrm{O}^2 x \mid x < y\}$, we have $\mathrm{O}^2 x \neq \mathrm{O}^2 y$, which proves (ii). To prove (iii), we shall need

(iv) $\vdash_{\mathrm{ZF}} \mathrm{Max}\,\mathrm{MP}\,x < \mathrm{MP}_0 x$;

(v) $\vdash_{\mathrm{ZF}} \mathrm{Max}\,\mathrm{MP}\,x \times \mathrm{Max}\,\mathrm{MP}\,x \subseteq x$.

Since $\mathrm{MP}_1 x < \mathrm{MP}_0 x$ and $\mathrm{MP}_2 x < \mathrm{MP}_0 x$, we have (iv), whence (v) by definition of MP_0. Let x and y be ordinals. By (ii) above and (vii) of §3.3, we have $\neg\exists z \forall y(\exists x(\mathrm{Ord}\,x \wedge y = \mathrm{O}^2 x)$. So in particular, there exist an ordinal z such that $\mathrm{O}^2 z$ does not belong to $\mathrm{S}\,\mathrm{Max}\langle x, y\rangle \times \mathrm{S}\,\mathrm{Max}\langle x, y\rangle$, and then both x and y belong to $\mathrm{Max}\,\mathrm{O}^2 z$. So by (v), $\langle x, y\rangle \in \{\mathrm{O}^2 w \mid w < z\}$, and this concludes the proof of (iii).

We prove the following additional results on O^2:

(vi) $\vdash_{\mathrm{ZF}} \dot{0} < \mathrm{MP}_1 x \to \langle \dot{0}, \mathrm{Max}\,\mathrm{MP}\,x\rangle \in x$;

(vii) $\vdash_{\mathrm{ZF}} \mathrm{Ord}\,x \to \mathrm{Ord}\,y \to \mathrm{Max}\,\mathrm{O}^2 x < \mathrm{Max}\,\mathrm{O}^2 y \to x < y$;

(viii) $\vdash_{\mathrm{ZF}} \mathrm{Ord}\,x \to \mathrm{Max}\,\mathrm{O}^2 x \leq x$;

(ix) $\vdash_{\mathrm{ZF}} \mathrm{Ord}\,x \to \dot{0} < \pi^2_1 \mathrm{O}^2 x \to \mathrm{Max}\,\mathrm{O}^2 x < x$.

Assume $\dot{0} < \mathrm{MP}_1 x$. Then by definition of MP_1, $\langle \dot{0}, y\rangle \in x$ for any $y < \mathrm{MP}_0 x$, so using (iv) we find $\langle \dot{0}, \mathrm{Max}\,\mathrm{MP}\,x\rangle \in x$, as in (vi). Let x and y be ordinals. If $\mathrm{Max}\,\mathrm{O}^2 x < \mathrm{Max}\,\mathrm{O}^2 y$, then by (v) and the definition of O^2, we have $\mathrm{O}^2 x \in \{\mathrm{O}^2 z | z < y\}$, whence $x < y$ by (ii). This proves (vii). We prove (viii) by transfinite induction on x. We assume that x is an ordinal such that $x < \mathrm{Max}\,\mathrm{O}^2 x$ and derive a contradiction. By (iii), there exists an ordinal y such that $\mathrm{O}^2 y = \langle \dot{0}, x\rangle$, so that $\mathrm{Max}\,\mathrm{O}^2 y = x$ and hence $\mathrm{Max}\,\mathrm{O}^2 y < \mathrm{Max}\,\mathrm{O}^2 x$. Thus by (vii), $y < x$, so $\mathrm{Max}\,\mathrm{O}^2 y \leq y$ by induction hypothesis. Hence $x \leq y$, but this contradicts $y < x$. Finally, assume that x is an ordinal such that $\dot{0} < \pi^2_1 \mathrm{O}^2 x$, i.e., $\dot{0} < \mathrm{MP}_1\{\mathrm{O}^2 y \mid y < x\}$. By (vi), $\langle \dot{0}, \mathrm{Max}\,\mathrm{O}^2 x\rangle \in \{\mathrm{O}^2 y \,|\, y < x\}$, so there exists $y < x$ such that $\mathrm{O}^2 y = \langle \dot{0}, \mathrm{Max}\,\mathrm{O}^2 x\rangle$. So $\mathrm{Max}\,\mathrm{O}^2 x = \mathrm{Max}\,\mathrm{O}^2 y$, but by (viii), $\mathrm{Max}\,\mathrm{O}^2 y \leq y$, so $\mathrm{Max}\,\mathrm{O}^2 x < x$.

Remark. This is the appropriate place to note that all the theorems we have derived in ZF until now did not use the infinity axiom, i.e., we could have replaced ZF by ZF_ω everywhere. We shall use this observation in ch. VII §2.4.

3.6 Infinity. We define the unary predicate symbol Lim by $\mathrm{Lim}\,x \leftrightarrow \mathrm{Ord}\,x \wedge \exists y(y \in x) \wedge \neg\exists y(\mathrm{Ord}\,y \wedge x = \mathrm{S}y)$. In English, we say that x is a *limit ordinal*.

We now prove $\vdash_{\mathrm{ZF}} \exists x\,\mathrm{Lim}\,x$. This will follow from the infinity axiom, the substitution axioms, and the $\exists$-introduction rule if we can prove

$$\vdash_{\mathrm{ZF}} \exists y(y \in x \wedge \forall z \neg(z \in y)) \wedge \forall y(y \in x \to \exists z(z \in x \wedge \forall w(w \in z \leftrightarrow w \in y \vee w = y))) \to \mathrm{Lim}\,\mathbf{a}$$

where $\mathbf{a}$ is $\mathrm{Un}\{z \mid z \in x \wedge \mathrm{Ord}\,z\}$. Assume that x is a set such that

$$\exists y(y \in x \wedge \forall z \neg(z \in y)), \text{ and} \tag{4}$$

$$\forall y(y \in x \to \exists z(z \in x \wedge \forall w(w \in z \leftrightarrow w \in y \vee w = y))). \tag{5}$$

Then $\dot{0} \in x$ by (4), and $\mathrm{S}\dot{0} \in x$ by (5), so $\dot{0} \in \mathbf{a}$ and in particular $\exists y(y \in \mathbf{a})$. From (iv) of §3.3, $\mathbf{a}$ is an ordinal. It remains to prove that $\neg\exists y(\mathrm{Ord}\,y \wedge \mathbf{a} = \mathrm{S}y)$. Assume that $\mathbf{a} = \mathrm{S}y$ for some ordinal y. Then $y \in \mathbf{a}$, so $y \in z$ for some ordinal z in x. Now by (5), $\mathrm{S}z \in x$, so $\mathrm{S}z \subseteq \mathbf{a}$. Since $\mathrm{S}y \in \mathrm{S}z$, $\mathrm{S}y \in \mathbf{a}$, whence $\mathbf{a} \in \mathbf{a}$. This contradicts the assumption that $\exists y(\mathrm{Ord}\,y \wedge \mathbf{a} = \mathrm{S}y)$. Thus, $\mathbf{a}$ is a limit ordinal.

By the theorem on ordinal definitions, we can define a constant ω with the defining axiom $y = \omega \leftrightarrow \mathrm{Lim}\,y \wedge \forall z(z \in y \to \neg\,\mathrm{Lim}\,z)$, or in abbreviated form $\omega = \mu y\,\mathrm{Lim}\,y$. Thus ω is the first limit ordinal.

3.7 The von Neumann hierarchy. We define the symbol Stg by transfinite recursion so that $\vdash_{\mathrm{ZF}} \mathrm{Ord}\,x \to \mathrm{Stg}\,x = \mathrm{P}\,\mathrm{Un}\{\mathrm{Stg}\,y | y < x\}$. Define also Rk by $y = \mathrm{Rk}\,x \leftrightarrow (\mathrm{Ord}\,y \wedge x \in \mathrm{Stg}\,y \wedge \forall z(z < y \to x \notin \mathrm{Stg}\,y)) \vee (\neg\exists z(\mathrm{Ord}\,z \wedge x \in \mathrm{Stg}\,z) \wedge y = \dot{0})$, i.e., $\mathrm{Rk}\,x$ is the first ordinal y such that $x \in \mathrm{Stg}\,y$ if such an ordinal exists. We shall prove that, in fact, such an ordinal always exists, that is,

(i) $\vdash_{ZF} \exists x(\mathrm{Ord}\, x \wedge y \in \mathrm{Stg}\, x)$.

We first prove

(ii) $\vdash_{ZF} \mathrm{Ord}\, x \to \mathrm{Tr}\,\mathrm{Stg}\, x$;

(iii) $\vdash_{ZF} \mathrm{Ord}\, x \to y < x \to \mathrm{Stg}\, y \subseteq \mathrm{Stg}\, x$;

(iv) $\vdash_{ZF} \exists x(\mathrm{Tr}\, x \wedge y \in x)$.

Now (ii) is proved by transfinite induction on x, noting that $\vdash_{ZF} \forall y(y \in x \to \mathrm{Tr}\, y) \to \mathrm{Tr}\,\mathrm{Un}\, x$ and $\vdash_{ZF} \mathrm{Tr}\, x \to \mathrm{Tr}\, \mathrm{P}x$, and (iii) follows at once from (ii) and the definition of Stg. To prove (iv), define $\mathbf{f}$ by transfinite recursion so that $\vdash_{ZF} \mathrm{Ord}\, x \to \mathbf{f}xy = \mathrm{Un}\{\mathrm{Un}\,\mathbf{f}zy \cup \{y\} \mid z < x\}$. We shall prove that $\vdash_{ZF} \mathrm{Tr}\,\mathbf{f}\omega y \wedge y \in \mathbf{f}\omega y$. Now $\vdash_{ZF} \mathrm{Ord}\, x \to \dot{0} < x \to y \in \mathbf{f}xy$ is clear from the definition, so we need only prove that $\mathbf{f}\omega y$ is transitive. Let z be a member of $\mathbf{f}\omega y$. Then z is a member of $\mathrm{Un}\,\mathbf{f}xy \cup \{y\}$ for some $x < \omega$, so $z \in \mathbf{f}\mathrm{S}xy$, and hence $z \subseteq \mathrm{Un}\,\mathbf{f}\mathrm{S}xy$. Since $\mathrm{S}x < \omega$, we get $z \subseteq \mathbf{f}\omega y$ as desired.

We now prove (i). Let y be any set, and using (iv) let z be a transitive set such that $y \in z$. Let $\mathbf{a}$ be $\{w \mid w \in z \wedge \neg\exists x(\mathrm{Ord}\, x \wedge w \in \mathrm{Stg}\, x)\}$. It will suffice to show that $\mathbf{a} = \dot{0}$. Assume the contrary, and using the regularity axiom† let x' be a member of $\mathbf{a}$ which has no member in common with $\mathbf{a}$. By transitivity of z, any member of x' is a member of z, and by choice of x' such a member does not belong to $\mathbf{a}$. Hence if $w \in x'$, then $w \in \mathrm{Stg}\, x$ for some ordinal x; in particular, $w \in \mathrm{Stg}\,\mathrm{Rk}\, w$. By (v) of §3.3, there exists an ordinal y' such that $\{\mathrm{Rk}\, w \mid w \in x'\} \subseteq y'$. If $w \in x'$, then $\mathrm{Rk}\, w < y'$, so by (iii), $w \in \mathrm{Stg}\, y'$. Thus $x' \subseteq \mathrm{Stg}\, y'$, and by definition of Stg we find $x' \in \mathrm{Stg}\,\mathrm{S}y'$, in contradiction with $x' \in \mathbf{a}$.

In English, $\mathrm{Rk}\,\mathbf{a}$ is the *rank* of $\mathbf{a}$. We now derive some more results on Stg and Rk.

(v) $\vdash_{ZF} \mathrm{Ord}\, x \to y \in \mathrm{Stg}\, x \leftrightarrow \mathrm{Rk}\, y \leq x$;

(vi) $\vdash_{ZF} \mathrm{Ord}\, x \to \exists z \forall y(\mathrm{Rk}\, y \leq x \to y \in z)$;

(vii) $\vdash_{ZF} x \in y \to \mathrm{Rk}\, x < \mathrm{Rk}\, y$;

(viii) $\vdash_{ZF} \mathrm{Rk}\, x = \mathrm{Un}\{\mathrm{S}\,\mathrm{Rk}\, y \mid y \in x\}$;

(ix) $\vdash_{ZF} \mathrm{Ord}\, x \to \mathrm{Rk}\, x = x$;

The implication from left to right in (v) is obvious, and the other implication follows at once from (iii). From (v) we get (vi). Let y be any set. We have $y \in \mathrm{Stg}\,\mathrm{Rk}\, y$, and $\mathrm{Stg}\,\mathrm{Rk}\, y = \mathrm{P}\,\mathrm{Un}\{\mathrm{Stg}\, z \mid z < \mathrm{Rk}\, y\}$. Thus if $x \in y$, then $x \in \mathrm{Stg}\, z$ for some $z < \mathrm{Rk}\, y$, and hence by (v), $\mathrm{Rk}\, x \leq z < \mathrm{Rk}\, y$, thereby proving (vii). Let x be a set, and let $\mathbf{a}$ be $\mathrm{Un}\{\mathrm{S}\,\mathrm{Rk}\, y \mid y \in x\}$. From (vii), $\mathrm{S}\,\mathrm{Rk}\, y \leq \mathrm{Rk}\, x$ for any $y \in x$, so $\mathbf{a} \leq \mathrm{Rk}\, x$. Now if $y \in x$, since $\mathrm{Rk}\, y < \mathrm{S}\,\mathrm{Rk}\, y$, we have $\mathrm{Rk}\, y < \mathbf{a}$, whence $y \in \mathrm{Un}\{\mathrm{Stg}\, z \mid z < \mathbf{a}\}$ by (v). From this and the definition of Stg, we find $x \in \mathrm{Stg}\,\mathbf{a}$, whence $\mathrm{Rk}\, x \leq \mathbf{a}$ by (v). This proves (viii). We prove (ix) by transfinite induction on x. By (viii), $\mathrm{Rk}\, x = \mathrm{Un}\{\mathrm{S}\,\mathrm{Rk}\, y \mid y \in x\}$, so by induction hypothesis $\mathrm{Rk}\, x = \mathrm{Un}\{\mathrm{S}y \mid y \in x\} = x$.

3.8 Similarity. We define the binary predicate symbol $\sim$ by $x \sim y \leftrightarrow \exists z(\mathrm{IFunc}\, z \wedge x = \mathrm{Dom}\, z \wedge y = \mathrm{Im}\, z)$, and we prove

(i) $\vdash_{ZF} x \sim x$;

(ii) $\vdash_{ZF} x \sim y \leftrightarrow y \sim x$;

(iii) $\vdash_{ZF} x \sim y \to y \sim z \to x \sim z$;

(iv) $\vdash_{ZF} \mathrm{S}x \sim \mathrm{S}y \leftrightarrow x \sim y$;

(v) $\vdash_{ZF} x \in \omega \to \neg(x \sim \mathrm{S}x)$;

(vi) $\vdash_{ZF} x \sim y \to \mathrm{P}x \sim \mathrm{P}y$.

Let $\mathbf{a}$ be $\{\langle y, y\rangle \mid y \in x\}$. Then $\vdash_{ZF} \mathrm{IFunc}\,\mathbf{a} \wedge x = \mathrm{Dom}\,\mathbf{a} \wedge x = \mathrm{Im}\,\mathbf{a}$, whence (i). Note that

$$\vdash_{ZF} \mathrm{IFunc}\, z \wedge x = \mathrm{Dom}\, z \wedge y = \mathrm{Im}\, z \to \mathrm{IFunc}\,\mathrm{Cnv}\, z \wedge y = \mathrm{Dom}\,\mathrm{Cnv}\, z \wedge x = \mathrm{Im}\,\mathrm{Cnv}\, z,$$

so $\vdash_{ZF} x \sim y \to y \sim x$. The other implication in (ii) is similar; (iii) follows from

$$\vdash_{ZF} \mathrm{IFunc}\, x' \wedge x = \mathrm{Dom}\, x' \wedge y = \mathrm{Im}\, x' \to \mathrm{IFunc}\, y' \wedge y = \mathrm{Dom}\, y' \wedge z = \mathrm{Im}\, y' \\ \to \mathrm{IFunc}(y' \circ x') \wedge x = \mathrm{Dom}(y' \circ x') \wedge z = \mathrm{Im}(y' \circ x').$$

† This is our first real use of the regularity axiom (see the remark in §3.1). It is in fact possible to derive in ZF_- the equivalence of the closure of the regularity axiom and the closure of (i). Thus the regularity axiom can be taken to mean that all sets can be obtained from the empty set by transfinite applications of P and Un.

Assume that z is an injective funtion with domain x and range y. Then $z \cup \{\langle y, x\rangle\}$ is an injective function with domain $\mathrm{S}x$ and range $\mathrm{S}y$. Conversely, let z be an injective function with domain $\mathrm{S}x$ and range $\mathrm{S}y$. If $z`x = y$, then $z \upharpoonright x$ is an injective function on $x$ with range $y$. Otherwise, it is easy to check that $\mathrm{Cnv}(\mathrm{Cnv}(z \upharpoonright x) \upharpoonright y) \cup \{\langle z`x, (\mathrm{Cnv}\, z)`y\rangle\}$ is an injective function with domain x and range y. To prove (v), we use transfinite induction on x. Let x be an ordinal. If $x = \dot{0}$, then $\mathrm{S}\dot{0} = \dot{1}$, and $\dot{0} \neq \dot{1}$ because $\dot{0} \in \dot{1}$. If $\dot{0} < x$ and $x < \omega$, then $x = \mathrm{S}y$ for some ordinal y, because ω is the first limit ordinal. Then $y < x$ and by induction hypothesis $\neg(y \sim \mathrm{S}y)$. Hence by (iv), $\neg(\mathrm{S}y \sim \mathrm{SS}y)$, that is, $\neg(x \sim \mathrm{S}x)$, as was to be shown. Finally, if $\omega \leq x$, the result is tautologically verified. To prove (vi), note that if z is an injective function on x with range y, then $\{\langle\{z`w' \mid w \in w\}, w\rangle \mid w \in \mathrm{P}x\}$ is an injective function on $\mathrm{P}x$ whose range is $\mathrm{P}y$.

3.9 The axiom of choice. We introduce the unary function symbol # with the axiom $y = \#x \leftrightarrow \mathrm{Ord}\, y \wedge x \sim y \wedge \forall z(z \in y \rightarrow \neg(x \sim z))$. Thus $\#x$ is the first ordinal y such that $x \sim y$. It turns out that we are unable to derive an existence condition for y in $\mathrm{Ord}\, y \wedge x \sim y \wedge \forall z(z \in y \rightarrow \neg(x \sim z))$, so we introduce a new axiom.

We define the binary predicate symbol Ch by $\mathrm{Ch}\, xy \leftrightarrow \mathrm{Func}\, x \wedge \mathrm{Dom}\, x = \mathrm{P}y - \{\dot{0}\} \wedge \forall z(z \in \mathrm{Dom}\, x \rightarrow x`z \in z)$; $\mathrm{Ch}\,\mathbf{ab}$ means that $\mathbf{a}$ is a *choice function* on $\mathbf{b}$. The *axiom of choice* is a translation of the formula $\forall x \exists y\, \mathrm{Ch}\, yx$ into ZF. For definiteness, we should now choose explicitly such a translation, but different choices yield equivalent theories by the theorem of ch. II §2.3, so the choice is irrelevant. We denote by ZFC the first-order theory obtained from ZF by the adjunction of the axiom of choice as an axiom. In ZFC, the existence condition for the symbol # follows from this theorem, also known as the *well-ordering theorem*:

ZERMELO'S THEOREM. $\vdash_{\mathrm{ZFC}} \exists y(\mathrm{IFunc}\, y \wedge \mathrm{Ord}\,\mathrm{Dom}\, y \wedge \mathrm{Im}\, y = x)$. More precisely,

$$\vdash_{\mathrm{ZF}} \mathrm{Ch}\, zx \rightarrow \exists y(\mathrm{IFunc}\, y \wedge \mathrm{Ord}\,\mathrm{Dom}\, y \wedge \mathrm{Im}\, y = x \\ \wedge \forall w(w \in \mathrm{Dom}\, y \rightarrow x - \{y`w' \mid w' < w\} \neq \dot{0} \wedge y`w = z`(x - \{y`w' \mid w' < w\})))$$

Proof. By transfinite recursion, we define a new ternary function symbol $\mathbf{f}$ such that $\vdash_{\mathrm{ZF}} \mathrm{Ord}\, w \rightarrow \mathbf{f}wxz = z`(x - \{\mathbf{f}w'xz \mid w' < w\})$. Let x be a set, and let z be a choice function on x. Then for any ordinal w, $\mathbf{f}wxz \in \mathrm{Im}\, z \cup \{\dot{0}\}$. Hence by (vii) of §3.3, there exist ordinals w and w' such that $w' < w$ and $\mathbf{f}wxz = \mathbf{f}w'xz$. Suppose that $x - \{\mathbf{f}w'xz \mid w' < w\}$ is not empty. Then $\mathbf{f}wxz \in x - \{\mathbf{f}w'xz \mid w' < w\}$, and in particular $\mathbf{f}wxy \neq \mathbf{f}w'xy$, contradicting $\mathbf{f}wxz = \mathbf{f}w'xz$. So $x - \{\mathbf{f}w'xz \mid w' < w\}$ must be empty, and hence $x \subseteq \{\mathbf{f}w'xz \mid w' < w\}$. Let y be the first ordinal such that $x \subseteq \{\mathbf{f}w'xz \mid w' < y\}$, and let $\mathbf{a}$ be $\{\langle \mathbf{f}w'xz, w'\rangle \mid w' < y\}$. We now derive that $\mathbf{a}$ is an injective function on y such that $\mathrm{Im}\,\mathbf{a} = x$. Clearly $x \subseteq \mathrm{Im}\,\mathbf{a}$ by the choice of y. If $y' < y$, then $x - \{\mathbf{f}w'xz \mid w' < y'\}$ is not empty by minimality of y, so $\mathbf{f}y'xz \in x - \{\mathbf{f}w'xz \mid w' < y'\}$. This shows that $\mathrm{Im}\,\mathbf{a} \subseteq x$, and hence $\mathrm{Im}\,\mathbf{a} = x$. Finally, $\mathbf{a}$ is injective by minimality of y, for if $w' < w$ and $w < y$, then $\mathbf{f}wxz \neq \mathbf{f}w'xz$, so $\mathbf{a}`w \neq \mathbf{a}`w'$. ☐

Thus in particular, $\vdash_{\mathrm{ZFC}} \exists y(\mathrm{Ord}\, y \wedge y \sim x)$, so by the theorem on ordinal definitions, # is a defined symbol of ZFC. In ZFC, we define the unary predicate symbols Card and ICard by $\mathrm{Card}\, x \leftrightarrow \exists y(x = \#y)$ and $\mathrm{ICard}\, x \leftrightarrow \mathrm{Card}\, x \wedge \omega \leq x$. The formula $\mathrm{Card}\,\mathbf{a}$ (resp. $\mathrm{ICard}\,\mathbf{a}$) means that $\mathbf{a}$ is a *cardinal* (resp. an *infinite cardinal*). We also say that $\#\mathbf{a}$ is the *cardinal of* $\mathbf{a}$.

3.10 Results on cardinals 1. We now derive basic results on cardinals.

(i) $\vdash_{\mathrm{ZFC}} x \sim \#x$;

(ii) $\vdash_{\mathrm{ZFC}} x \sim y \leftrightarrow \#x = \#y$;

(iii) $\vdash_{\mathrm{ZFC}} \mathrm{Ord}\, x \rightarrow \#x \leq x$;

(iv) $\vdash_{\mathrm{ZFC}} \mathrm{Card}\, x \rightarrow \#x = x$;

(v) $\vdash_{\mathrm{ZFC}} \mathrm{Ord}\, x \rightarrow y \subseteq x \rightarrow \exists z(\mathrm{Ord}\, z \wedge z \leq x \wedge z \sim y)$;

(vi) $\vdash_{\mathrm{ZFC}} x \subseteq y \rightarrow \#x \leq \#y$;

(vii) $\vdash_{\mathrm{ZFC}} \mathrm{Func}\, x \rightarrow \#\,\mathrm{Im}\, x \leq \#\,\mathrm{Dom}\, x$;

(viii) $\vdash_{\mathrm{ZFC}} x \in \omega \rightarrow \mathrm{Card}\, x$;

(ix) $\vdash_{\mathrm{ZFC}} \mathrm{Card}\, \omega$;

(x) $\vdash_{\mathrm{ZFC}} \#x = y \rightarrow \#\mathrm{P}x = \#\mathrm{P}y$.

From the definition of # we have (i), whence (ii) by symmetry and transitivity of $\sim$. Since $\#x$ is the first ordinal similar to x and since $x \sim x$, (iii) holds. Now assume that x is a cardinal, i.e., that $x = \#y$ for some y. Then $x \sim y$, so $\#x = \#y$ by (ii), and hence $\#x = x$. The other implication in (iv) is obvious. We now prove (v). Let x be an ordinal and $y \subseteq x$. Let $\mathbf{a}$ be $\{\langle \mu w(w \in z), z\rangle \mid z \in \mathrm{P}y - \{\dot{0}\}\}$. Clearly $\mathrm{Ch}\,\mathbf{a}y$. By Zermelo's theorem, there exists w such that $\mathrm{IFunc}\,w \wedge \mathrm{Ord}\,\mathrm{Dom}\,w \wedge \mathrm{Im}\,w = y$ and for any $x' \in \mathrm{Dom}\,w$, $y - \{w\text{‘}x'' \mid x'' < x'\} \neq \dot{0}$ and $w\text{‘}x' = \mathbf{a}\text{‘}(y - \{w\text{‘}x'' \mid x'' < x'\})$, that is,

$$w\text{‘}x' = \mu z(z \in y - \{w\text{‘}x'' \mid x'' < x'\}). \tag{6}$$

In particular, $\mathrm{Dom}\,w \sim y$, so it will suffice to prove that $\mathrm{Dom}\,w \leq x$. Assume that $x < \mathrm{Dom}\,w$. Then $w\text{‘}x \in y$, so $w\text{‘}x < x$. Let $\mathbf{b}$ be $\mu z(w\text{‘}z < z)$. Then $\mathbf{b} < \mathrm{Dom}\,w$, $w\text{‘}\mathbf{b} < \mathrm{Dom}\,w$, and since $\mathrm{IFunc}\,w$, $w\text{‘}\mathbf{b} \in y - \{w\text{‘}x'' \mid x'' < w\text{‘}\mathbf{b}\}$. By (6), $w\text{‘}w\text{‘}\mathbf{b}$ is the first ordinal in $y - \{w\text{‘}x'' \mid x'' < w\text{‘}\mathbf{b}\}$, so $w\text{‘}w\text{‘}\mathbf{b} \leq w\text{‘}\mathbf{b}$, whence $w\text{‘}w\text{‘}\mathbf{b} < w\text{‘}\mathbf{b}$ because w is injective and $w\text{‘}\mathbf{b} < \mathbf{b}$. This contradicts the fact that $\mathbf{b}$ is the first ordinal such that $w\text{‘}\mathbf{b} < \mathbf{b}$, and proves (v).

Assume $x \subseteq y$. By (i) there exists an injective function z on y whose image is $\#y$. Then $z \restriction x$ is an injective function on x whose image is a subset of $\#y$. By (v) and the transitivity of $\sim$, $x \sim w$ for some ordinal $w \leq \#y$. Then $\#x \leq w$, so $\#x \leq \#y$. This proves (vi). Let x be a function and let z be a choice function on $\mathrm{Dom}\,x$. Then $\{\langle z\text{‘}\{w \mid w \in \mathrm{Dom}\,x \wedge x\text{‘}w = y\}, y\rangle \mid y \in \mathrm{Im}\,x\}$ is an injective function on $\mathrm{Im}\,x$ whose image is a subset of $\mathrm{Dom}\,x$. So (vii) follows from (vi). We prove (viii) by transfinite induction on x. By (iii), $\#\dot{0} \leq \dot{0}$, so $\#\dot{0} = \dot{0}$ and hence $\mathrm{Card}\,\dot{0}$. Assume that $\dot{0} < x$ and $x < \omega$. Then $x = \mathrm{S}y$ for some ordinal y, by definition of ω. By induction hypothesis, $\mathrm{Card}\,y$, so by (iv), $\#y = y$. By (vi), $\#y \leq \#x$, so $y \leq \#x$. By (v) of §3.8 and by (ii), we have $\#x \neq \#y$. Thus $y < \#x$, and hence $x \leq \#x$. By (iii), we obtain $x = \#x$, whence $\mathrm{Card}\,x$. If $\omega < x$, there is nothing to prove. To prove (ix), assume that $\neg\,\mathrm{Card}\,\omega$. Then by (iii), $\#\omega < \omega$. Because ω is a limit ordinal, $\mathrm{S}\#\omega < \omega$, so $\#\mathrm{S}\#\omega = \mathrm{S}\#\omega$ by (viii) and (iv). Thus by (vi), $\mathrm{S}\#\omega \leq \#\omega$, which contradicts (v) of §3.8. From (vi) of §3.8, we obtain at once (x).

3.11 Results on cardinals 2. We define in ZFC $\oplus xy = \#((x \times \{\dot{0}\}) \cup (y \times \{\dot{1}\}))$ and $\otimes xy = \#(x \times y)$, with the usual abbreviations $(\mathbf{a} \oplus \mathbf{b})$ and $(\mathbf{a} \otimes \mathbf{b})$. Then:

(i) $\vdash_{\mathrm{ZFC}} x \cap y = \dot{0} \to \#(x \cup y) = \#x \oplus \#y$;

(ii) $\vdash_{\mathrm{ZFC}} \#(x \times y) = \#x \otimes \#y$;

(iii) $\vdash_{\mathrm{ZFC}} \#(x \cup y) \leq \#x \oplus \#y$;

(iv) $\vdash_{\mathrm{ZFC}} x \subseteq x' \to y \subseteq y' \to x \oplus y \leq x' \oplus y' \wedge x \otimes y \leq x' \otimes y'$;

(v) $\vdash_{\mathrm{ZFC}} \mathrm{Card}\,x \to x \leq x \oplus x$;

(vi) $\vdash_{\mathrm{ZFC}} \mathrm{Card}\,x \to \dot{2} \leq x \to x \oplus x \leq x \otimes x$;

(vii) $\vdash_{\mathrm{ZFC}} \mathrm{Card}\,x \to \forall y(y \in z \to \#y \leq x) \to \#\,\mathrm{Un}\,z \leq \#z \otimes x$;

(viii) $\vdash_{\mathrm{ZFC}} x \in \omega \to y \in \omega \to x \oplus y \in \omega \wedge x \otimes y \in \omega$;

(ix) $\vdash_{\mathrm{ZFC}} x \in \omega \to y \in \omega \to x \oplus \mathrm{S}y = \mathrm{S}(x \oplus y) \wedge x \otimes \mathrm{S}y = (x \otimes y) \oplus x$;

(x) $\vdash_{\mathrm{ZFC}} \mathrm{ICard}\,x \to \mathrm{Max}\,\mathrm{O}^2 x = x$;

(xi) $\vdash_{\mathrm{ZFC}} \mathrm{ICard}\,x \to x \oplus x = x \wedge x \otimes x = x$;

(xii) $\vdash_{\mathrm{ZFC}} \mathrm{ICard}\,x \to \mathrm{ICard}\,y \to x \oplus y = \mathrm{Max}\langle x, y\rangle \wedge x \otimes y = \mathrm{Max}\langle x, y\rangle$;

(xiii) for any $(m+n)$-ary function symbol $\mathbf{f}$ of a good extension T of ZFC, $\vdash_T \mathrm{ICard}\,x \to \#y \leq x \to \#\{\mathbf{f}x^n y^n \mid x_1 \in y \wedge \cdots \wedge x_n \in y\} \leq x$.

Let x' be an injective function on x with range $\#x \times \{\dot{0}\}$ and let y' be an injective function on y with range $\#y \times \{\dot{1}\}$. If $\langle z, w\rangle$ and $\langle z', w\rangle$ belong to $x' \cup y'$ with $z \neq z'$, then one of them must belong to x' and the other to y'. This implies $w \in x \cap y$. Thus if $x \cap y = \dot{0}$, then $x' \cup y'$ is a function on $x \cup y$, which is clearly injective and with range $(\#x \times \{\dot{0}\}) \cup (\#y \times \{\dot{1}\})$, as required to prove (i). To prove (ii), it suffices to note that if $x \sim x'$ and $y \sim y'$, then $x \times y \sim x' \times y'$. From (i) we obtain (iii) by noting that $x \cup y = x \cup (y - x)$ and $x \cap (y - x) = \dot{0}$. Assume that $x \subseteq x'$ and $y \subseteq y'$. Then $(x \times \{0\}) \cup (y \times \{1\}) \subseteq (x' \times \{0\}) \cup (y' \times \{1\})$ and $x \times y \subseteq x' \times y'$, whence (iv). The set $\{\langle w, w, \dot{0}\rangle \mid w \in x\}$ is an injective function on x whose range is included in $x \oplus x$, whence (v). Noting that $x \oplus x = x \otimes \dot{2}$, (vi) follows from (iv).

We now prove (vii). Assume that x is a cardinal and z a set all of whose members have cardinal at most x, and let w be a choice function on $\mathrm{P}(\mathrm{Un}\,z \times x)$. Let $\mathbf{a}$ be $\{y' \in \mathrm{P}(\mathrm{Un}\,z \times x) \mid \mathrm{Func}\,y' \wedge \mathrm{Dom}\,y' =$

$\#y \wedge \operatorname{Im} y' = y\}$. Note that for all $y \in z$, $\mathbf{a}$ is not empty. Thus for $y \in z$, $w`\mathbf{a}$ is an injective function on $\#y$ with range $y$. Let $\mathbf{b}$ be $\{\langle y, x'\rangle \mid y \in z \wedge x' < \#y\}$, and let $\mathbf{c}$ be $\{\langle (w`\mathbf{a})`x', y, x'\rangle \mid \langle y, x'\rangle \in \mathbf{b}\}$. Then $\mathbf{c}$ is a function on $\mathbf{b}$ whose range is $\operatorname{Un} z$. Thus $\# \operatorname{Un} z \leq \#\mathbf{b}$. Since $\mathbf{b} \subseteq z \times x$ and using (ii), we find $\# \operatorname{Un} z \leq \#z \otimes x$.

Let $\mathbf{a}$ be $(x \times \{\dot{0}\}) \cup (y \times \{\dot{1}\})$. We shall first prove

$$\vdash_{\mathrm{ZFC}} x \in \omega \to y \in \omega \to x \oplus \mathrm{S}y \sim \mathrm{S}(x \oplus y) \text{ and} \tag{7}$$

$$\vdash_{\mathrm{ZFC}} x \in \omega \to y \in \omega \to x \otimes \mathrm{S}y \sim (x \otimes y) \oplus x. \tag{8}$$

Assume that $x \in \omega$ and $y \in \omega$. Then $\{\langle w, w\rangle \mid w \in \mathbf{a}\} \cup \{\langle \mathbf{a}, y, \dot{1}\rangle\}$ is an injective function on $\mathrm{S}\mathbf{a}$ with range $(x \times \{\dot{0}\}) \cup (\mathrm{S}y \times \{\dot{1}\})$, which proves (7). The set $\{\langle w, w, \dot{0}\rangle \mid w \in x \times y\} \cup \{\langle\langle w, y\rangle, w, \dot{1}\rangle \mid w \in x\}$ is an injective function on $x \times \mathrm{S}y$ whose range is $((x \times y) \times \{\dot{0}\}) \cup (x \times \{\dot{1}\})$, whence (8). We now prove (viii) by transfinite induction on y. If $y = \dot{0}$, then $x \oplus y = x$ and $x \otimes y = \dot{0}$, so the result holds. Assume that $y < \omega$ and $y = \mathrm{S}y'$ for some ordinal y'. Then by (7), $x \oplus y \sim \mathrm{S}(x \oplus y')$, and by induction hypothesis $x \oplus y'$, and hence $\mathrm{S}(x \oplus y')$, belongs to ω, so $x \oplus y < \omega$. Thus the first part of (viii) is proved. By (8), we have $x \otimes y \sim (x \otimes y') \oplus x$, and using the induction hypothesis and the first part, this implies $x \otimes y < \omega$. Now (ix) follows from (viii), (7) and (8).

Let x be an infinite cardinal. By (v) and (vi), $x \leq x \otimes x$. Assume that $\operatorname{Max} \mathrm{O}^2 x = x$. Then by (v) of §3.5 and the definition of O^2, we have $x \times x \subseteq \{\mathrm{O}^2 y \mid y < x\}$. Now $\{\langle \mathrm{O}^2 y, y\rangle \mid y < x\}$ is an injective function on x whose range is $\{\mathrm{O}^2 y \mid y < x\}$. Hence we find $x \otimes x \leq x$. In summary, we have proved

$$\vdash_{\mathrm{ZFC}} \mathrm{ICard}\, x \to \operatorname{Max} \mathrm{O}^2 x = x \to x \otimes x = x. \tag{9}$$

We now prove (x) by tranfinite induction on x. We assume that $\operatorname{Max} \mathrm{O}^2 x \neq x$ and derive a contradiction. By (viii) of §3.5, this implies $\operatorname{Max} \mathrm{O}^2 x < x$. Set $\mathbf{a}$ to be $\operatorname{Max} \mathrm{O}^2 x$. Let $y < x$. Using (vi) of §3.5, we find $\operatorname{Max} \mathrm{O}^2 y \leq \mathbf{a}$, so $\operatorname{Max} \mathrm{O}^2 y < \mathrm{S}\mathbf{a}$. From this we deduce that $\mathrm{O}^2 y \in \mathrm{S}\mathbf{a} \times \mathrm{S}\mathbf{a}$. Since $x = \#x = \#\{\mathrm{O}^2 y \mid y < x\}$, we find

$$x \leq \mathrm{S}\mathbf{a} \otimes \mathrm{S}\mathbf{a}. \tag{10}$$

Now suppose that $\mathbf{a} < \omega$. Then by (viii), $\mathrm{S}\mathbf{a} \otimes \mathrm{S}\mathbf{a} < \omega$, and since $\omega \leq x$, this contradicts (10). So we must have $\omega \leq \mathbf{a}$, whence $\omega \leq \#\mathbf{a}$. Using $\mathbf{a} < x$, we also have $\#\mathbf{a} < x$. Thus we may apply the induction hypothesis to $\#\mathbf{a}$, and this yields $\operatorname{Max} \mathrm{O}^2 \#\mathbf{a} = \#\mathbf{a}$. By (9), we obtain $\#\mathbf{a} = \#\mathbf{a} \otimes \#\mathbf{a}$. By (i), $\#\mathrm{S}\mathbf{a} = \#\mathbf{a} \oplus \dot{1}$ and by (iv) and (vi), $\#\mathbf{a} \oplus \dot{1} \leq \#\mathbf{a} \otimes \#\mathbf{a} = \#\mathbf{a}$, so $\#\mathrm{S}\mathbf{a} = \#\mathbf{a}$. Hence $\mathrm{S}\mathbf{a} \otimes \mathrm{S}\mathbf{a} = \#\mathbf{a} \otimes \#\mathbf{a} = \#\mathbf{a} \leq \mathbf{a} < x$. This contradicts (10).

The second part of (xi) follows from (x) and (9). The first part follows from the second one with (v) and (vi). Let x and y be infinite cardinals. Using (iv) and (xi), we find $\operatorname{Max}\langle x, y\rangle = \operatorname{Max}\langle x \oplus \dot{0}, \dot{0} \oplus y\rangle \leq x \oplus y \leq \operatorname{Max}\langle x, y\rangle \oplus \operatorname{Max}\langle x, y\rangle = \operatorname{Max}\langle x, y\rangle$ and $\operatorname{Max}\langle x, y\rangle = \operatorname{Max}\langle x \otimes \dot{1}, \dot{1} \otimes x\rangle \leq x \otimes y \leq \operatorname{Max}\langle x, y\rangle \otimes \operatorname{Max}\langle x, y\rangle = \operatorname{Max}\langle x, y\rangle$. These prove (xii).

To prove (xiii), define $\mathbf{f}'$ by $\mathbf{f}'xy^n = \mathbf{f}\pi_1^m x \ldots \pi_m^m x y^n$. Let x be an infinite cardinal and y a set of cardinal at most x. Then $\{\mathbf{f}x^n y^n \mid x_1 \in y \wedge \cdots \wedge x_m \in y\} = \{\mathbf{f}'wy^n \mid w \in \times_m y \ldots y\}$. But this set is the range of the function $\{\langle \mathbf{f}'wy^n, w\rangle \mid w \in \times_m y \ldots y\}$ whose domain is $\times_m y \ldots y$. By (vii) of §3.10, $\#\{\mathbf{f}x^n y^n \mid x_1 \in y \wedge \cdots \wedge x_m \in y\} \leq \#\times_m y \ldots y$. On the other hand, $\#\times_m y \ldots y \leq x$ by (xi) and induction on m, so we find (xiii).

3.12 Alephs. We define

$$y = \mathbf{g}x'x \leftrightarrow (\mathrm{Ord}\, x \wedge \mathrm{ICard}\, y \wedge \forall z(z < x \to y \neq x'`z) \wedge \forall z(z < y \to \neg(\mathrm{ICard}\, z \wedge \forall w(w < x \to z \neq x'`w)))) \vee (\neg \mathrm{Ord}\, x \wedge y = \dot{0}).$$

Thus for x an ordinal, $\mathbf{g}x'x$ is the first ordinal which is a cardinal greater than ω and not equal to a member of $\operatorname{Im}(x' \restriction x)$. To prove that this is a valid definition in ZFC, we must check that $\vdash_{\mathrm{ZFC}} \mathrm{Ord}\, x \to \exists y(\mathrm{ICard}\, y \wedge y \notin \operatorname{Im}(x' \restriction x))$. We know that $\vdash_{\mathrm{ZFC}} \exists y(\mathrm{Ord}\, y \wedge y \notin \operatorname{Im}(x' \restriction x))$, so by transitivity of ordinals it will suffice to prove that any ordinal belongs to a cardinal, i.e., $\vdash_{\mathrm{ZFC}} \mathrm{Ord}\, x \to \exists y(\mathrm{Card}\, y \wedge x < y)$. This follows from the following theorem.

CANTOR'S THEOREM. $\vdash_{\mathrm{ZFC}} \mathrm{Ord}\, x \to x < \#\mathrm{P}x$.

Proof. We first prove $\vdash_{\mathrm{ZFC}} \mathrm{Card}\, x \to x < \#\mathrm{P}x$. Let x be a cardinal. The set $\{\langle\{y\}, y\rangle \mid y \in x\}$ is an injective function on x whose image is included in $\mathrm{P}x$, so $x \leq \#\mathrm{P}x$. Assume $x = \#\mathrm{P}x$. Then there exist an injective mapping z on x whose image is $\mathrm{P}x$. Let $\mathbf{a}$ be $\{y \mid y < x \wedge y \notin z`y\}$. Clearly $\mathbf{a} \in \mathrm{P}x$, and hence there exists

$y < x$ such that $z`y = \mathbf{a}$. Then $y \in \mathbf{a} \leftrightarrow y \notin \mathbf{a}$, which is a contradiction. Assume now that x is any ordinal, and that $\#\mathrm{P}x \le x$. By (vi) of §3.8 and (iv) of §3.10, $\#\mathrm{P}\#x = \#\mathrm{P}x = \#\#\mathrm{P}x$, and since $\#\#\mathrm{P}x \le \#x$ by the hypothesis and (vi) of §3.10, we obtain $\#\mathrm{P}\#x \le \#x$, which contradicts the first part. □

We define the unary function symbol $\aleph$ by transfinite recursion so that $\vdash_{\mathrm{ZFC}}\mathrm{Ord}\,x \to \aleph x = \mathbf{g}\{\langle \aleph y, y\rangle \mid y < x\}x$. We now prove that $\aleph$ is in fact a one to one correspondence between ordinals and infinite cardinals.

(i) $\vdash_{\mathrm{ZFC}}\mathrm{Ord}\,x \to y < x \to \aleph y < \aleph x$;

(ii) $\vdash_{\mathrm{ZFC}}\mathrm{ICard}\,x \leftrightarrow \exists y(\mathrm{Ord}\,y \wedge x = \aleph y)$;

(iii) $\vdash_{\mathrm{ZFC}}\aleph\dot{0} = \omega$;

(iv) $\vdash_{\mathrm{ZFC}}\mathrm{Ord}\,x \to \aleph\mathrm{S}x \le \#\mathrm{P}\aleph x$.

Let x be an ordinal and $x < y$. By definition of $\aleph$, $\aleph x$ is not equal to $\aleph y$ and does not belong to $\{\aleph z \mid z < y\}$. Since $\aleph y$ is the first infinite cardinal which does not belong to $\{\aleph z \mid z < y\}$, we obtain $\aleph y \le \aleph x$, whence (i). The implication from right to left in (ii) is immediate from the definition of $\aleph$. Let x be an infinite cardinal. By (i) above and (vii) of §3.3, we have $\neg\exists z\forall y(\exists w(\mathrm{Ord}\,w \wedge y = \aleph w) \to y \in z)$. In particular, $\neg\forall y(\exists w(\mathrm{Ord}\,w \wedge y = \aleph w) \to y \in x)$, that is, $\aleph w \notin x$ for some ordinal w. Then $x \le \aleph w$. Hence either $x = \aleph w$ or $x < \aleph w$. In the latter case, $x \in \{\aleph w' \mid w' < w\}$ by definition of $\aleph$. Thus in both cases $x = \aleph y$ for some ordinal y. For (iii), recall that ω is a cardinal, so it is the first infinite cardinal by definition of an infinite cardinal; (iv) is obvious from Cantor's theorem and the definition of $\aleph$.

3.13 The functional closure theorem. In this paragraph, we prove that we may form the "closure" of any set under given function symbols. In ZFC this can be done so that the cardinal of the closure of x is at most $\mathrm{Max}\langle \#x, \aleph\dot{0}\rangle$. Moreover, the closure is "minimal" in the sense that any set which includes x and is closed under those function symbols includes the closure of x.

Functional Closure Theorem. Let T be a good extension of ZFC in which there are an n-ary function symbol $\mathbf{h}$ and function symbols $\mathbf{f}_1, \ldots, \mathbf{f}_k$, each $\mathbf{f}_i$ of index $m_i + n$. Then there is a defined n-ary function symbol $\mathbf{h}'$ of T such that

(i) $\vdash_T \mathbf{h}y^n \subseteq \mathbf{h}'y^n$;

(ii) $\vdash_T \mathrm{Ord}\,x \to \#\mathbf{h}y^n \le \aleph x \to \#\mathbf{h}'y^n \le \aleph x$;

(iii) for each i, $\vdash_T x_1 \in \mathbf{h}'y^n \to \cdots \to x_{m_i} \in \mathbf{h}'y^n \to \mathbf{f}_i x^{m_i} y^n \in \mathbf{h}'y^n$;

(iv) $\vdash_T \mathbf{h}y^n \subseteq x \to \{\mathbf{f}_1 x^{m_1} y^n \mid x_1 \in x \wedge \cdots \wedge x_{m_1} \in x\} \subseteq x \to \cdots \to \{\mathbf{f}_k x^{m_k} y^n \mid x_1 \in x \wedge \cdots \wedge x_{m_k} \in x\} \subseteq x \to \mathbf{h}'y^n \subseteq x$.

Proof. Define $\mathbf{f}$ by

$$\mathbf{f}xy^n = \mathrm{Un}\{x, \{\mathbf{f}_1 x^{m_1} y^n \mid x_1 \in x \wedge \cdots \wedge x_{m_1} \in x\}, \ldots, \{\mathbf{f}_k x^{m_k} y^n \mid x_1 \in x \wedge \cdots \wedge x_{m_k} \in x\}\},$$

and define a function symbol $\mathbf{g}$ by transfinite recursion so that $\vdash_T \mathbf{g}\dot{0}y^n = \mathbf{h}y^n$, $\vdash_T \mathrm{Ord}\,y \to y = \mathrm{S}y' \to \mathbf{g}yy^n = \mathbf{f}\mathbf{g}y'y^n y^n$, and $\vdash_T \mathrm{Lim}\,y \to \mathbf{g}yy^n = \mathrm{Un}\{\mathbf{g}y'y^n \mid y' < y\}$. Since $x \subseteq \mathbf{f}xy^n$, we have $y < \omega \to y' < y \to \mathbf{g}y'y^n \subseteq \mathbf{g}yy^n$. Let us prove that the $\mathbf{h}'$ defined by $\mathbf{h}'y^n = \mathbf{g}\omega y^n$ has the desired properties. Since $\mathbf{h}y^n = \mathbf{g}\dot{0}y^n$, $\mathbf{h}y^n \subseteq \mathbf{g}\dot{1}y^n$, and so $\mathbf{h}y^n \subseteq \mathbf{h}'y^n$. Let z be an ordinal such that $\#\mathbf{h}y^n \le \aleph z$. By (iii), (xi), and (xiii) of §3.11, we have

$$\vdash_{\mathrm{ZFC}}\#w \le \aleph z \to \#\mathbf{f}wy^n \le \aleph z. \tag{11}$$

We prove by transfinite induction that $y < \omega \to \#\mathbf{g}yy^n \le \aleph z$. This holds if y is $\dot{0}$ because $\#\mathbf{h}y^n \le \aleph z$. If $y = \mathrm{S}y'$ and $y' < \omega$, then $\#\mathbf{g}y'y^n \le \aleph z$ by induction hypothesis, so $\#\mathbf{g}yy^n \le \aleph z$ by (11). If $\omega \le y$, the result is tautologically satisfied. By (vii) of §3.11, we find $\#\mathbf{h}'y^n \le \aleph z \otimes \omega \le \aleph z \otimes \aleph z = \aleph z$. This proves (ii). Suppose that $x_1, \ldots, x_{m_i}$ are members of $\mathbf{h}'y^n$. Then each x_j belongs to $\mathbf{g}z_j y^n$ for some $z_j < \omega$. Let y be the greatest ordinal among $z_1, \ldots, z_{m_i}$. Then $x_1, \ldots, x_{m_i}$ are members of $\mathbf{g}yy^n$, and hence $\mathbf{f}_i x^{m_i} y^n \in \mathbf{f}\mathbf{h}'yy^n y^n = \mathbf{h}'\mathrm{S}yy^n$. Since $\mathrm{S}y < \omega$, we have $\mathbf{f}_i x^{m_i} y^n \in \mathbf{h}'y^n$. Finally, assume that $\mathbf{h}y^n \subseteq x$ and that $\{\mathbf{f}_i x^{m_i} y^n \mid x_1 \in x \wedge \cdots \wedge x_{m_i} \in x\} \subseteq x$ for each i. We prove by transfinite induction on y that $y < \omega \to \mathbf{g}yy^n \subseteq x$, from which (iv) follows. If $y = \dot{0}$, this is assumed. So suppose that $y = \mathrm{S}y'$ for some $y' < \omega$. Since $y \subseteq x \to \mathbf{f}yy^n \subseteq x$ by the closure conditions on x, using the induction hypothesis we find $\mathbf{g}yy^n = \mathbf{f}\mathbf{g}y'y^n y^n \subseteq x$. □

3.14 The continuum hypothesis. The *generalized continuum hypothesis* is a translation into ZFC of the formula $\forall x(\mathrm{Ord}\, x \rightarrow \#\mathrm{P}\aleph x \leq \aleph \mathrm{S}x)$. The *continuum hypothesis* is a translation into ZFC of the formula $\#\mathrm{P}\aleph\dot{0} \leq \aleph\dot{1}$. Obviously the continuum hypothesis is inferrable from the generalized continuum hypothesis in ZFC.

Chapter Seven
The Consistency Proofs

§1 Simple interpretations of ZF

1.1 Mostowski collapsing. In this paragraph, T is a good extension of ZF. Assume that there is a unary predicate symbol $\mathbf{q}$ in T such that $\vdash_T \exists x \mathbf{q}x$. Since $L(\mathrm{ZF})$ has no function symbols, the simple interpretation I of $L(\mathrm{ZF})$ in $L(T)$ defined by $\mathbf{q}$ is an interpretation of $L(\mathrm{ZF})$ in T. We sometimes write $\mathbf{A_q}$ and $\mathbf{A^q}$ for $\mathbf{A}_I$ and $\mathbf{A}^I$. Note that I is also an interpretation of $L(\mathrm{ZF})$ in any extension of T.

A particular case is when $\vdash_T \mathbf{q}x \leftrightarrow x \in \mathbf{e}$ with $\vdash_T \exists x(x \in \mathbf{e})$. We then call the simple interpretation defined by $\mathbf{q}$ the *simple interpretation* $\mathbf{e}$. If only $\mathbf{e}$ is given, then we should define $\mathbf{q}$ in T by $\mathbf{q}x \leftrightarrow x \in \mathbf{e}$ before we can speak of the interpretation $\mathbf{e}$. But in practice, it is of course not necessary to pass to an extension by definitions: we may just replace $\mathbf{qa}$ by $\mathbf{a} \in \mathbf{e}$ everywhere. The notations $\mathbf{A_e}$ and $\mathbf{A^e}$ are used in that sense.

To simplify, when a formula $\mathbf{B}$ clearly satisfies $\vdash_T \mathbf{B} \leftrightarrow \mathbf{A}^I$, we also call $\mathbf{B}$ the interpretation of $\mathbf{A}$ by I. For example, the variables free in $\mathbf{A}$ need not be arranged in reverse alphabetical order when forming $\mathbf{A}^I$, and we may use $\forall \mathbf{x}(\mathrm{U}_I\mathbf{x} \to \mathbf{C}_I)$ as the interpretation by I of a closed formula $\forall \mathbf{x}\mathbf{C}$, even though it really is $\forall \mathbf{x}(\mathrm{U}_I\mathbf{x} \to \neg\neg\mathbf{C}_I)$.

A simple interpretation I of $L(\mathrm{ZF})$ in T is said to be *transitive* if $x \in y$ is complete in x for I, i.e., if $\vdash_T \mathrm{U}_I y \to x \in y \to \mathrm{U}_I x$.

MOSTOWSKI COLLAPSING THEOREM. Let I be an interpretation of $L(\mathrm{ZF})$ in T such that $=_I$ is $=$ and such that the interpretation by I of the extensionality axiom is a theorem of T. Suppose that there is a unary function symbol $\mathbf{h}$ in T such that

(i) $\vdash_T \mathrm{Ord}\,\mathbf{h}x$, i.e., $\mathbf{h}$ is an ordinal function symbol;
(ii) $\vdash_T \mathrm{Ord}\,x \to \exists y \forall z(\mathbf{h}z \le x \to z \in y)$;
(iii) $\vdash_T x \in_I y \to \mathbf{h}x < \mathbf{h}y$.

Then I is isomorphic to a transitive simple interpretation of $L(\mathrm{ZF})$ in an extension by definitions of T.

Proof. In T, define a unary function symbol $\mathbf{g}$ by transfinite recursion on $\mathbf{h}x$ so that $\vdash_T \mathbf{g}x = \{\mathbf{g}w \mid \mathrm{U}_I w \wedge w \in_I x\}$. This is a valid definition by the hypotheses (i)–(iii). Define the unary predicate symbol $\mathbf{q}$ by $\mathbf{q}x \leftrightarrow \exists y(\mathrm{U}_I y \wedge x = \mathbf{g}y)$. From $\vdash_T \exists x \mathrm{U}_I x$ we derive $\vdash_T \exists x \mathbf{q}x$. We let J be the simple interpretation of $L(\mathrm{ZF})$ defined by $\mathbf{q}$, and we prove that J is transitive and that $\mathbf{g}$ is an isomorphism from I to J. Let x and y be such that $\mathbf{q}y$ and $x \in y$. Then for some z with $\mathrm{U}_I z$, $y = \mathbf{g}z$, whence $x \in \mathbf{g}z$. By definition of $\mathbf{g}$, it follows that $x = \mathbf{g}w$ for some w such that $\mathrm{U}_I w$ and $w \in_I z$. In particular, $\mathbf{q}x$. Thus, J is transitive. To prove that $\mathbf{g}$ is an isomorphism from I to J, we must prove (i), (ii), and (iii) of ch. II §3.4. But (i) is given, so we need only prove

$$\vdash_T \mathrm{U}_I x \to \mathrm{U}_I y \to x = y \leftrightarrow \mathbf{g}x = \mathbf{g}y \text{ and} \tag{1}$$

$$\vdash_T \mathrm{U}_I x \to \mathrm{U}_I y \to x \in_I y \leftrightarrow \mathbf{g}x \in \mathbf{g}y. \tag{2}$$

We prove (1) by transfinite induction on $\mathbf{h}x$. The implication from left to right is just an equality axiom. Suppose that $\mathrm{U}_I x$ and $\mathrm{U}_I y$ and that $\mathbf{g}x = \mathbf{g}y$. The interpretation by I of the extensionality axiom is $\forall z(\mathrm{U}_I z \to z \in_I x \leftrightarrow z \in_I y) \to x = y$, which holds by hypothesis. Hence we need only derive that if $\mathrm{U}_I z$, then $z \in_I x \leftrightarrow z \in_I y$. By symmetry, it suffices to prove the implication from left to right. Suppose then that $z \in_I x$ for some z such that $\mathrm{U}_I z$. By definition of $\mathbf{g}$, $\mathbf{g}z \in \mathbf{g}x$. Since $\mathbf{g}x = \mathbf{g}y$, and again by definition of $\mathbf{g}$, $\mathbf{g}z = \mathbf{g}w$ for some w such that $\mathrm{U}_I w$ and $w \in_I y$. Since $\mathbf{h}z < \mathbf{h}x$, the induction hypothesis yields $z = w$, so $z \in_I y$.

The implication from left to right in (2) is obvious by the definition of $\mathbf{g}$. Assume that $\mathrm{U}_I x$, $\mathrm{U}_I y$, and $\mathbf{g}x \in \mathbf{g}y$. By definition of $\mathbf{g}$, $\mathbf{g}x = \mathbf{g}z$ for some z such that $\mathrm{U}_I z$ and $z \in_I y$. By (1), $x = z$, so $x \in_I y$. □

1.2 Interpretations of ZF. In this paragraph, we let T be a good extension of ZF and I a transitive simple interpretation of $L(\mathrm{ZF})$ in T. Our goal is to give sufficient conditions for I to be an interpretation of ZF in T. In fact, we shall give sufficient conditions for the interpretation by I of each axiom of ZF to be a theorem of T.

LEMMA 1. The interpretations by I of the extensionality axiom and the regularity axiom are theorems of T.

Proof. From the extensionality axiom and using the transitivity of I, we find $\vdash_T \mathrm{U}_I x \to \mathrm{U}_I y \to \forall z(\mathrm{U}_I z \to z \in x \leftrightarrow z \in y) \to x = y$, which is the interpretation by I of the extensionality axiom. By the regularity axiom, the conjunction of $\mathrm{U}_I x$ and $\exists y(\mathrm{U}_I y \wedge y \in x)$ implies $\exists y(y \in x \wedge \neg\exists z(\mathrm{U}_I z \wedge z \in x \wedge z \in y))$. Using the transitivity of I, we obtain $\vdash_T \mathrm{U}_I x \wedge \exists y(\mathrm{U}_I y \wedge y \in x) \to \exists y(\mathrm{U}_I y \wedge y \in x \wedge \neg\exists z(\mathrm{U}_I z \wedge z \in x \wedge z \in y))$, which is the interpretation by I of the regularity axiom. □

LEMMA 2. Suppose that $\vdash_T \mathrm{U}_I x \to \exists w(\mathrm{U}_I w \wedge \forall y(\mathrm{U}_I y \to y \subseteq x \to y \in w))$. Then the interpretation by I of the power set axiom is a theorem of T.

Proof. The interpretation of the power set axiom is $\mathrm{U}_I x \to \exists w(\mathrm{U}_I w \wedge \forall y(\mathrm{U}_I y \to \forall z(\mathrm{U}_I z \to z \in y \to z \in x) \to y \in w)$. By the transitivity of I, this is equivalent to the hypothesis. □

LEMMA 3. Suppose that $\vdash_T \mathrm{U}_I \omega$. Then the interpretation by I of the infinity axiom is a theorem of T.

Proof. By the substitution axioms it will suffice to derive

$$\begin{aligned}&\mathrm{U}_I\omega \wedge \exists y(\mathrm{U}_I y \wedge y \in \omega \wedge \forall z(\mathrm{U}_I z \to \neg z \in y))\\ &\qquad\qquad \wedge \forall y(\mathrm{U}_I y \to y \in \omega \to \exists z(\mathrm{U}_I z \wedge z \in \omega \wedge \forall w(\mathrm{U}_I w \to w \in z \leftrightarrow w \in y \vee w = y))).\end{aligned}$$

This follows by transitivity of I from $\vdash_T \mathrm{U}_I\omega$ and $\vdash_T \exists y(y \in \omega \wedge \forall z \neg z \in y) \wedge \forall y(y \in \omega \to \exists z(z \in \omega \wedge \forall w(w \in z \leftrightarrow w \in y \vee w = y)))$. □

LEMMA 4. Suppose that any defined $(n+1)$-ary function symbol $\mathbf{f}$ of T with a defining axiom of the form $\mathbf{f}yx_1 \dots x_n = \{x \mid x \in y \wedge \mathbf{A}_I\}$ is I-invariant. Then the interpretation by I of each subset axiom of ZF is a theorem of T.

Proof. By the version theorem it will suffice to consider a subset axiom of the form $\exists z \forall x(x \in z \leftrightarrow x \in y \wedge \mathbf{A})$, where y and z do not occur in $\mathbf{A}$ and where $x_1, \dots, x_n$ are the variables other than x free in $\mathbf{A}$. Its interpretation by I is

$$\mathrm{U}_I x \to \mathrm{U}_I x_1 \to \cdots \to \mathrm{U}_I x_n \to \exists z(\mathrm{U}_I z \wedge \forall x(\mathrm{U}_I x \to x \in z \leftrightarrow x \in y \wedge \mathbf{A}_I)). \tag{3}$$

Let $\mathbf{f}$ be defined by $\mathbf{f}yx_1 \dots x_n = \{x \mid x \in y \wedge \mathbf{A}_I\}$. Assume that $\mathrm{U}_I y \wedge \mathrm{U}_I x_1 \wedge \cdots \wedge \mathrm{U}_I x_n$; then $\mathrm{U}_I \mathbf{f}yx_1 \dots x_n$ by I-invariance of $\mathbf{f}$. From the definition of $\mathbf{f}$, we also have

$$\forall x(\mathrm{U}_I x \to x \in \mathbf{f}yx_1 \dots x_n \leftrightarrow x \in y \wedge \mathbf{A}_I), \tag{4}$$

so (3) holds by the substitution axioms. □

LEMMA 5. Suppose that, for any I-invariant defined $(n+1)$-ary function symbol $\mathbf{f}$ of T, $\vdash_T \mathrm{U}_I y \to \mathrm{U}_I x_1 \to \cdots \to \mathrm{U}_I x_n \to \exists z(\mathrm{U}_I z \wedge \forall x(x \in y \to \mathbf{f}xx_1 \dots x_n \subseteq z))$. Then the interpretation by I of each replacement axiom of ZF is a theorem of T.

Proof. It suffices to consider a replacement axiom of the form $\forall x \exists z \forall y(\mathbf{A} \leftrightarrow y \in z) \to \exists z \forall y(\exists x(x \in w \wedge \mathbf{A}) \to y \in z)$ where z and w do not occur in $\mathbf{A}$, and $x_1, \dots, x_n$ are the variables other than x and y free in $\mathbf{A}$. Its interpretation by I is

$$\begin{aligned}&\mathrm{U}_I w \to \mathrm{U}_I x_1 \to \cdots \to \mathrm{U}_I x_n \to \forall x(\mathrm{U}_I x \to \exists z(\mathrm{U}_I z \wedge \forall y(\mathrm{U}_I y \to \mathbf{A}_I \leftrightarrow y \in z)))\\ &\qquad\qquad\qquad\qquad \to \exists z(\mathrm{U}_I z \wedge \forall y(\mathrm{U}_I y \to \exists x(\mathrm{U}_I x \wedge x \in w \wedge \mathbf{A}_I) \to y \in z))\end{aligned}$$

Define $\mathbf{f}$ by

$$w = \mathbf{f}xx_1 \ldots x_n \leftrightarrow (\exists z(\mathrm{U}_I z \wedge \forall y(y \in z \leftrightarrow \mathrm{U}_I y \wedge \mathbf{A}_I) \wedge w = z)) \\ \vee (\neg\exists z(\mathrm{U}_I z \wedge \forall y(y \in z \leftrightarrow \mathrm{U}_I y \wedge \mathbf{A}_I)) \wedge w = x).$$

Then $\mathbf{f}$ is obviously I-invariant. Assume that $\mathrm{U}_I w, \mathrm{U}_I x_1, \ldots, \mathrm{U}_I x_n$ and that

$$\forall x(\mathrm{U}_I x \rightarrow \exists z(\mathrm{U}_I z \wedge \forall y(\mathrm{U}_I y \rightarrow \mathbf{A}_I \leftrightarrow y \in z))). \tag{5}$$

By the hypothesis of the lemma, we have $\exists z(\mathrm{U}_I z \wedge \forall x(x \in w \rightarrow \mathbf{f}xx_1 \ldots x_n \subseteq z))$. Thus it will suffice to prove that $\mathrm{U}_I z \wedge \forall x(x \in w \rightarrow \mathbf{f}xx_1 \ldots x_n \subseteq z)$ implies $\mathrm{U}_I z \wedge \forall y(\mathrm{U}_I y \rightarrow \exists x(\mathrm{U}_I x \wedge x \in w \wedge \mathbf{A}_I) \rightarrow y \in z)$. So assume that $\mathrm{U}_I z \wedge \forall x(x \in w \rightarrow \mathbf{f}xx_1 \ldots x_n \subseteq z)$ and let y be such that $\mathrm{U}_I y$ and $\mathrm{U}_I x \wedge x \in w \wedge \mathbf{A}_I$ for some x. We must prove that $y \in z$. By (5) and the transitivity of I, we have $\exists z(\mathrm{U}_I z \wedge \forall y(\mathbf{A}_I \leftrightarrow y \in z))$, whence $\forall y(\mathbf{A}_I \leftrightarrow y \in \mathbf{f}xx_1 \ldots x_n)$ by definition of $\mathbf{f}$. Now since $\mathbf{A}_I$, we have $y \in \mathbf{f}xx_1 \ldots x_n$, and since $x \in w$, $\mathbf{f}xx_1 \ldots x_n \subseteq z$. So $y \in z$. □

1.3 Absoluteness of defined symbols. We now assume that I is a transitive simple interpretation of ZF in T, where T is a good extension of ZF. As in most results that we shall prove in this section, the hypothesis that I be a transitive simple interpretation of ZF is often too strong: sometimes we do not use the transivity of I or we only need the interpretations of a few axioms of ZF. But we shall have no need of a more precise analysis than that which is given. Recall that by the interpretation extension theorem, I can be extended to an interpretation of any extension by definitions of ZF in a suitable extension by definitions of T (since it is possible to define a constant in T, and we agree to use the constant $\dot{0}$), which we continue to denote by I as we do not usually distinguish between ZF and its extensions by definitions.

Note that $=$ and $\in$ are absolute for I since I is simple. Moreover $x = y$ is complete in x for I, and since I is transitive, $x \in y$ is complete in x for I. We now use the results of ch. II §3.5 to prove that some defined symbols of ZF are absolute for I. In particular, we shall see that the general principles on set formation proved in Chapter VI yield absolute symbols when the "input" is absolute. To prove that a function symbol $\mathbf{f}$ is absolute, it will suffice, by Lemma 8 of ch. II §3.5, to derive a formula of the form $\mathbf{f}x_1 \ldots x_n = \mathbf{a}$ where $\mathbf{a}$ is absolute for I, or a formula of the form $y = \mathbf{f}x_1 \ldots x_n \leftrightarrow \mathbf{A}$ where $\mathbf{A}$ is absolute for I. Similarly, to prove that a predicate symbol $\mathbf{p}$ is absolute, it will suffice, by the same Lemma, to derive an equivalence of the form $\mathbf{p}x_1 \ldots x_n \leftrightarrow \mathbf{A}$ where $\mathbf{A}$ is absolute for I. We shall thus make a list of such formulae, and in all cases the fact that the formula is a theorem of ZF will be clear given the defining axiom of the symbol. In most cases the formula will also give at once the desired absoluteness by the general principles of ch. II §3.5, the completeness in x of $x \in y$, and the preceding results in the list. First note that since $\mathbf{x} \in \mathbf{y}$ is complete in $\mathbf{x}$ for I, $\mathbf{x} \in \mathbf{y} \wedge \mathbf{B}$ and $\neg(\mathbf{x} \in \mathbf{y} \rightarrow \mathbf{B})$ are complete in $\mathbf{x}$ for I. So by Lemma 5 of ch. II §3.5, $\exists\mathbf{x}(\mathbf{x} \in \mathbf{y} \wedge \mathbf{B})$ and $\forall\mathbf{x}(\mathbf{x} \in \mathbf{y} \rightarrow \mathbf{B})$ are absolute for I if $\mathbf{B}$ is absolute for I. From now on we use these facts and the lemmas of ch. II §3.5 without mention.

Lemma 1. If $\mathbf{f}$ is defined by $\mathbf{f}\mathbf{y}_1 \ldots \mathbf{y}_n = \{\mathbf{x} | \mathbf{A}\}$ as in the first theorem on set definitions and if $\mathbf{A}$ is absolute for I and complete in $\mathbf{x}$ for I, then $\mathbf{f}$ is absolute for I. If $\mathbf{g}$ is defined by $\mathbf{g}\mathbf{y}_1 \ldots \mathbf{y}_n = \{\mathbf{f}\mathbf{x}_1 \ldots \mathbf{x}_m\mathbf{y}'_1 \ldots \mathbf{y}'_k | \mathbf{A}\}$ as in the third theorem on set definitions, if $\mathbf{A}$ is absolute for I and complete in $\mathbf{x}_1, \ldots, \mathbf{x}_m$ for I, and if $\mathbf{f}$ is absolute for I, then $\mathbf{g}$ is absolute for I.

Proof. $\mathbf{x} \in \mathbf{y} \leftrightarrow \mathbf{A}$ is absolute for I, so it will suffice to prove that $\neg(\mathbf{x} \in \mathbf{y} \leftrightarrow \mathbf{A})$ is complete in $\mathbf{x}$ for I. The latter is tautologically equivalent to $(\mathbf{x} \in \mathbf{y} \wedge \neg\mathbf{A}) \vee (\mathbf{x} \notin \mathbf{y} \wedge \mathbf{A})$ which is complete in $\mathbf{x}$ by completeness in $\mathbf{x}$ of $\mathbf{x} \in \mathbf{y}$ and of $\mathbf{A}$. If $\mathbf{A}$ is absolute and complete in $\mathbf{x}_1, \ldots, \mathbf{x}_m$, then $\mathbf{A} \wedge \mathbf{y} = \mathbf{f}\mathbf{x}_1 \ldots \mathbf{x}_m\mathbf{y}'_1 \ldots \mathbf{y}'_k$ is absolute and complete in $\mathbf{x}_1, \ldots, \mathbf{x}_m, \mathbf{y}$ by absoluteness of $\mathbf{f}$. Hence we find as above that $\neg(\mathbf{y} \in \mathbf{w} \leftrightarrow \exists\mathbf{x}_1 \ldots \exists\mathbf{x}_m(\mathbf{A} \wedge \mathbf{y} = \mathbf{f}\mathbf{x}_1 \ldots \mathbf{x}_m\mathbf{y}'_1 \ldots \mathbf{y}'_k))$ is absolute and complete in $\mathbf{y}$. □

(i) $\dot{0} = \{y \mid y \neq y\}$;
(ii) $x \subseteq y \leftrightarrow \forall z(z \in x \rightarrow z \in y)$;
(iii) $\{\}_2 xy = \{w \mid w = x \vee w = y\}$;
(iv) $\{\}_1 x = \{\}_2 xx$;
(v) $\langle\rangle_2 xy = \{\}_2 \{\}_1 x \{\}_2 xy$;
(vi) $\langle\rangle_1 x = x$;

(vii) for $n \geq 3$, $\between_n x_1 \ldots x_n = \between_2 x_1 \between_{n-1} x_2 \ldots x_n$;

(viii) $\mathrm{Un}\, w = \{y \mid \exists x(x \in w \wedge y \in x)\}$.

For (viii) observe that $\exists x(x \in w \wedge y \in x)$ is complete in y by transitivity of I.

(ix) for $n \geq 3$, $\{\}_n x_1 \ldots x_n = \mathrm{Un}\{\}_2\{\}_1 x_1 \{\}_{n-1} x_2 \ldots x_n$;

(x) $\cup xy = \mathrm{Un}\{\}_2 xy$;

(xi) $\cap xy = \{w \mid w \in x \wedge w \in y\}$;

(xii) $-xy = \{w \mid w \in x \wedge w \notin y\}$;

(xiii) $\mathrm{S}x = \cup x\{\}_1 x$;

(xiv) $\dot{1} = \mathrm{S}\dot{0}$, $\dot{2} = \mathrm{S}\dot{1}$, $\dot{3} = \mathrm{S}\dot{2}$, $\dot{4} = \mathrm{S}\dot{3}$, $\dot{5} = \mathrm{S}\dot{4}$, $\dot{6} = \mathrm{S}\dot{5}$, $\dot{7} = \mathrm{S}\dot{6}$, $\dot{8} = \mathrm{S}\dot{7}$, $\dot{9} = \mathrm{S}\dot{8}$;

(xv) $\times_2 xy = \{w \mid \exists x' \exists y'(w = \between_2 x'y' \wedge x' \in x \wedge y' \in y)\}$;

For (xv) we must prove that $\exists x' \exists y'(w = \between_2 x'y' \wedge x' \in x \wedge y' \in y)$ is complete in w for I. This follows from the transitivity of I and the absoluteness of $\between_2$.

(xvi) for $n \geq 3$, $y = \times_n x_1 \ldots x_n \leftrightarrow y = \times_2 x_1 \times_{n-1} x_2 \ldots x_n$;

(xvii) for $n \geq 1$ and for $1 \leq i \leq n$, $y = \pi^n_i x \leftrightarrow (\exists x_1 \ldots \exists x_n(x = \between_n x_1 \ldots x_n \wedge y = x_i)) \vee (\neg \exists x_1 \ldots \exists x_n(x = \between_n x_1 \ldots x_n) \wedge y = \dot{0})$.

For (xvii) we must prove that $x = \between_n x_1 \ldots x_n$ is complete in $x_1, \ldots, x_n$ for I. This is obvious if $n = 1$, and if $n = 2$ it follows from the transitivity of I since $\vdash_{\mathrm{ZF}} x_1 \in \{x_1\} \wedge x_2 \in \{x_1, x_2\}$ and $\vdash_{\mathrm{ZF}} \{x_1\} \in \between_2 x_1 x_2 \wedge \{x_1, x_2\} \in \between_2 x_1 x_2$. The general case is proved by induction.

(xviii) $\mathrm{Dom}\, x = \{\pi^2_2 w \mid w \in x\}$;

(xix) $\mathrm{Im}\, x = \{\pi^2_1 w \mid w \in x\}$;

(xx) $\restriction xy = \{\langle x', x'' \rangle \mid x'' \in y \wedge \langle x', x'' \rangle \in x\}$;

(xxi) $\mathrm{Func}\, x \leftrightarrow x \subseteq \mathrm{Im}\, x \times \mathrm{Dom}\, x \wedge \forall y \forall y' \forall z(\langle y, z \rangle \in x \to \langle y', z \rangle \in x \to y = y')$;

(xxii) $\mathrm{IFunc}\, x \leftrightarrow \mathrm{Func}\, x \wedge \forall y \forall z \forall z'(\langle y, z \rangle \in x \to \langle y, z' \rangle \in x \to z = z')$;

(xxiii) $z = {}^{\backprime}xy \leftrightarrow ((\mathrm{Func}\, x \wedge y \in \mathrm{Dom}\, x) \wedge \langle z, y \rangle \in x) \vee (\neg(\mathrm{Func}\, x \wedge \neg y \in \mathrm{Dom}\, x) \wedge z = \dot{0})$;

(xxiv) $\circ xy = \{\langle x', y' \rangle \mid \exists z'(\langle x', z' \rangle \in x \wedge \langle z', y' \rangle \in y)\}$.

The absoluteness in (xxiv) follows from the completeness in x', y', z' of $\langle x', z' \rangle \in x \wedge \langle z', y' \rangle \in y$.

(xxv) $\mathrm{Tr}\, x \leftrightarrow \forall y \forall z(y \in x \to z \in y \to z \in x)$;

(xxvi) $\mathrm{Ord}\, x \leftrightarrow \mathrm{Tr}\, x \wedge \forall y(y \in x \to \mathrm{Tr}\, y)$;

LEMMA 2. If $\mathbf{f}$ is defined by $\mathbf{f}\mathbf{x}_1 \ldots \mathbf{x}_n = \mu\mathbf{y}\mathbf{A}$ where $\mathbf{A}$ is absolute for I, then $\mathbf{f}$ is absolute for I.

Proof. The actual defining axiom of $\mathbf{f}$ is $\mathbf{y} = \mathbf{f}\mathbf{x}_1 \ldots \mathbf{x}_n \leftrightarrow \mathrm{Ord}\, \mathbf{y} \wedge \mathbf{A} \wedge \forall \mathbf{z}(\mathbf{z} \in \mathbf{y} \to \neg \mathbf{A}[\mathbf{y}|\mathbf{z}])$ for some suitable $\mathbf{z}$, and its right-hand side is absolute. □

LEMMA 3. Let $\mathbf{g}$ be a defined $(n+2)$-ary function symbol and let $\mathbf{f}$ be defined using $\mathbf{g}$ as in the principle of transfinite recursion. If $\mathbf{g}$ is absolute for I, so is $\mathbf{f}$.

Proof. Recall that $\mathbf{f}$ is such that $\vdash_T (\mathrm{Ord}\, x \wedge \mathbf{f}xy^n = \mathbf{g}\{\langle \mathbf{f}x'y^n, x' \rangle \mid x' < x\}xy^n) \vee (\neg\, \mathrm{Ord}\, x \wedge \mathbf{f}xy^n = \dot{0})$. Thus if $\mathbf{h}$ is defined by $y = \mathbf{h}xy^n \leftrightarrow (\mathrm{Ord}\, x \wedge \mathrm{Func}\, y \wedge x = \mathrm{Dom}\, y \wedge \forall z(z \in x \to y^{\backprime}z = \mathbf{g}\{\langle y^{\backprime}z', z' \rangle \mid z' < z\}zy^n)) \vee (\neg\, \mathrm{Ord}\, x \wedge \mathbf{y} = \dot{0})$, then $\vdash_T \mathbf{f}xy^n = (\mathbf{h}\mathrm{S}xy^n)^{\backprime}x$. But $\forall z(z \in x \to y^{\backprime}z = \mathbf{g}\{\langle y^{\backprime}z', z' \rangle \mid z' < z\}zy^n)$ is absolute. So $\mathbf{h}$ is absolute for I, and hence $\mathbf{f}$ is absolute for I. □

(xxviii) $\mathrm{Max}\, x = \pi^2_1 x \cup \pi^2_2 x$;

(xxix) $\mathrm{MP}\, x = \langle \mu y \exists z(z \in \mu w \neg(w \times w \subseteq x) \wedge \langle y, z \rangle \notin x), \mu y(\langle \mu y' \exists z(z \in \mu w \neg(w \times w \subseteq x) \wedge \langle y', z \rangle \notin x), y \rangle \notin x) \rangle$;

(xxx) O^2 is defined by transfinite recursion using $\mathbf{g}zx = \mathrm{MP}\{z^{\backprime}y \mid y < x\}$;

(xxxi) $\mathrm{Lim}\, x \leftrightarrow \mathrm{Ord}\, x \wedge \neg x = 0 \wedge \neg \exists y(\mathrm{Ord}\, y \wedge x = \mathrm{S}y)$;

(xxxii) $\omega = \mu y\, \mathrm{Lim}\, y$.

Remark. Recall that, by the remark in ch. VI §3.3, the absoluteness of all the symbols listed above except (xxxii) holds not only for transitive simple interpretations of ZF, but more generally for transitive simple interpretations of ZF_ω.

§2 The predicate of constructibility

2.1 Introduction. We shall define in ZF a unary predicate symbol L, called the *predicate of constructibility*, and we shall consider the simple interpretation of $L(\mathrm{ZF})$ defined by L. Intuitively, we would like to define the constructible sets by transfinite induction as follows. We first define $\mathbf{f}$ by transfinite recursion so that $\mathbf{f}x = \mathrm{Un}\{\mathrm{Def}\, y \mid y < x\}$ for x an ordinal, where $\mathrm{Def}\, y$ is the the set of subsets of y which are characterized by a formula with parameters and quantifiers restricted to sets in y, i.e., the set of all sets of the form $\{z \mid z \in y \wedge \mathbf{A}^y\}$ with y not occurring in $\mathbf{A}$. Then a set is constructible if it lies in $\mathbf{f}x$ for some ordinal x. This definition ensures that the interpretations by L of the subset axioms hold. Moreover, the interpretations of the other axioms can be proved to hold as well by simple arguments. Unfortunately, a more precise look at this "definition" brings us to the conclusion that it is flawed. It is indeed possible to define a unary function symbol Def in ZF such that $\vdash_{\mathrm{ZF}} \mathrm{Def}\, x \subseteq \mathrm{P}x$ and for any formula $\mathbf{A}$ of ZF with parameters and quantifiers restricted to a variable x, $\vdash_{\mathrm{ZF}} y \subseteq x \wedge \forall z(z \in y \leftrightarrow \mathbf{A}) \to y \in \mathrm{Def}\, x$. But it makes no sense to ask that for every $y \in \mathrm{Def}\, x$ there exists a formula $\mathbf{A}$ as above such that $z \in y \leftrightarrow \mathbf{A}$. The formal definition of L will thus necessarily differ from this naïve definition, but it will retain all of the above properties. We shall even be able to prove that the interpretations by L of the axiom of choice and the generalized continuum hypothesis hold. We shall then use the interpretation theorem to obtain a result on consistency. Note that the proof of this result will be entirely finitary. Here, instead of defining the symbol Def, constructible sets will be defined more directly as the image of a function symbol on ordinals. The hard part will be to build that function symbol so that it satisfies our requirements. The property of Def that we mentioned above can be found in the theorem on definability in §2.3.

The method given here is essentially the original one of Gödel [4], and we use the same notations with a few minor differences. (Gödel's original proof was written for the first-order theory NBG, but all his arguments translate into ZF in a straightforward manner.)

2.2 Definition of L. We define the binary function symbols $\mathfrak{F}_1, \ldots, \mathfrak{F}_9$, called the *Gödel symbols*, as follows (all the definitions are legit according to the third theorem on definitions or the proposition 1 of ch. II §2.2):

(i) $z = \mathfrak{F}_1 xy \leftrightarrow \forall w(w \in z \leftrightarrow \exists x' \exists y'(w = \langle x', y' \rangle \wedge w \in x \wedge y' \in x'))$;

(ii) $z = \mathfrak{F}_2 xy \leftrightarrow \forall w(w \in z \leftrightarrow \exists x'(w = \langle x', x' \rangle \wedge w \in x))$;

(iii) $z = \mathfrak{F}_3 xy \leftrightarrow \forall w(w \in z \leftrightarrow \exists x' \exists y'(w = \langle x', y' \rangle \wedge w \in x \wedge x' \in y))$;

(iv) $z = \mathfrak{F}_4 xy \leftrightarrow \forall w(w \in z \leftrightarrow \exists x' \exists y'(w = \langle x', y' \rangle \wedge w \in x \wedge y' \in y))$;

(v) $z = \mathfrak{F}_5 xy \leftrightarrow \forall w(w \in z \leftrightarrow \exists x' \exists y'(w = \langle x', y' \rangle \wedge w \in x \wedge \langle y', x' \rangle \in y))$;

(vi) $z = \mathfrak{F}_6 xy \leftrightarrow \forall w(w \in z \leftrightarrow \exists x' \exists y' \exists z'(w = \langle x', y', z' \rangle \wedge w \in x \wedge \langle y', x', z' \rangle \in y))$;

(vii) $z = \mathfrak{F}_7 xy \leftrightarrow \forall w(w \in z \leftrightarrow \exists x' \exists y' \exists z'(w = \langle x', y', z' \rangle \wedge w \in x \wedge \langle z', x', y' \rangle \in y))$;

(viii) $z = \mathfrak{F}_8 xy \leftrightarrow z = x - y$;

(ix) $z = \mathfrak{F}_9 xy \leftrightarrow z = x \cap \mathrm{Dom}\, y$.

Note that the $\mathfrak{F}_i$ are absolute for transitive simple interpretations of ZF_ω, and that $\vdash_{\mathrm{ZF}} \mathfrak{F}_i xy \subseteq x$. We then define the unary function symbols J_0, J_1, J_2 as follows:

(x) $\mathrm{J}_0 x = \pi^3_1 \mathrm{O}^3 x$;

(xi) $\mathrm{J}_1 x = \pi^3_2 \mathrm{O}^3 x$;

(xii) $\mathrm{J}_2 x = \pi^3_3 \mathrm{O}^3 x$.

Then the J_i are absolute for transitive simple interpetations of ZF_ω. By (viii) and (ix) of ch. VI §3.5, we have $\vdash_{ZF}\mathrm{Ord}\,x \to \dot{0} < J_0x \to J_1x < x \land J_2x < x$. Hence we may define a function symbol C by transfinite recursion so that

$$\begin{aligned}\vdash_{ZF} y = Cx \leftrightarrow{} & (\mathrm{Ord}\,x \land J_0x = \dot{0} \land \forall w(w \in y \leftrightarrow \exists z(z < x \land w = Cz)))\\ & \lor (\mathrm{Ord}\,x \land J_0x = \dot{1} \land y = \mathfrak{F}_1CJ_1xCJ_2x)\\ & \lor (\mathrm{Ord}\,x \land J_0x = \dot{2} \land y = \mathfrak{F}_2CJ_1xCJ_2x)\\ & \lor (\mathrm{Ord}\,x \land J_0x = \dot{3} \land y = \mathfrak{F}_3CJ_1xCJ_2x)\\ & \lor (\mathrm{Ord}\,x \land J_0x = \dot{4} \land y = \mathfrak{F}_4CJ_1xCJ_2x)\\ & \lor (\mathrm{Ord}\,x \land J_0x = \dot{5} \land y = \mathfrak{F}_5CJ_1xCJ_2x)\\ & \lor (\mathrm{Ord}\,x \land J_0x = \dot{6} \land y = \mathfrak{F}_6CJ_1xCJ_2x)\\ & \lor (\mathrm{Ord}\,x \land J_0x = \dot{7} \land y = \mathfrak{F}_7CJ_1xCJ_2x)\\ & \lor (\mathrm{Ord}\,x \land J_0x = \dot{8} \land y = \mathfrak{F}_8CJ_1xCJ_2x)\\ & \lor (\mathrm{Ord}\,x \land J_0x = \dot{9} \land y = \mathfrak{F}_9CJ_1xCJ_2x)\\ & \lor (\mathrm{Ord}\,x \land \dot{9} < J_0x \land y = \{CJ_1x, CJ_2x\})\\ & \lor (\neg\,\mathrm{Ord}\,x \land y = \dot{0}).\end{aligned}$$

Finally, we define L by $Lx \leftrightarrow \exists y(\mathrm{Ord}\,y \land x = Cy)$. If x is constructible, the first ordinal y such that $x = Cy$ is called the *order* of x. Formally, we define Od by $y = \mathrm{Od}\,x \leftrightarrow (\mathrm{Ord}\,y \land x = Cy \land \forall z(z < y \to x \neq Cz)) \lor (\neg Lx \land y = \dot{0})$. Observe that $\vdash_{ZF}Lx \to Ly \to x \neq y \to \mathrm{Od}\,x \neq \mathrm{Od}\,y$. Finally, we define C^* by $C^*x = \{Cy \mid y < x\}$.

2.3 Constructibility and definability. In this paragraph we review the fundamental properties of L. The main result is the theorem on definability which says that "definable" sets are constructible.

Lemma 1. $\vdash_{ZF}Lx \to y \in x \to Ly \land \mathrm{Od}\,y \in \mathrm{Od}\,x$.

Proof. Suppose that x is constructible and that $y \in x$. We derive the result using the priniciple of transfinite induction on $\mathrm{Od}\,x$. If $J_0\,\mathrm{Od}\,x = 0$, then $x = \{Cz \mid z < \mathrm{Od}\,x\}$. So $y = Cz$ for some $z < \mathrm{Od}\,x$, and hence Ly and $\mathrm{Od}\,y \le z < \mathrm{Od}\,x$. If $J_0\,\mathrm{Od}\,x = \dot{1}$, then $x = \mathfrak{F}_1CJ_1\,\mathrm{Od}\,xCJ_2\,\mathrm{Od}\,x$; so $x \subseteq CJ_1\,\mathrm{Od}\,x$. Thus y belongs to $CJ_1\,\mathrm{Od}\,x$ and $J_1\,\mathrm{Od}\,x < \mathrm{Od}\,x$. By induction hypothesis, Ly and $\mathrm{Od}\,y < J_1\,\mathrm{Od}\,x < \mathrm{Od}\,x$. We proceed similary for $\dot{2}, \ldots, \dot{9}$. Finally, if $\dot{9} < J_0\,\mathrm{Od}\,x$, then $y = CJ_1\,\mathrm{Od}\,x$ or $y = CJ_2\,\mathrm{Od}\,x$, but in either case Ly and $\mathrm{Od}\,y < \mathrm{Od}\,x$. □

Lemma 2. $\vdash_{ZF}\forall y(y \in x \to Ly) \to \exists y(Ly \land x \subseteq y)$.

Proof. Assume that $\forall y(y \in x \to Ly)$ and using (v) of ch. VI §3.3 let z be an ordinal such that $y \in x \to \mathrm{Od}\,y < z$. Then $x \subseteq \{Cx' \mid x' < z\}$. There exists an ordinal w satisfying $z \le w$ such that $O^2w = \langle\dot{0}, z\rangle$, whence $\{Cx' \mid x' < z\} \subseteq \{Cx' \mid x' < w\}$. Since $J_0w = 0$, $Cw = \{Cx' \mid x' < w\}$. Consequently, $x \subseteq Cw$. □

From the two Lemmas we now derive some properties of stability of L.

(i) For $1 \le i \le 9$, $\vdash_{ZF}Lx \to Ly \to L\mathfrak{F}_ixy$;

(ii) $\vdash_{ZF}Lx \to Ly \to L\{x, y\}$;

(iii) $\vdash_{ZF}Lx_1 \to \cdots \to Lx_n \to L\langle x_1, \ldots, x_n\rangle$.

Suppose that $x = Cz$ and $y = Cw$ for some ordinals z and w. There exists an ordinal x' such that $J_0x' = \dot{1}$, $J_1x' = z$, and $J_2x' = w$. So $\mathfrak{F}_1xy = Cx'$. Similarly for $\mathfrak{F}_2, \ldots, \mathfrak{F}_9$. This proves (i). Under the same hypotheses, we may choose an ordinal x' such that $J_0x' = S\dot{9}$, $J_1x' = z$, and $J_2x' = w$. Then $\{x, y\} = Cx'$. By induction using (ii), we find (iii).

(iv) $\vdash_{ZF}Lx_1 \to \cdots \to Lx_n \to L(x_1 \times \cdots \times x_n)$;

(v) $\vdash_{ZF}Lx \to Ly \to L(x \times^1 y)$;

(vi) $\vdash_{ZF}Lx \to Ly \to L(x \times^2 y)$.

For (iv), it suffices to derive $\mathrm{L}x \to \mathrm{L}y \to \mathrm{L}(x\times y)$. Let x and y be constructible sets. By (iii), z is constructible for all $z \in x\times y$. So by Lemma 2, there is a constructible set w such that $x\times y \subseteq w$. But then $x\times y = \mathfrak{F}_4\mathfrak{F}_3 wxy$, so $\mathrm{L}(x \times y)$. Similarly, there is a constructible set w such that $x \times^1 y \subseteq w$ (resp. $x \times^2 y \subseteq w$). Then $x \times^1 y = \mathfrak{F}_6 w(x \times y)$ (resp. $x \times^2 y = \mathfrak{F}_7 w(x \times y)$), whence (v) (resp. (vi)).

(vii) $\vdash_{\mathrm{ZF}} \mathrm{L}x \to \mathrm{L}\,\mathrm{Cnv}\,x$.

Let $\langle y, z\rangle$ be a member of x. Then by Lemma 1, $\{y, z\}$ and hence y and z are constructible. By (iii), $\langle z, y\rangle$ is contructible. By Lemma 2, there exists a constructible set w such that $\mathrm{Cnv}\,x \subseteq w$, and then $\mathrm{Cnv}\,x = \mathfrak{F}_5 wx$ is constructible.

(viii) $\vdash_{\mathrm{ZF}} \mathrm{L}x \to \mathrm{L}y \to \mathrm{L}(x \cup y)$;

(ix) $\vdash_{\mathrm{ZF}} \mathrm{L}x \to \mathrm{L}y \to \mathrm{L}(x \cap y)$;

(x) $\vdash_{\mathrm{ZF}} \mathrm{L}x_1 \to \cdots \to \mathrm{L}x_n \to \mathrm{L}\{x_1, \dots, x_n\}$.

Let x and y be constructible. By Lemmas 1 and 2, there is a constructible set z such that $x \cup y \subseteq z$. Then $x \cup y = \mathfrak{F}_8 w\mathfrak{F}_8\mathfrak{F}_8 wxy$ and $x \cap y = \mathfrak{F}_8 x\mathfrak{F}_8 xy$, so both are constructible. Since $\{x_1, \dots, x_n\} = \{x_1, x_1\} \cup \{x_2, \dots, x_n\}$, (x) follows from (ii) and (viii) by induction.

(xi) $\vdash_{\mathrm{ZF}} \mathrm{L}x \to \mathrm{L}\,\mathrm{Dom}\,x$;

(xii) $\vdash_{\mathrm{ZF}} \mathrm{L}x \to \mathrm{L}\,\mathrm{Im}\,x$.

Let x be constructible. By Lemma 1, $y \in \mathrm{Dom}\,x$ implies $\mathrm{L}y$. So by Lemma 2, there exists a constructible set z such that $\mathrm{Dom}\,x \subseteq z$. Then $\mathrm{Dom}\,x = \mathfrak{F}_9 zx$. Finally, $\mathrm{Im}\,x = \mathrm{Dom}\,\mathrm{Cnv}\,x$.

LEMMA 3. For any $n \geq 1, 1 \leq i \leq n, 1 \leq j \leq n$, $\vdash_{\mathrm{ZF}} \mathrm{L}x \to \mathrm{L}y \to \exists z(\mathrm{L}z \wedge \forall x_1 \dots \forall x_n(x_1 \in x \to \cdots \to x_n \in x \to \langle x_i, x_j\rangle \in y \leftrightarrow \langle x_1, \dots, x_n\rangle \in z))$.

Proof. Assume that x and y are constructible. We shall exhibit a constructible set z satsfying the theorem; this set will be seen in each case to be constructible on the basis of (i)–(xii), and we shall not mention it. We first note that if $i = j$, then $\times_n x \dots x\,\mathrm{Dom}\,\mathfrak{F}_2 yyx \dots x$, where $\mathrm{Dom}\,\mathfrak{F}_2 yy$ stands at the i-th place, is as desired. In particular, the result is proved for $n = 1$. If $i = 1$ and $j = 2$, then the set $(\times_{n-2} x \dots x) \times^2 y$ (read y if $n = 2$) is as desired. We suppose from now on that $n \geq 2$, and we prove the result by induction on n. We shall distinguish the following cases:

(i) $i = 1$ and $j > 2$;

(ii) $i > 1$ and $i < j$;

(iii) $j < i$.

By induction hypothesis, there is a constructible set w such that for all $x_1, x_3, \dots, x_n$ in x, $\langle x_1, x_j\rangle \in y \leftrightarrow \langle x_1, x_3, \dots, x_n\rangle \in w$. For (i), we may then take z to be $x \times^1 w$. By induction hypothesis, there exists a constructible set w such that for all $x_2, \dots, x_n$ in x, $\langle x_i, x_j\rangle \in y \leftrightarrow \langle x_2, \dots, x_n\rangle \in w$. Hence we may take $x \times w$ for z in case (ii). For (iii), since $\mathrm{Cnv}\,y$ is constructible, there exists by the preceding cases a constructible set z such that for all $x_1, \dots, x_n$ in x, $\langle x_j, x_i\rangle \in \mathrm{Cnv}\,y \leftrightarrow \langle x_1, \dots, x_n\rangle \in z$. Clearly z is as required. □

THEOREM ON DEFINABILITY. Let $\mathbf{x}_1, \dots, \mathbf{x}_n$, $\mathbf{x}$, and $\mathbf{y}$ be distinct variables. For any formula $\mathbf{A}$ of $L(\mathrm{ZF})$ with free variables among $\mathbf{x}_1, \dots, \mathbf{x}_n$, $\vdash_{\mathrm{ZF}} \mathrm{L}\mathbf{x} \to \exists\mathbf{y}(\mathrm{L}\mathbf{y} \wedge \forall\mathbf{x}_1 \dots \forall\mathbf{x}_n(\mathbf{x}_1 \in \mathbf{x} \to \cdots \to \mathbf{x}_n \in \mathbf{x} \to \langle\mathbf{x}_1, \dots, \mathbf{x}_n\rangle \in \mathbf{y} \leftrightarrow \mathbf{A}_{\mathrm{L}}))$.

Proof. We proceed by induction on the length of $\mathbf{A}$, and we assume (without loss of generality) that $\mathbf{x}_1, \dots, \mathbf{x}_n$, $\mathbf{x}$, and $\mathbf{y}$ are $x_1, \dots, x_n$, x, and y, respectively. If $\mathbf{A}$ is atomic, then $\mathbf{A}$ is of the form $x_i = x_j$ or $x_i \in x_j$ for some i and j, and $\mathbf{A}_{\mathrm{L}}$ is $\mathbf{A}$. Suppose that $\mathbf{A}$ is $x_i = x_j$ (resp. $x_i \in x_j$), and let $\mathbf{a}$ be $\mathfrak{F}_2(x_i \times x_i)x_i$ (resp. $\mathfrak{F}_1(x_i \times x_i)x_i$). Then for x_i and x_j in x, $\langle x_i, x_j\rangle \in \mathbf{a} \leftrightarrow \mathbf{A}_{\mathrm{L}}$, and by Lemma 3, there is a constructible set y such that for x_i and x_j in x, $\langle x_i, x_j\rangle \in \mathbf{a} \leftrightarrow \langle x_1, \dots, x_n\rangle \in y$. Thus, y satisfies the theorem.

Suppose that $\mathbf{A}$ is $\mathbf{B} \vee \mathbf{C}$; then $\mathbf{A}_{\mathrm{L}}$ is $\mathbf{B}_{\mathrm{L}} \vee \mathbf{C}_{\mathrm{L}}$. By induction hypothesis, there are constructible sets y_1 and y_2 such that for all $x_1, \dots, x_n$ in x, $\langle x_1, \dots, x_n\rangle \in y_1 \leftrightarrow \mathbf{B}_{\mathrm{L}}$ and $\langle x_1, \dots, x_n\rangle \in y_2 \leftrightarrow \mathbf{C}_{\mathrm{L}}$; then $y_1 \cup y_2$ is constructible and such that for all $x_1, \dots, x_n$ in x, $\langle x_1, \dots, x_n\rangle \in y_1 \cup y_2 \leftrightarrow \mathbf{A}_{\mathrm{L}}$. Suppose that $\mathbf{A}$ is $\neg\mathbf{B}$, so that $\mathbf{A}_{\mathrm{L}}$ is $\neg\mathbf{B}_{\mathrm{L}}$. By induction hypothesis, there is a constructible set z such that for all $x_1, \ldots, x_n$ in x, $\langle x_1, \ldots, x_n\rangle \in z \leftrightarrow \mathbf{B}_{\mathrm{L}}$. We then see that the constructible set $(\times_n x \ldots x) - z$ is as required.

Finally, suppose that $\mathbf{A}$ is $\exists z\mathbf{B}$, so that $\mathbf{A}_{\mathrm{L}}$ is $\exists z(\mathrm{L}z \wedge \mathbf{B}_{\mathrm{L}})$. Define an n-ary function symbol $\mathbf{f}$ by

$$y = \mathbf{f}x_1 \ldots x_n \leftrightarrow (\mathrm{L}y \wedge \mathbf{B}_{\mathrm{L}}[z|y] \wedge \forall w(\mathrm{L}w \to \mathrm{Od}\, w < \mathrm{Od}\, y \to \neg\mathbf{B}_{\mathrm{L}}[z|w])) \vee (\neg\exists z(\mathrm{L}z \wedge \mathbf{B}_{\mathrm{L}}) \wedge y = \mathrm{C}\dot{0}),$$

i.e., $\mathbf{f}x_1 \ldots x_n$ denotes the constructible set z of smallest order such that $\mathbf{B}_{\mathrm{L}}$ if such a set exists, and $\mathrm{C}\dot{0}$ otherwise. Then clearly $\mathbf{f}x_1 \ldots x_n$ is constructible, so by Lemma 2 there is a constructible set w' such that $x \cup \{\mathbf{f}x_1 \ldots x_n \mid x_1 \in x \wedge \cdots \wedge x_n \in x\} \subseteq w'$. By induction hypothesis, there is a constructible set w such that for all $z, x_1, \ldots, x_n$ in w', $\langle z, x_1, \ldots, x_n\rangle \in w \leftrightarrow \mathbf{B}_{\mathrm{L}}$. Let $\mathbf{a}$ be $\mathrm{Dom}(w \cap \times_{n+1} w' \ldots w')$, and let us verify that $\mathbf{a}$ satisfies the theorem, namely that for all $x_1, \ldots, x_n$ in x, $\langle x_1, \ldots, x_n\rangle \in \mathbf{a} \leftrightarrow \mathbf{A}_{\mathrm{L}}$. Fix $x_1, \ldots, x_n$ in x. Now $\langle x_1, \ldots, x_n\rangle \in \mathbf{a}$ if and only if $\exists z(z \in w' \wedge \langle z, x_1, \ldots, x_n\rangle \in w \cap \times_{n+1} w' \ldots w')$. Since $x \subseteq w'$, this is the case if and only if $\exists z(z \in w' \wedge \mathbf{B}_{\mathrm{L}})$, and since $\{\mathbf{f}x_1 \ldots x_n \mid x_1 \in x \wedge \cdots \wedge x_n \in x\} \subseteq w'$, this holds if and only if $\exists z(\mathrm{L}z \wedge \mathbf{B}_{\mathrm{L}})$, as was to be shown. ☐

2.4 L is an interpretation of ZF. We now turn to the proof that L is an interpretation of ZF.

LEMMA 1. The simple interpretation defined by L is an interpretation of ZF_ω.

Proof. We have $\vdash_{\mathrm{ZF}} \mathrm{LC}x$, so $\vdash_{\mathrm{ZF}} \exists x\mathrm{L}x$. This means that L is an interpretation of $L(\mathrm{ZF})$ in (an extension by definitions of) ZF. Note that by Lemma 1 of §2.3, L is transitive. Thus by Lemma 1 of §1.2, the interpretations of the extensionality axiom and the regularity axiom hold.

Suppose that y is constructible. Then by Lemma 2 of §2.3, there exists a constructible set z such that $\mathrm{L}x \wedge x \subseteq y \to x \in z$. By Lemma 2 of §1.2, the interpretation of the power set axiom holds.

We now derive the interpretations of the subset axioms. Suppose that the variables free in $\mathbf{A}$ are x, $x_1, \ldots, x_n$. Let $x, x_1, \ldots, x_n$ be constructible. Then $x \cup \{x_1, \ldots, x_n\}$ is constructible. By the theorem on definability, there exists a constructible set w such that for all $z \in x$, $\langle z, x_1, \ldots, x_n\rangle \in w \leftrightarrow \mathbf{A}_{\mathrm{L}}$. Then $z \in \mathrm{Im}(w \cap (x \times \{x_1\} \times \cdots \times \{x_n\})) \leftrightarrow z \in x \wedge \mathbf{A}_{\mathrm{L}}$, so by extensionality the set $\mathrm{Im}(w \cap (x \times \{x_1\} \times \cdots \times \{x_n\}))$ equals $\{z \mid z \in x \wedge \mathbf{A}_{\mathrm{L}}\}$, which is thus constructible. Hence by Lemma 4 of §1.2, the interpretation of each subset axiom holds.

To derive the interpetations of the replacement axioms, suppose that $\mathbf{f}$ is a defined L-invariant $(n+1)$-ary function symbol, and let $y, x_1, \ldots, x_n$ be constructible sets. Then every member of $\mathrm{Un}\{\mathbf{f}xx_1 \ldots x_n \mid x \in y\}$ is constructible, so by Lemma 2 of §2.3 there exists a constructible set z such that $\mathrm{Un}\{\mathbf{f}xx_1 \ldots x_n \mid x \in y\} \subseteq z$. Thus for all $x \in y$, $\mathbf{f}xx_1 \ldots x_n \subseteq z$. By Lemma 5 of §1.2, the interpretation of each replacement axiom holds. ☐

LEMMA 2. $\vdash_{\mathrm{ZF}} \mathrm{Ord}\, x \to \mathrm{L}x$.

Proof. By the results of §1.3 and Lemma 1, Ord is absolute for L. Recall that we have derived $\exists x(\mathrm{Ord}\, x \wedge x \notin y)$ in ZF_ω. So by Lemma 1 its interpretation by L holds, namely $\mathrm{L}y \to \exists x(\mathrm{L}x \wedge \mathrm{Ord}_{\mathrm{L}}\, x \wedge \neg(x \in_{\mathrm{L}} y))$. Since Ord is absolute for L and $\in_{\mathrm{L}}$ is $\in$, we have

$$\vdash_{\mathrm{ZF}} \mathrm{L}y \to \exists x(\mathrm{L}x \wedge \mathrm{Ord}\, x \wedge x \notin y). \tag{1}$$

We now prove $\vdash_{\mathrm{ZF}} \mathrm{Ord}\, x \to \mathrm{L}x$ by transfinite induction. Let x be an ordinal. By induction hypothesis, every member of x is constructible. So there is a constructible set y such that $x \subseteq y$. By (1), there is a constructible ordinal z such that $z \notin y$. Now if $z < x$, then $z \in y$. So $x \leq z$, and hence x is constructible by transitivity of L. ☐

THEOREM. The simple interpretation defined by L is an interpretation of ZF.

Proof. By Lemma 2, $\vdash_{\mathrm{ZF}} \mathrm{L}\omega$. So by Lemma 3 of §1.2, the interpretation of the infinity axiom holds. ☐

LEMMA 3. C is absolute for transitive simple interpretations of ZF_ω.

Proof. C is defined by transfinite recursion so that $\vdash_{\mathrm{ZF}} \mathrm{Ord}\, x \to \mathrm{C}x = \mathbf{g}\{\langle \mathrm{C}y, y\rangle \mid y < x\}x$ where $\mathbf{g}$ is the binary function symbol defined by

$$\begin{aligned} y = \mathbf{g}x'x \leftrightarrow{} & (\mathrm{Ord}\, x \wedge \mathrm{J}_0 x = \dot{0} \wedge \forall w(w \in y \leftrightarrow \exists z(z \in x \wedge w = x'`z))) \\ & \vee (\mathrm{Ord}\, x \wedge \mathrm{J}_0 x = \dot{1} \wedge y = \mathfrak{F}_1 x'`\mathrm{J}_1 x x'`\mathrm{J}_2 x) \vee \cdots \vee (\mathrm{Ord}\, x \wedge \dot{9} < \mathrm{J}_0 x \wedge y = \{x'`\mathrm{J}_1 x, x'`\mathrm{J}_2 x\}) \\ & \vee (\neg\, \mathrm{Ord}\, x \wedge y = \dot{0}). \end{aligned}$$

Since all the symbols in the right-hand side are absolute for transitive simple interpretations of ZF_ω, $\mathbf{g}$, and hence C, is absolute for such interpretations. □

2.5 The axiom of constructibilty. The *axiom of constructibility* is a translation of $\forall x \mathrm{L}x$ into ZF. We denote by ZFL the first-order theory obtained from ZF by the adjunction of the axiom of constructibility. Note that $\vdash_{\mathrm{ZFL}} \mathrm{Ord}\, x \to \exists z \forall y(\mathrm{Od}\, y \leq x \to y \in z)$ and $\vdash_{\mathrm{ZFL}} \mathrm{Od}\, x = \mu y(x = \mathrm{C}y)$. Thus by Lemma 2 of §1.3 and Lemma 3 of §2.4, Od is absolute for transitive simple interpretations of ZFL.

Theorem 1. If ZFL is inconsistent, so is ZF.

Proof. We already know that L is an interpretation of ZF in an extension by definitions of ZF. By the interpretation theorem, it remains to show that the interpretation by L of the axiom of constructibility is a theorem. This interpretation is $\forall x(\mathrm{L}x \to \mathrm{L}_{\mathrm{L}}x)$. Thus it will suffice to prove that L is absolute for L. But L was defined by $\mathrm{L}x \leftrightarrow \exists y(\mathrm{Ord}\, y \wedge x = \mathrm{C}y)$. Because $\vdash_{\mathrm{ZF}} \mathrm{Ord}\, x \to \mathrm{L}x$, $\mathrm{Ord}\, y \wedge x = \mathrm{C}y$ is complete in y for L. Hence L is absolute for L by Lemma 3 of §2.4. □

Theorem 2. The axiom of choice is a theorem of ZFL.

Proof. In ZFL, we define a unary function symbol Inf by $y = \mathrm{Inf}\, x \leftrightarrow \exists z(\mathrm{Ord}\, z \wedge y = \mathrm{C}z \wedge (x = \dot{0} \vee \mathrm{C}z \in x) \wedge \forall w(\mathrm{Ord}\, w \to w < z \to \neg(x = \dot{0} \vee \mathrm{C}w \in x)))$. Thus $\mathrm{Inf}\, x$ is the image by C of the first ordinal z such that $x = \dot{0} \vee \mathrm{C}z \in x$. To prove that this is a valid definition, we must derive $\exists z(\mathrm{Ord}\, z \wedge x = \dot{0} \vee \mathrm{C}z \in x)$. If x is $\dot{0}$, we can choose z to be $\dot{0}$. Otherwise, since $\forall x \mathrm{L}x$, any member of x is constructible, so we can find such a z. If x does not equal $\dot{0}$, then $\mathrm{Inf}\, x \in x$. Thus the set $\{\langle \mathrm{Inf}\, y, y\rangle \mid y \in \mathrm{P}x - \{\dot{0}\}\}$ is a choice function on x. So $\forall x \exists y\, \mathrm{Ch}\, yx$ is derivable in ZFL. □

In a less formal setting, the proof of the above theorem can be summarized as follows: in ZFL, the universe of all sets is well-ordered via Od, and this provides a canonical way of defining a choice function on a set. Note that in a first-order theory for sets in which the notion of class is defined, such as NBG, together with the axiom of constructibility, the above proof shows in fact that there is an explicit (i.e., definable) choice function on the class of all sets.

Since the axiom of choice is a theorem of ZFL, # is a defined symbol of ZFL.

§3 The cardinality theorem

3.1 The reflection principle. The goal of this section is to prove a formal version for the first-order theory ZFC of the famous result of model theory known as the (downward) Löwenheim–Skolem theorem. By a "formal version", we mean of course that both the formulation and the proof of this theorem will be completely finitary. We first prove a weaker result known as the reflection principle (in ZF). In this paragraph, we let T be a good extension of ZF in which there is a constant $\mathbf{e}_0$, and we let Γ consist of finitely many *instantiations* of T.

Reflection Principle. There is a defined constant $\mathbf{e}$ of T such that $\vdash_T \mathbf{e}_0 \subseteq \mathbf{e}$, $\vdash_T \mathrm{Tr}\, \mathbf{e}$, and for every instantiation $\exists \mathbf{y}\mathbf{B}$ of Γ with free variables $\mathbf{x}_1, \ldots, \mathbf{x}_n$, $\vdash_T \mathbf{x}_1 \in \mathbf{e} \to \cdots \to \mathbf{x}_n \in \mathbf{e} \to \exists \mathbf{y}\mathbf{B} \to \exists \mathbf{y}(\mathbf{y} \in \mathbf{e} \wedge \mathbf{B})$.

Proof. Let $\mathbf{A}$ be a formula of Γ of the form $\exists \mathbf{y}\mathbf{B}$ with free variables $\mathbf{x}_1, \ldots, \mathbf{x}_n$ in reverse alphabetical order. Choose $\mathbf{z}$ not free in $\mathbf{B}$ and define in T a function symbol $\mathbf{f}_\mathbf{A}$ by

$$\mathbf{z} = \mathbf{f}_\mathbf{A}\mathbf{x}_1 \ldots \mathbf{x}_n \leftrightarrow (\exists \mathbf{y}(\mathbf{B} \wedge \mathrm{Rk}\, \mathbf{y} = \mathbf{z}) \wedge \forall \mathbf{w}(\mathrm{Rk}\, \mathbf{w} < \mathbf{z} \to \neg \mathbf{B}[\mathbf{y}|\mathbf{w}])) \vee (\neg \mathbf{A} \wedge \mathbf{z} = \dot{0}).$$

Thus $\mathbf{f}_\mathbf{A}\mathbf{x}_1 \ldots \mathbf{x}_n$ denotes the smallest rank of a set $\mathbf{y}$ such that $\mathbf{B}$, if such a set exists. We also define in T a unary function symbol $\mathbf{g}_\mathbf{A}$ by $\mathbf{g}_\mathbf{A}\mathbf{x} = \mathrm{Un}\{\mathbf{f}_\mathbf{A}\mathbf{x}_1 \ldots \mathbf{x}_n \mid \mathbf{x}_1 \in \mathrm{Stg}\, \mathbf{x} \wedge \cdots \wedge \mathbf{x}_n \in \mathrm{Stg}\, \mathbf{x}\}$, for some $\mathbf{x}$ not free in $\mathbf{A}$. This definition implies that if $\mathbf{x}_1, \ldots, \mathbf{x}_n$ belong to $\mathrm{Stg}\, \mathbf{x}$ and if there exists $\mathbf{y}$ such that $\mathbf{B}$, then we can find such a $\mathbf{y}$ in $\mathrm{Stg}\, \mathbf{g}_\mathbf{A}\mathbf{x}$.

Say Γ consists of $\mathbf{A}_1, \ldots, \mathbf{A}_n$. We define $\mathbf{g}$ by $\mathbf{g}x = \mathrm{Un}\{\mathbf{g}_{\mathbf{A}_1}x, \ldots, \mathbf{g}_{\mathbf{A}_n}x, \mathrm{S}x, \mathrm{Rk}\, \mathbf{e}_0\}$. Finally, we define a unary function symbol $\mathbf{h}$ by transfinite recursion so that $\vdash_T \mathrm{Ord}\, x \to \mathbf{h}x = \mathrm{Un}\{\mathbf{g}\mathbf{h}y \mid y < x\}$. Note that $\vdash_T \mathrm{Ord}\, x \to \mathrm{Ord}\, \mathbf{h}x$ and, by definition of $\mathbf{g}$, $\vdash_T \mathrm{Ord}\, x \to y < x \to \mathbf{h}y < \mathbf{h}x$.

We claim that the constant $\mathbf{e}$ defined by $\mathbf{e} = \mathrm{Un}\{\mathrm{Stg}\, \mathbf{h}z \mid z < \omega\}$ has the required properties. Clearly $\vdash_T \mathbf{e}_0 \subseteq \mathbf{e}$ and $\vdash_T \mathrm{Tr}\, \mathbf{e}$. Let $\exists \mathbf{y}\mathbf{B}$ be an instantiation of Γ with free variables $\mathbf{x}_1, \ldots, \mathbf{x}_n$. We derive the second assertion in English, and we assume that $\mathbf{y}, \mathbf{x}_1, \ldots, \mathbf{x}_n$ are $y, x_1, \ldots, x_n$. Suppose that $x_1, \ldots, x_n$ belong to $\mathbf{e}$. Then they belong to $\mathrm{Stg}\, \mathbf{h}z$ for some $z < \omega$. Assume that $\mathbf{A}$ holds. Then by definition of $\mathbf{g}$, there exists $y \in \mathrm{Stg}\, \mathbf{g}\mathbf{h}z$ such that $\mathbf{B}$. By definition of $\mathbf{h}$, $\mathbf{g}\mathbf{h}z \leq \mathbf{h}\mathrm{S}z$, and since $\mathrm{S}z < \omega$, we find $y \in \mathbf{e}$. □

Note that if Δ consists of finitely many formulae of T and if any subformula of a formula in Δ is in Δ, then applying the reflection principle to the instantiations of Δ yields a constant $\mathbf{e}$ with the property that the formulae of Δ are absolute for the transitive simple interpretation $\mathbf{e}$ (by Lemma 7 of ch. II §3.5). It is because of this fact that the reflection principle is thus named.

3.2 The cardinality theorem. In this paragraph, we let Γ be a collection of formulae of $L(\mathrm{ZF})$, and T an extension of ZFC *whose language is* $L(\mathrm{ZF})$. We denote by T_0 the first-order theory obtained from T by the adjunction of a new constant $\mathbf{e}_0$, and as new axioms $\exists x(x \in \mathbf{e}_0)$ and a translation of $\mathrm{Tr}\,\mathbf{e}_0$. Form T_1 from T_0 by adding a new constant $\mathbf{e}_1$.

We let T^Γ be obtained from T_1 by the adjunction of the following nonlogical axioms:

(i) a translation of $\mathbf{e}_0 \subseteq \mathbf{e}_1$;

(ii) a translation of $\mathrm{Ord}\, x \to \#\mathbf{e}_0 \leq \aleph x \to \#\mathbf{e}_1 \leq \aleph x$;

(iii) for every instantiation $\exists\mathbf{y}\mathbf{B}$ of Γ with free variables $\mathbf{x}_1, \ldots, \mathbf{x}_n$ in reverse alphabetical order, $\mathbf{x}_1 \in \mathbf{e}_1 \to \cdots \to \mathbf{x}_n \in \mathbf{e}_1 \to \exists\mathbf{y}\mathbf{B} \to \exists\mathbf{y}(\mathbf{y} \in \mathbf{e}_1 \wedge \mathbf{B})$.

Axioms (i) and (ii) mean that $\mathbf{e}_1$ includes $\mathbf{e}_0$ and that the cardinality of $\mathbf{e}_1$ is no greater than the cardinality of $\mathbf{e}_0$. The formula $\#\mathbf{e}_1 \leq \mathrm{Max}\langle \aleph\dot{0}, \#\mathbf{e}_0\rangle$ is derivable from these two axioms. The axioms of (iii) mean that whenever a formula of Γ holds which has parameters in $\mathbf{e}_1$ and asserts the existence of some set, there must be such a set member of $\mathbf{e}_1$. Note that all of the above extensions are good extensions, since the only nondefined added symbols are constants.

Lemma 1. T^Γ is a conservative extension of T_0.

Proof. Let $\mathbf{C}$ be a formula of T_0 such that $\vdash_{T^\Gamma} \mathbf{C}$. By the reduction theorem, there are formulae $\mathbf{D}_1, \ldots, \mathbf{D}_k$ among the closures of the axioms (i)–(iii) added to form T^Γ such that $\vdash_{T_1} \mathbf{D}_1 \wedge \cdots \wedge \mathbf{D}_k \to \mathbf{C}$. Let $\mathbf{x}$ be a variable not occurring in $\mathbf{C}, \mathbf{D}_1, \ldots, \mathbf{D}_k$, and for each i let $\mathbf{D}'_i$ be obtained from $\mathbf{D}_i$ by replacing each occurrence of $\mathbf{e}_1$ by $\mathbf{x}$. Then by the theorem on constants, $\vdash_{T_0} \mathbf{D}'_1 \wedge \cdots \wedge \mathbf{D}'_k \to \mathbf{C}$, whence $\vdash_{T_0} \exists\mathbf{x}(\mathbf{D}'_1 \wedge \cdots \wedge \mathbf{D}'_k) \to \mathbf{C}$ by the $\exists$-introduction rule. Thus it will suffice to prove

$$\vdash_{T_0} \exists\mathbf{x}(\mathbf{D}'_1 \wedge \cdots \wedge \mathbf{D}'_k). \tag{1}$$

Let $\mathbf{e}$ be obtained by the reflection principle applied to Δ, T_0, and $\mathbf{e}_0$, where Δ consists of the finitely many instantiations of Γ whose associated axioms are among $\mathbf{D}_1, \ldots, \mathbf{D}_k$. Then $\vdash_{T_0} \mathbf{e}_0 \subseteq \mathbf{e}$ and for each $\exists\mathbf{y}\mathbf{B}$ in Δ with free variables $\mathbf{x}_1, \ldots, \mathbf{x}_n$,

$$\vdash_{T_0} \mathbf{x}_1 \in \mathbf{e} \to \cdots \to \mathbf{x}_n \in \mathbf{e} \to \exists\mathbf{y}\mathbf{B} \to \exists\mathbf{y}(\mathbf{y} \in \mathbf{e} \wedge \mathbf{B}). \tag{2}$$

Choose variables $\mathbf{z}$ and $\mathbf{w}$ not occurring in any formula of Δ and distinct from x. For $\mathbf{A}$ a formula of Δ of the form $\exists\mathbf{y}\mathbf{B}$ with free variables $\mathbf{x}_1, \ldots, \mathbf{x}_n$, we define an $(n+1)$-ary function symbol $\mathrm{Sk}_\mathbf{A}$ by

$$\mathbf{y} = \mathrm{Sk}_\mathbf{A}\mathbf{x}_1 \ldots \mathbf{x}_n\mathbf{w} \leftrightarrow (\mathbf{A} \wedge \mathrm{Ch}\,\mathbf{w}\mathbf{e} \wedge \mathbf{y} = \mathbf{w}`\{\mathbf{z} \mid \mathbf{z} \in \mathbf{e} \wedge \mathbf{B}[\mathbf{y}|\mathbf{z}]\}) \vee (\neg(\mathbf{A} \wedge \mathrm{Ch}\,\mathbf{w}\mathbf{e}) \wedge \mathbf{y} = \dot{0}).$$

By the functional closure theorem, we can define in T_0 a unary function symbol $\mathbf{h}$ such that

$$\vdash_{T_0} \mathbf{e}_0 \subseteq \mathbf{h}\mathbf{w}, \tag{3}$$

$$\vdash_{T_0} \mathrm{Ord}\, x \to \#\mathbf{e}_0 \leq \aleph x \to \#\mathbf{h}\mathbf{w} \leq \aleph x, \text{ and} \tag{4}$$

$$\vdash_{T_0} \mathbf{x}_1 \in \mathbf{h}\mathbf{w} \to \cdots \to \mathbf{x}_n \in \mathbf{h}\mathbf{w} \to \mathrm{Sk}_\mathbf{A}\mathbf{x}_1 \ldots \mathbf{x}_n\mathbf{w} \in \mathbf{h}\mathbf{w} \tag{5}$$

for every $\mathbf{A}$ in Δ with free variables $\mathbf{x}_1, \ldots, \mathbf{x}_n$. By the definition of $\mathrm{Sk}_\mathbf{A}$ and the fact that $\vdash_{T_0} \dot{0} \in \mathbf{e}$, we clearly have $\vdash_{T_0} \mathbf{x}_1 \in \mathbf{e} \to \cdots \to \mathbf{x}_n \in \mathbf{e} \to \mathrm{Sk}_\mathbf{A}\mathbf{x}_1 \ldots \mathbf{x}_n\mathbf{w} \in \mathbf{e}$ for such $\mathbf{A}$, so by (iv) of the functional closure theorem, $\vdash_{T_0} \mathbf{h}\mathbf{w} \subseteq \mathbf{e}$. Thus by (2), $\vdash_{T_0} \mathbf{x}_1 \in \mathbf{h}\mathbf{w} \to \cdots \to \mathbf{x}_n \in \mathbf{h}\mathbf{w} \to \mathbf{A} \to \{\mathbf{z} \mid \mathbf{z} \in \mathbf{e} \wedge \mathbf{B}[\mathbf{y}|\mathbf{z}]\} \neq \dot{0}$, so by the definition of $\mathrm{Sk}_\mathbf{A}$,

$$\vdash_{T_0} \mathrm{Ch}\,\mathbf{w}\mathbf{e} \to \mathbf{x}_1 \in \mathbf{h}\mathbf{w} \to \cdots \to \mathbf{x}_n \in \mathbf{h}\mathbf{w} \to \mathbf{A} \to \mathbf{B}[\mathbf{y}|\mathrm{Sk}_\mathbf{A}\mathbf{x}_1 \ldots \mathbf{x}_n\mathbf{w}]. \tag{6}$$

By (5), (6), and the substitution axioms,

$$\vdash_{T_0} \mathrm{Ch}\,\mathbf{w}\mathbf{e} \to \mathbf{x}_1 \in \mathbf{h}\mathbf{w} \to \cdots \to \mathbf{x}_n \in \mathbf{h}\mathbf{w} \to \mathbf{A} \to \exists\mathbf{y}(\mathbf{y} \in \mathbf{h}\mathbf{w} \wedge \mathbf{B}). \tag{7}$$

Using (3), (4), and (7) with the closure theorem and the substitution axioms, we obtain $\vdash_{T_0} \mathrm{Ch}\,\mathbf{w}\mathbf{e} \to \exists\mathbf{x}(\mathbf{D}'_1 \wedge \cdots \wedge \mathbf{D}'_k)$. Since $\vdash_{T_0} \exists\mathbf{w}\,\mathrm{Ch}\,\mathbf{w}\mathbf{e}$ by the axiom of choice and the substitution rule, we find (1) using the $\exists$-introduction rule and the detachment rule. □

Remark. By the functional extension theorem, we would have obtained a conservative extension of T_0 by simply adding for each $\mathbf{A}$ in Γ of the form $\exists\mathbf{yB}$ with free variables $\mathbf{x}_1, \ldots, \mathbf{x}_n$ the n-ary function symbol $\mathrm{Sk}_{\mathbf{A}}$ and the axiom $\mathbf{A} \to \mathbf{B}[\mathbf{y}|\mathrm{Sk}_{\mathbf{A}}\mathbf{x}_1 \ldots \mathbf{x}_n]$. An application of the functional closure theorem on finitely many $\mathrm{Sk}_{\mathbf{A}}$ would then yield at once Lemma 1 without using the reflection principle. Unfortunately, such a simplification does not work, because the adjunction of the symbols $\mathrm{Sk}_{\mathbf{A}}$ as above does not yield a good extension of T. In fact, we can precisely see that some replacement axioms featuring the new symbols which must be used in the proof of the functional closure theorem fail to be derivable. Thus the more complicated proofs above were really necessary, as was the direct use of the axiom of choice in the proof of Lemma 1. Note however that a much simpler proof of Lemma 1 is possible if the axiom of constructibilty holds in T, for then we can define $\mathrm{Sk}_{\mathbf{A}}\mathbf{x}_1 \ldots \mathbf{x}_n$ to be the set of smallest order such that $\mathbf{B}$ whenever such a set exists, and apply the functional closure theorem.

Lemma 2. Suppose that any subformula of a formula in Γ is in Γ. Then the formulae of Γ are absolute for the simple interpretation $\mathbf{e}_1$ in T^Γ. Explicitely, if $\mathbf{A}$ is in Γ and $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are the variables free in $\mathbf{A}$, then $\vdash_{T^\Gamma} \mathbf{x}_1 \in \mathbf{e}_1 \to \cdots \to \mathbf{x}_n \in \mathbf{e}_1 \to \mathbf{A} \leftrightarrow \mathbf{A}_{\mathbf{e}_1}$.

Proof. This follows from the axioms of (iii) and Lemma 7 of ch. II §3.5. □

Cardinality Theorem. Suppose that all the formulae of Γ are closed. Then in some good conservative extension T' of T_0, there is a constant $\mathbf{e}$ such that

(i) $\vdash_{T'} \mathbf{e}_0 \subseteq \mathbf{e}$;
(ii) $\vdash_{T'} \mathrm{Tr}\,\mathbf{e}$;
(iii) $\vdash_{T'} \mathrm{Ord}\, x \to \#\mathbf{e}_0 \leq \aleph x \to \#\mathbf{e} \leq \aleph x$;
(iv) $\vdash_{T'} \mathbf{A} \leftrightarrow \mathbf{A}_{\mathbf{e}}$ for every formula $\mathbf{A}$ of Γ.

Proof. We let Δ consist of all the subformulae of the formulae of Γ and all the subformulae of the extensionality axiom. We shall build T' as an extension by definitions of T^Δ. So by Lemma 1, T' will be a conservative extension of T_0. It remains to define $\mathbf{e}$ suitably. Since by Lemma 2 the extensionality axiom is absolute for $\mathbf{e}_1$ in T^Δ, it follows that the interpretation by $\mathbf{e}_1$ of the extensionality axiom is a theorem of T^Δ. Recall that the ordinal function symbol Rk satisfies $\vdash_{\mathrm{ZF}} \mathrm{Ord}\, x \to \exists y \forall z(\mathrm{Rk}\, z \leq x \to z \in y)$ and $\vdash_{\mathrm{ZF}} x \in y \to \mathrm{Rk}\, x < \mathrm{Rk}\, y$. We can then apply the Mostowski collapsing theorem: the function symbol $\mathbf{g}$ defined by $\mathbf{g}x = \{\mathbf{g}w \mid w \in \mathbf{e}_1 \wedge w \in x\}$ is an isomorphism from the interpretation $\mathbf{e}_1$ to the transitive simple interpretation $\mathbf{e}$ where $\mathbf{e}$ is defined by $\mathbf{e} = \{\mathbf{g}y \mid y \in \mathbf{e}_1\}$. Because $\vdash_{T'} x \in \mathbf{e}_1 \to y \in \mathbf{e}_1 \to x = y \leftrightarrow \mathbf{g}x = \mathbf{g}y$, we see that $\{\langle \mathbf{g}y, y\rangle \mid y \in \mathbf{e}_1\}$ is an injective function whose image is $\mathbf{e}$, so $\#\mathbf{e} = \#\mathbf{e}_1$. It remains to prove (i) and (iv). Because the formulae of Γ are closed, $\vdash_{T'} \mathbf{A}_{\mathbf{e}_1} \leftrightarrow \mathbf{A}_{\mathbf{e}}$ for all $\mathbf{A}$ in Γ by the isomorphism extension theorem, and since by Lemma 2 $\vdash_{T'} \mathbf{A} \leftrightarrow \mathbf{A}_{\mathbf{e}_1}$, we obtain $\vdash_{T'} \mathbf{A} \leftrightarrow \mathbf{A}_{\mathbf{e}}$. This proves (iv). To prove (i), it will suffice to prove that $\vdash_{T'} x \in \mathbf{e}_0 \to \mathbf{g}x = x$. We proceed by transfinite induction on $\mathrm{Rk}\, x$. Suppose $x \in \mathbf{e}_0$. If $y \in x$, then $\mathrm{Rk}\, y < \mathrm{Rk}\, x$ and by transitivity of $\mathbf{e}_0$, $y \in \mathbf{e}_0$. So we may apply the induction hypothesis to y, which yields $\mathbf{g}y = y$. Now by definition of $\mathbf{g}$, $\mathbf{g}x = \{\mathbf{g}y \mid y \in \mathbf{e}_1 \wedge y \in x\} = \{y \mid y \in \mathbf{e}_1 \wedge y \in x\}$. Since $y \in x$ implies $y \in \mathbf{e}_0$ and hence $y \in \mathbf{e}_1$, $\mathbf{g}x = \{y \mid y \in x\} = x$. □

A particular case of the cardinality theorem is when Γ contains the closures of the axioms of T. Then $\mathbf{e}$ is a transitive simple interpretation of T in T'. For let $\mathbf{A}$ be a nonlogical axiom of T and $\mathbf{B}$ its closure. Then by (iv) and the closure theorem, $\vdash_{T'} \mathbf{B}_{\mathbf{e}}$, so by prenex operations and the closure theorem, $\vdash_{T'} \mathbf{A}^{\mathbf{e}}$.

3.3 The countable interpretation. We obtain an important corollary to the cardinality theorem. We let Γ, T, and T_0 be as in §3.2.

Corollary. Suppose that the formulae of Γ are closed. Let U be obtained from T by the adjunction of a constant $\mathbf{e}$ and the following axioms:

(i) $\exists x(x \in \mathbf{e})$;
(ii) a translation of $\mathrm{Tr}\,\mathbf{e}$;
(iii) a translation of $\#\mathbf{e} \leq \aleph\dot{0}$;
(iv) $\mathbf{A} \leftrightarrow \mathbf{A}_{\mathbf{e}}$ for every formula $\mathbf{A}$ of Γ.

Then U is a conservative extension of T.

Proof. Let T' be as in the cardinality theorem. Form T'' from T' by the adjunction of the axiom $\mathbf{e}_0 = \dot{1}$. Then clearly (i)–(iv) are theorems of T''. This means that T'' is an extension of U. Let $\mathbf{A}$ be a formula of T such that $\vdash_U \mathbf{A}$. Then $\vdash_{T''} \mathbf{A}$. By the deduction theorem, $\vdash_{T'} \mathbf{e}_0 = \dot{1} \to \mathbf{A}$. Since T' is a conservative extension of T_0, $\vdash_{T_0} \mathbf{e}_0 = \dot{1} \to \mathbf{A}$. Hence by the deduction theorem, $\vdash_T \exists x(x \in \mathbf{x}) \to \mathrm{Tr}\,\mathbf{x} \to \mathbf{x} = \dot{1} \to \mathbf{A}$ for some $\mathbf{x}$ not free in $\mathbf{A}$ and distinct from x. By the substitution rule, $\vdash_T \exists x(x \in \dot{1}) \to \mathrm{Tr}\,\dot{1} \to \dot{1} = \dot{1} \to \mathbf{A}$, so $\vdash_T \mathbf{A}$. □

If Γ contains the closures of the axioms of T, we find as in §3.2 that $\mathbf{e}$ is a transitive simple interpretation of T in its conservative extension U, and $\vdash_U \mathbf{e} \leq \aleph\dot{0}$.

§4 The generalized continuum hypothesis

We now have the necessary tools to prove that the generalized continuum hypothesis is derivable in ZF from the axiom of constructibility.

Lemma 1. $\vdash_{\mathrm{ZFL}} \mathrm{Ord}\,x \to \#\mathrm{C}^*\aleph x = \aleph x$.

Proof. Let x be an ordinal. The set $\{\langle \mathrm{C}y, y\rangle \mid y < \aleph x\}$ is a function on $\aleph x$ whose image is $\mathrm{C}^*\aleph x$. Hence $\#\mathrm{C}^*\aleph x \leq \#\aleph x$, so $\#\mathrm{C}^*\aleph x \leq \aleph x$. Thus it remains to derive the existence of an injective function from $\aleph x$ to $\mathrm{C}^*\aleph x$. Let $\mathbf{f}$ be defined by $z = \mathbf{f}y \leftrightarrow (\mathrm{Ord}\,y \wedge \mathrm{O}^2 z = \langle \dot{0}, y\rangle) \vee (\neg\,\mathrm{Ord}\,y \wedge z = \dot{0})$, and let $\mathbf{a}$ be $\{\langle \mathrm{C}^*\mathbf{f}y, y\rangle \mid y < \aleph x\}$. This is a valid definition from the results of ch. VI §3.5. Note that $\mathrm{C}^*\mathbf{f}y = \mathrm{C}\mathbf{f}y$ by the definition of C. If $y < \aleph x$, then $y < \mathrm{Max}\,\mathrm{O}^2\aleph x$ by (x) of ch. VI §3.11, so $\mathrm{Max}\,\mathrm{O}^2\mathbf{f}y = y < \mathrm{Max}\,\mathrm{O}^2\aleph x$. Hence $\mathbf{f}y < \aleph x$ by (vii) of ch. VI §3.5. This proves that $\mathrm{Im}\,\mathbf{a} \subseteq \mathrm{C}^*\aleph x$. Let y and y' be ordinals such that $y' < y$ and $y < \aleph x$. Then $\mathbf{f}y \neq \mathbf{f}y'$. So either $\mathbf{f}y < \mathbf{f}y'$ or $\mathbf{f}y' < \mathbf{f}y$. In the first case, $\mathbf{a}`y = \mathrm{C}^*\mathbf{f}y = \mathrm{C}\mathbf{f}y \in \mathrm{C}^*\mathbf{f}y' = \mathbf{a}`y'$. In the second case, we find similarly $\mathbf{a}`y' \in \mathbf{a}`y$. But in both cases, $\mathbf{a}`y \neq \mathbf{a}`y'$ by ch. VI §3.1 (iii), so $\mathbf{a}$ is injective. □

Lemma 2. $\vdash_{\mathrm{ZFL}} \mathrm{Ord}\,x \to \mathrm{P}\mathrm{C}^*\aleph x \subseteq \mathrm{C}^*\aleph\mathrm{S}x$.

Proof. Let Γ consists of the closures of the axioms of ZFL. Let T_0, T', $\mathbf{e}_0$, $\mathbf{e}$ be as in the cardinality theorem when T is ZFL. Then $\mathbf{e}$ is a transitive simple interpretation of ZFL in T'. Since Od is absolute for any transitive simple interpretation of ZFL (cf. §2.5), Od is absolute for $\mathbf{e}$. In particular, Od is $\mathbf{e}$-invariant, so $\vdash_{T'} y \in \mathbf{e} \to \mathrm{Od}\,y \in \mathbf{e}$. By transitivity of $\mathbf{e}$, $\vdash_{T'} y \in \mathbf{e} \to \mathrm{Od}\,y \subseteq \mathbf{e}$, so $\vdash_{T'} y \in \mathbf{e} \to \#\,\mathrm{Od}\,y \leq \#\mathbf{e}$. Using (i) and (iii) of the cardinality theorem, we find

$$\vdash_{T'} \mathrm{Ord}\,x \to \#\mathbf{e}_0 \leq \aleph x \to y \in \mathbf{e}_0 \to \#\,\mathrm{Od}\,y \leq \aleph x. \tag{1}$$

Let x be an ordinal such that $\#\mathbf{e}_0 \leq \aleph x$ and let $y \in \mathbf{e}_0$. Suppose that $y \notin \mathrm{C}^*\aleph\mathrm{S}x$. Then $\aleph\mathrm{S}x \leq \mathrm{Od}\,y$, so $\aleph x < \#\,\mathrm{Od}\,y$, but this contradicts (1). Hence $y \in \mathrm{C}^*\aleph\mathrm{S}x$. In summary, we have $\vdash_{T'} \mathrm{Ord}\,x \to \#\mathbf{e}_0 \leq \aleph x \to \mathbf{e}_0 \subseteq \mathrm{C}^*\aleph\mathrm{S}x$. Since T' is a conservative extension of T_0, this is a theorem of T_0 as well. Thus by the deduction theorem,

$$\vdash_{\mathrm{ZFL}} \exists x(x \in y) \to \mathrm{Tr}\,y \to \mathrm{Ord}\,x \to \#y \leq \aleph x \to y \subseteq \mathrm{C}^*\aleph\mathrm{S}x. \tag{2}$$

Now let x be an ordinal and let y belong to $\mathrm{P}\mathrm{C}^*\aleph x$, i.e., $y \subseteq \mathrm{C}^*\aleph x$. Let $\mathbf{a}$ be $\mathrm{C}^*\aleph x \cup \{y\}$. By Lemma 1 of §2.3, $\mathrm{C}^*\aleph x$ is transitive. Since $y \subseteq \mathrm{C}^*\aleph x$, $\mathbf{a}$ is transitive. By Lemma 1, $\#\mathbf{a} \leq \#\mathrm{C}^*\aleph x \oplus \#\{y\} = \aleph x \oplus \dot{1} = \aleph x$. Thus $\mathbf{a}$ is a nonempty transitive set of cardinal at most $\aleph x$, so by (2), $\mathbf{a} \subseteq \mathrm{C}^*\aleph\mathrm{S}x$. In particular, $y \in \mathrm{C}^*\aleph\mathrm{S}x$, which proves the lemma. □

Theorem. The generalized continuum hypothesis is a theorem of ZFL.

Proof. By (x) of ch. VI §3.10 and Lemma 1, $\vdash_{\mathrm{ZFL}} \mathrm{Ord}\,x \to \#\mathrm{P}\mathrm{C}^*\aleph x = \#\mathrm{P}\aleph x$. Then by Lemma 2, $\vdash_{\mathrm{ZFL}} \mathrm{Ord}\,x \to \#\mathrm{P}\aleph x \leq \#\mathrm{C}^*\aleph\mathrm{S}x$, and again by Lemma 1, $\vdash_{\mathrm{ZFL}} \mathrm{Ord}\,x \to \#\mathrm{P}\aleph x \leq \aleph\mathrm{S}x$. Taking the closure, we find the generalized continuum hypothesis. □

Bibliography

[1] G. Boolos, *The Logic of Provability*, Cambridge University Press, 1993

[2] P. J. Cohen, *Set theory and the continuum hypothesis*, Benjamin, New York, 1966

[3] S. Feferman, *Arithmetization of metamathematics in a general setting*, Fundamenta Mathematicae, **49** (1960), 35–92

[4] K. Gödel, *The consistency of the axiom of choice and of the generalized continuum hypothesis with the axioms of set theory*, Princeton University Press, Princeton NJ, 1970

[5] E. Mendelson, *Introduction to Mathematical Logic*, Third Edition, Pacific Grove CA, 1987

[6] J. R. Shoenfield, *Mathematical Logic*, A. K. Peters, Natick MA, 2001

Index