0. Introduction

The theory of hereditarily finite sets, called here HF, and described in the Appendix, has been investigated by Ackermann, Beth, Givant and Tarski ([A], [Be] and [TG]). HF is very closely related to PA (Peano Arithmetic); in fact, they are definitionally equivalent (they have a common extension by definitions $\mathcal{T}$—see Section 10). Thus it appears that the general mathematical significance of establishing Gödel's incompleteness theorems with reference to one of these theories is the same as for the other.

Unlike most other authors working in this field, we have chosen HF, rather than PA, for this task. This was done because HF is much better suited for the purpose of describing its own meta-theory. The reason lies in the great expressive power of set theory. A set-theoretical description of the language of HF (and then of the whole meta-theory) presents itself in a wholly natural fashion, once it has been decided how to code (represent) by constant terms the 7 basic (primitive) symbols and the variables of the first-order language of HF. It is at hand to code the variables $x_1, x_2, \ldots$ simply by the ordinals $1, 2, \ldots$ The constant 0 can be coded as 0, and the remaining 6 symbols as n-tuples of 0's, say $\in$ as $\langle 0, 0\rangle$, etc. And here ends the arbitrariness of coding, which is so unpleasant when languages are arithmetized. Because terms, formulas and proofs are built from these already coded basic symbols and variables by the formation of n-tuples or sequences and such processes can be faithfully and naturally replicated by set-theoretical constructions. By contrast, if natural (Gödel) numbers are used as codes, there is no natural way of coding a sequence of these (numbers) as one number. (Mostly this is done by means of the Chinese Remainder Theorem or by using the uniqueness of prime power decomposition.) The necessity of coding proofs appropriately was pointed out by S. Feferman; he showed that for a less than natural coding, Gödel's results will fail (Theorems 5.9, 4.10 in [F]). We need not face such a problem here: A sequence of codes of formulas that represents a proof, i.e., a sequence of finite sets, is a finite set in its own right and there is no need to code it again!

The simplicity (and possibly elegance) of working towards Gödel's results (6.5 and 9.8) in the realm of finite sets offers one further advantage: With marginally more effort, one can present all arguments completely, without omissions. As far as we know, such a degree of completeness in proving Gödel's results has never been attained before. To reiterate: All published proofs of Gödel's incompleteness theorems contain gaps, omissions or references to key results in other publications. The omissions might not be gross, but nevertheless presenting a potential stumbling block for the non-specialist. Thus, we have made our exposition very detailed, hoping to make it thereby accessible to the widest possible readership. Gödel's theorems certainly deserve this sort of "popularization"; they have generated much interest, even among non-mathematicians, and shed much light on our understanding (or non-understanding) of the real nature of the game of mathematics.

It has been known since long that the more difficult (second) incompleteness theorem is easily deduced, once one has established for each formula α the theorem (Hilbert–Bernays derivabiltity condition)

$$\mathrm{Pf}(\ulcorner\alpha\urcorner) \to \mathrm{Pf}(\ulcorner\mathrm{Pf}(\ulcorner\alpha\urcorner)\urcorner),$$

where $\ulcorner\alpha\urcorner$ is the code of α and $\mathrm{Pf}(x)$ is the formula which states that x is the code of a provable formula. To show this, we followed in broad outline the work of Boolos [Bo]. The latter pertained to PA and as such it had to be, in the words of its author, "incomplete and irremediably messy" (p. 16 in [Bo]).

We have added to the main body of the paper Sections 10 and 11: The first deals with the relationship between PA and HF and in the second we address what might be called the "real world unprovability" of a theorem.

The author would like to thank Jan Mycielski for suggesting this line of research, to Don Monk for a detailed verification of correctness (and pointing out inaccuracies), and also to the other participants of the so-called Ulam seminar at the University of Colorado in Boulder for listening to the first presentation of this work.

1. The language of HF and its coding

The theory of finite sets, i.e. HF, is presented in the language with the logical constants $=, \vee, \neg, \exists$, the variables $x_1, x_2, \ldots$ and the non-logical constants 0, $\in$ (binary relation) and $\triangleleft$ (binary operation). (We shall treat $\wedge, \to$ and $\forall$ as defined symbols.)

The following axiomatization of HF was taken from A. Tarski and S. Givant [TG] (though we modified it slightly). There are two axioms and one scheme. The first axiom defines 0 as the set deprived of elements, the second axiom defines $\triangleleft$ as the operation of adjoining an element to a set, and the axiom scheme describes proofs by induction, very much like in the case of PA.

HF1. $z = 0 \leftrightarrow \forall x(x \notin z)$.

HF2. $z = x \triangleleft y \leftrightarrow \forall u(u \in z \leftrightarrow u \in x \vee u = y)$.

HF3. $\alpha(0) \wedge \forall x \forall y[\alpha(x) \wedge \alpha(y) \to \alpha(x \triangleleft y)] \to \forall x \alpha(x)$.

The assumption in HF3 is that α is any formula in the first-order language based on 0, $\in$ and $\triangleleft$, moreover α contains a freely occurring variable z such that x and y are substitutable for z, whilst $\alpha(x)$, $\alpha(y)$ and $\alpha(x \triangleleft y)$ denote the effects of substitutions. (We shall use lower case Latin letters, $k, l, m, n, r, s, t, u, v, w, x, y, z$, mostly to denote any conveniently chosen variables; e.g., in the above axioms we can think of x, y, z as x_1, x_2, x_3.)

A detailed development of the theory, to the extent that is needed for proving the incompleteness theorems, is given in the Appendix (references to it will be prefixed with Ap). To point out some differences between HF and the Zermelo–Fränkel set theory ZF, let us note that in HF we have *terms*, due to the presence of the binary function symbol $\triangleleft$, whereas in ZF there are none. We shall also use the fact that the universe of finite sets is well ordered by an HF-definable relation $<$ (Ap.5). No such ordering can be defined in ZF.

The only way to produce terms in HF is by repeated applications of $\lhd$. We can read $x \lhd y$ as "x eats y" since this set arises from x by adopting a (new, or not) element y. Of course, x can eat itself; the result is $x \lhd x$ and this is the *successor* of x, denoted by x^+. The first few ordinals are then $0,\ 0^+ = 1,\ 1^+ = 2,\ \ldots$ etc. (Ap.2).

All the *constant terms* of HF are obtained from terms with variables by replacing in the latter all the variables by 0. The family of constant terms will be denoted by $\mathbb{C}$. The elements of $\mathbb{C}$ will be used as codes, and also for the construction of the standard model $\mathfrak{S}$ of HF (Ap.6).

We call $0 \lhd x$ a *singleton* and denote it by $\{x\}$. Similarly, $(0 \lhd x) \lhd y$ will be abbreviated as $\{x, y\}$ and $((0 \lhd x) \lhd y) \lhd z$ as $\{x, y, z\}$. Thus $\{x\}, \{x, y\}, \{x, y, z\}$ *are terms.* The *ordered pair* $\langle x, y\rangle = \{\{x\}, \{x, y\}\}$, *triple* $\langle x, y, z\rangle = \langle x, \langle y, z\rangle\rangle$, and, in general, any *n-tuple* are composites of terms and hence these are terms themselves.

We wish to apply now the theory HF to describe its own meta-theory, i.e., various relations between the words of the language of HF. For this purpose we assign to the main objects of this language (variables, logical and non-logical symbols, terms and formulas) their *codes.* What kind of entity should a code be? Since we wish to apply set-theoretical formulas to codes in order to describe the meta-theory, it appears that a code should be a set, or at least an entity closely and naturally associated to a set. However in the proof that follows we have to assign to the code $\ulcorner\varphi\urcorner$ of a formula φ the code of this code, i.e., $\ulcorner(\ulcorner\varphi\urcorner)\urcorner$ (this is accomplished by the function H in the proof of Gödel's Diagonal Lemma in Section 6). Since only language objects are coded, $\ulcorner\varphi\urcorner$ must be one of these. The simplest language objects naturally associated to sets are the constant terms (associated to the sets they name). Thus constant terms will be selected here as codes. But there is a restriction. Two constant terms could be just different names for the same set, like $\{\{0\}, 0\}$ and $\{0, \{0\}\}$. To avoid this situation we shall select codes from a family Γ of constant terms such that for any two distinct terms σ,τ in Γ one has $\vdash \sigma \neq \tau$.

Definition 1.1. We denote by Γ_0 the least family of constant terms such that

(a) 0 is in Γ_0.

(b) If σ is in Γ_0 then so is $\sigma \lhd \sigma$.

Further, we denote by Γ the least family of terms satisfying $\Gamma_0 \subseteq \Gamma$ and such that

(c) If σ,τ are in Γ then so is the term $\langle\sigma, \tau\rangle$.

Theorem 1.2. *For any two distinct* (*as strings of symbols*) *terms* σ, τ *in* Γ *one has* $\vdash \sigma \neq \tau$.

Proof. Let us say that two constant terms σ, τ *are identified in* HF if $\sigma = \tau$ is a theorem. By Ap.6.4, exactly one of the formulas $\sigma = \tau$, $\sigma \neq \tau$ is a theorem. So our task is to prove that distinct terms in Γ are never identified in HF. Let σ,τ be distinct. We consider three cases:

Case 1: Both σ, τ are in Γ_0. It is easy to see that the terms belonging to Γ_0 are names of ordinal numbers, moreover $\sigma \lhd \sigma$ is a name of the successor of the number named by σ (Ap.2.2 and 2.3(a)). Thus, if τ is longer than σ then the ordinal number named by τ is obtained from the one named by σ by several applications of the successor operation, which means that it is larger than the ordinal named by σ. It follows that σ, τ are not identified in HF.

Case 2: Exactly one of σ, τ is in Γ_0. Suppose σ is in Γ_0. Since σ names an ordinal, either σ is 0 or $\vdash 0 \in \sigma$ (see Ap.2.5). On the other hand, τ is an ordered pair. But it is easily proved that an ordered pair (i.e., $\{\{x\}, \{x, y\}\}$) is never the empty set nor does it contain 0 among its elements. Thus σ, τ are not identified in HF.

Case 3: Both σ, τ are not in Γ_0. We prove by induction on the length of σ the following implication:

If σ, τ are distinct terms in $\Gamma \setminus \Gamma_0$ then σ, τ are not identified in HF.

The inductive assumption is that the above implication holds if σ, τ are replaced by σ', τ', where σ' is shorter than σ. Let σ, τ be distinct terms in $\Gamma \setminus \Gamma_0$. Suppose σ is $\langle \sigma', \sigma'' \rangle$ and τ is $\langle \tau', \tau'' \rangle$. Clearly, σ' is not the same term as τ', or σ'' is not the same term as τ''. We assume the first possibility (if the second holds, the proof is analogous). If σ' or τ' is in Γ_0 then one of the Cases 1, 2 applies and we see that σ' and τ' are not identified in HF. If both σ', τ' are not in Γ_0 then the inductive hypothesis yields that σ' and τ' are not identified in HF. It follows that $\langle \sigma', \sigma'' \rangle = \langle \tau', \tau'' \rangle$ is not a theorem and thus σ, τ are not are not identified in HF. ∎

We begin coding by assigning to each variable x_k and each of the basic 7 language symbols $0, \in, \lhd, =, \vee, \neg, \exists$ their codes, denoted by $\ulcorner x_k \urcorner, \ulcorner 0 \urcorner, \ulcorner \in \urcorner, \ulcorner \lhd \urcorner, \ulcorner = \urcorner, \ulcorner \vee \urcorner, \ulcorner \neg \urcorner, \ulcorner \exists \urcorner$. These are the following terms belonging to Γ:

$$\begin{aligned}
&\ulcorner 0 \urcorner = 0,\\
&\ulcorner \in \urcorner = \langle 0, 0 \rangle,\\
&\ulcorner \lhd \urcorner = \langle 0, 0, 0 \rangle,\\
&\ulcorner = \urcorner = \langle 0, 0, 0, 0 \rangle,\\
&\ulcorner \vee \urcorner = \langle 0, 0, 0, 0, 0 \rangle,\\
&\ulcorner \neg \urcorner = \langle 0, 0, 0, 0, 0, 0 \rangle,\\
&\ulcorner \exists \urcorner = \langle 0, 0, 0, 0, 0, 0, 0 \rangle,\\
&\ulcorner x_1 \urcorner = 0 \lhd 0,\ \ulcorner x_2 \urcorner = \ulcorner x_1 \urcorner \lhd \ulcorner x_1 \urcorner,\ \text{and in general: } \ulcorner x_{k+} \urcorner = \ulcorner x_k \urcorner \lhd \ulcorner x_k \urcorner.
\end{aligned}$$

Thus the code $\ulcorner x_k \urcorner$ of x_k is a term representing the ordinal number k (see Ap.2.3(a)). The coding of terms and formulas now proceeds as follows:

Terms: The terms 0 and x_k have been coded already. Other terms are coded so that if μ and τ have been coded, then

$$\ulcorner \mu \lhd \tau \urcorner = \langle \ulcorner \lhd \urcorner, \ulcorner \mu \urcorner, \ulcorner \tau \urcorner \rangle.$$

Atomic formulas: The codes of atomic formulas (for terms μ, τ) are

$$\ulcorner \mu = \tau \urcorner = \langle \ulcorner = \urcorner, \ulcorner \mu \urcorner, \ulcorner \tau \urcorner \rangle \quad \text{and} \quad \ulcorner \mu \in \tau \urcorner = \langle \ulcorner \in \urcorner, \ulcorner \mu \urcorner, \ulcorner \tau \urcorner \rangle.$$

Non-atomic formulas: For non-atomic formulas, if α, β are formulas already coded, we put

$$\begin{aligned} \ulcorner \neg \alpha \urcorner &= \langle \ulcorner \neg \urcorner, \ulcorner \alpha \urcorner \rangle, \\ \ulcorner \alpha \vee \beta \urcorner &= \langle \ulcorner \vee \urcorner, \ulcorner \alpha \urcorner, \ulcorner \beta \urcorner \rangle, \\ \ulcorner \exists x_k \alpha \urcorner &= \langle \ulcorner \exists \urcorner, \ulcorner x_k \urcorner, \ulcorner \alpha \urcorner \rangle. \end{aligned}$$

2. Σ-formulas

We wish now to describe the property of a formula φ being provable by a set-theoretical condition applicable to the code $\ulcorner \varphi \urcorner$. This means that we should find a formula $\mathrm{Pf}(x)$ of HF such that

$$\vdash \varphi \quad \text{iff} \quad \vdash \mathrm{Pf}(\ulcorner \varphi \urcorner)$$

for any formula φ. Let us call the above the *proof formalization condition.* The description of the proof process not being unique, there are many candidates for $\mathrm{Pf}(x)$. After having selected $\mathrm{Pf}(x)$, in order to establish the proof formalization condition, we shall invoke a model $\mathfrak{S}$ of HF which we call the standard model. It will be not too hard then to argue convincingly that φ is a theorem iff $\mathrm{Pf}(\ulcorner \varphi \urcorner)$ is true in $\mathfrak{S}$, i.e.,

$$\vdash \varphi \quad \text{iff} \quad \mathfrak{S} \models \mathrm{Pf}(\ulcorner \varphi \urcorner).$$

However not every sentence true in $\mathfrak{S}$ is provable in HF (for example, the sentence δ in Gödel's First Incompleteness Theorem 6.5). Fortunately there is a large class of sentences for which truth in $\mathfrak{S}$ is equivalent to provability in HF. These are the Σ-sentences introduced in this section. Hence we need to choose $\mathrm{Pf}(x)$ so that $\mathrm{Pf}(\ulcorner \varphi \urcorner)$ is a Σ-sentence.

We first introduce the strict Σ-formulas, i.e., formulas that involve only bounded universal quantifiers, the variables and the symbols $\in, \vee, \wedge, \exists$.

Definition 2.1. The class $\mathbf{\Sigma}$ of *strict Σ-formulas* is the smallest class of formulas such that:

(1) The atomic formulas $x_i \in x_j$ belong to $\mathbf{\Sigma}$ for all variables x_i, x_j.
(2) If α, β are in $\mathbf{\Sigma}$, then so are $\alpha \vee \beta$ and $\alpha \wedge \beta$.
(3) If α is in $\mathbf{\Sigma}$, then so are $\exists x_i \alpha$ and $\forall (x_i \in x_j) \alpha$ for any distinct variables x_i, x_j.

If α is a strict Σ-formula and β is provably in HF equivalent to α then β will be called simply a *Σ-formula.*

Here $\forall (x_i \in x_j) \alpha$ is an abbreviation of $\forall x_i (x_i \in x_j \to \alpha)$ (which in turn is an abbreviation of $\neg \exists x_i (x_i \in x_j \wedge \neg \alpha)$). So the only permissible occurrences of $\to$ and $\neg$ in a strict Σ-formula are those which are introduced by the abbreviations $\forall (x_i \in x_j) \alpha$.

Lemma 2.2. *The formulas $x \subseteq y$, $x = y$ $z = x \triangleleft y$, $x = 0$, $x \in 0$, $0 \in x$ and all atomic formulas are Σ-formulas.*

Proof. We have the theorems:

(a) $x \subseteq y \leftrightarrow \forall(u \in x)[u \in y]$.

(b) $x = y \leftrightarrow x \subseteq y \wedge y \subseteq x$.

(c) $z = x \triangleleft y \leftrightarrow \forall(u \in z)[u \in x \vee u = y] \wedge (x \subseteq z) \wedge (y \in z)$.

(d) $x = 0 \leftrightarrow \forall(u \in x)[u \in u]$ (see Ap.1.1 and 1.13).

(e) $x \in 0 \leftrightarrow x \in x$ (see Ap.1.13).

(g) $0 \in x \leftrightarrow \exists y(y = 0 \wedge y \in x)$.

It follows from 2.1(1), (b) and (c) that every atomic formula is a Σ-formula. (E.g., $z = (x \triangleleft u) \triangleleft y$ is equivalent to $\exists v(z = v \triangleleft y \wedge v = x \triangleleft u)$.) ∎

Our way of coding terms (in Section 1) implies that, to code a longer term, one has to perform a *sequence* of coding operations, applied to its sub-terms. The same observation applies to coding formulas. Thus it is clear that formulas such as Term(x) or Form(x) (saying that x is the code of a term or a formula) must involve the sub-formulas Seq(s, k), LstSeq(s, k, y) described in the next lemma. We need to know that these are Σ-formulas.

Lemma 2.3. *Each of the following formulas is a Σ-formula*: Ord(x), Seq(s,k), Seq(s,k,y); *their respective meanings are*:

(a) Ord(x): *x is an ordinal number*,

(b) Seq(s, k): *s is a sequence of length $k > 0$ (i.e., s is a function whose domain is the ordinal number $k > 0$)*,

(c) LstSeq(s, k, y): *s is a sequence of length $k > 0$ and s terminates (ends) with y.*

Proof.

(a) $\mathrm{Ord}(x) \leftrightarrow \forall(y \in x)\{[y \subseteq x] \wedge \forall(z \in y)[z \subseteq y]\}$
(see 2.2 and Ap.2.1).

(b) $\mathrm{Seq}(s, k) \leftrightarrow \mathrm{Ord}(k) \wedge 0 \in k \wedge \forall(n \in k)\exists u[\langle n, u\rangle \in s] \wedge$
$\forall(y, z \in s)\exists(m, n \in k)\exists u\exists v[y = \langle m, u\rangle \wedge z = \langle n, v\rangle \wedge (m \neq n \vee u = v)]$.

(c) $\mathrm{LstSeq}(s, k, y) \leftrightarrow \mathrm{Seq}(s, k) \wedge \exists n(n \triangleleft n = k \wedge \langle n, y\rangle \in s)$ (see 2.2). ∎

In (b) the sub-formula $m \neq n \vee u = v$ replaces the implication $m = n \rightarrow u = v$ (which is not allowed in a strict Σ-formula). Observe that $m \neq n$ is a Σ-formula if m, n are ordinals, as it is equivalent to $m \in n \vee n \in m$ (see Ap.2.5).

In Ap.6 we shall construct the *standard model* $\mathfrak{S}$ of HF. Its elements are all the sets that can be written down using only $\{\,,\}$ and 0. For example: $0, \{0\}, \{0, \{0\}\}$, $\{0, \{\{\{0\}\}\}\}$. Each of these is represented by a constant term, but in general, not uniquely (e.g., the terms $(x \triangleleft y) \triangleleft z$ and $(x \triangleleft z) \triangleleft y$ are provably equal in HF).

A *Σ-sentence* is a Σ-formula that is a sentence. The key property of these sentences is that any true Σ-sentence is provable. Here "true" means "valid in the standard model" $\mathfrak{S}$ of HF and provability is in HF. In order to establish this property, we need the class of $\Sigma_{\triangleleft}$-formulas:

Definition 2.4. By a *$\Sigma_{\triangleleft}$-formula* we shall mean a formula that is made up of terms, the symbols $0, =, \in, \vee, \wedge, \exists x_i$ and $\forall(x_i \in \tau)$, where τ is any term that does not contain x_i.

For a $\Sigma_{\triangleleft}$-formula α, its *length* $\lambda(\alpha)$ is defined to be the total number of occurrences of the symbols $\vee, \wedge, \exists$ and $\forall$.

Let us observe that the length of a $\Sigma_{\triangleleft}$-formula remains unchanged if a variable occurring in the formula is replaced by a term (assuming that after this replacement we again get a formula). Evidently, every strict Σ-formula is a $\Sigma_{\triangleleft}$-formula.

THEOREM 2.5. *For every Σ-sentence α,*

$$(\S) \qquad \mathfrak{S} \models \alpha \;\Rightarrow\; \vdash \alpha.$$

Proof. Since every strict Σ-sentence is a $\Sigma_{\triangleleft}$-sentence, it will suffice to prove (§) for every $\Sigma_{\triangleleft}$-sentence α. We proceed by induction on the length $\lambda(\alpha)$.

Suppose $\lambda(\alpha) = 0$. As all logical operators (i.e., $\neg, \rightarrow, \vee, \wedge, \exists, \forall$) are then excluded from α, α must be of the form $\sigma \in \tau$ or $\sigma = \tau$, where $\sigma, \tau \in \mathbb{C}$. For a sentence of this kind, (§) follows directly from the definition of $\in$ and $=$ in the structure $\mathfrak{S}$ (see Ap.6.6).

Now suppose (§) has been shown for all $\Sigma_{\triangleleft}$-sentences β (in place of α) such that $\lambda(\beta) < \lambda(\alpha)$. We deduce that then (§) holds for the $\Sigma_{\triangleleft}$-sentence α, by considering the following cases:

- α is $\beta \vee \gamma$. In this case β and γ are $\Sigma_{\triangleleft}$-sentences and, by the inductive assumption,
$$\mathfrak{S} \models \alpha \;\Rightarrow\; (\mathfrak{S} \models \beta \text{ OR } \mathfrak{S} \models \gamma) \;\Rightarrow\; (\vdash \beta \text{ OR} \vdash \gamma) \;\Rightarrow\; \vdash \beta \vee \gamma \;\Rightarrow\; \vdash \alpha.$$
- α is $\beta \wedge \gamma$. The reasoning is analogous.
- α is $\exists x \beta(x)$. If $\mathfrak{S} \models \alpha$ then $\mathfrak{S} \models \beta(\tau)$ for some $\tau \in \mathbb{C}$. Since $\lambda(\beta(\tau)) = \lambda(\beta) < \lambda(\alpha)$, the inductive assumption yields
$$\mathfrak{S} \models \beta(\tau) \;\Rightarrow\; \vdash \beta(\tau) \;\Rightarrow\; \vdash \exists x \beta(x)$$
(by the rules of logic). Thus $\mathfrak{S} \models \alpha \;\Rightarrow\; \vdash \alpha$.
- α is $\forall(x \in \tau)\beta(x)$. Since there are no free variables in α and x does not occur in τ, we must have $\tau \in \mathbb{C}$. Thus, by Ap.6.2, there are $\tau_1, \ldots, \tau_m \in \mathbb{C}$ such that
$$(\S\S) \qquad \vdash \forall(x \in \tau)\beta(x) \leftrightarrow \beta(\tau_1) \wedge \ldots \wedge \beta(\tau_m).$$
Since $\lambda(\beta(\tau_j)) = \lambda(\beta(x)) < \lambda(\alpha)$ for all j, we can apply the inductive assumption, thus getting:
$$\mathfrak{S} \models \alpha \;\Rightarrow\; \mathfrak{S} \models \beta(\tau_1) \text{ AND } \ldots \text{ AND } \mathfrak{S} \models \beta(\tau_m) \;\Rightarrow$$
$$\vdash \beta(\tau_1) \text{ AND } \ldots \text{ AND} \vdash \beta(\tau_m) \;\Rightarrow\; \vdash \beta(\tau_1) \wedge \ldots \wedge \beta(\tau_m) \;\Rightarrow\; \vdash \alpha,$$
by (§§). ■

3. An HF-description of relations between formulas

In this section we cover the more tedious part of the work needed for producing the Σ-formula $\mathrm{Pf}(x)$. This amounts to a direct verification that all basic relations between variables, terms and formulas that underlie the idea of a *proof* can be described by means of Σ-formulas "talking" only about codes. All of these 25 formulas will become sub-formulas of $\mathrm{Pf}(x)$.

We shall introduce each of the formulas **1–25** below by stating first its intended meaning (in the meta-theory of HF) and then writing the formula itself. If any of the formulas $\mathrm{Seq}(s,k)$, $\mathrm{LstSeq}(s,k,t)$ (both introduced above) is used then, for any $m \in k$ (i.e., $m < k$), we shall write $\varphi(s_m)$ to abbreviate $\exists y[\langle m,y\rangle \in s \wedge \varphi(y)]$ (see Ap.3.2). Hence, if φ is a Σ-formula, then so is $\varphi(s_m)$.

1. $\mathrm{Var}(x)$
Means: x is the code of a variable.
$\mathrm{Ord}(x) \wedge 0 \in x$.

2. $\mathrm{SeqTerm}(s,k,t)$
Means: s is a sequence of length k and s ends with the code t of a term.
$\mathrm{LstSeq}(s,k,t) \wedge \forall(l \in k)\{s_l = 0 \vee \mathrm{Var}(s_l) \vee \exists(m,n \in l)[s_l = \langle \ulcorner\triangleleft\urcorner, s_m, s_n\rangle]\}$.

3. $\mathrm{Term}(t)$
Means: t is the code of a term.
$\exists k \exists s\, \mathrm{SeqTerm}(s,k,t)$.

4. $\mathrm{Const}(t)$
Means: t is the code of a constant term.
$\exists k \exists s[\mathrm{LstSeq}(s,k,t) \wedge \forall(l \in k)\{s_l = 0 \vee \exists(m,n \in l)[s_l = \langle \ulcorner\triangleleft\urcorner, s_m, s_n\rangle]\}]$.

5. $\mathrm{NecSeqTerm}(s,k,t)$
Means: s is a sequence of length k whose last element t is the code of a term and all terms whose codes appear in s (as some s_l) are sub-terms of the term coded by t (i.e., each s_l is *necessary* for the formation of the code t).
$\mathrm{SeqTerm}(s,k,t) \wedge \forall(l \in k)\{l^+ = k \vee$
$\exists(n,m \in k)[(l \in m) \wedge (s_m = \langle \ulcorner\triangleleft\urcorner, s_l, s_n\rangle \vee s_m = \langle \ulcorner\triangleleft\urcorner, s_n, s_l\rangle)]\}$.
In other words, each s_l, except the last, is used in creating some s_m further on.

Lemma 3.1. $\vdash \mathrm{Term}(t) \to \exists k \exists s\, \mathrm{NecSeqTerm}(s,k,t)$.

Proof. Assume $\mathrm{Term}(t)$, i.e., $\mathrm{SeqTerm}(s,k,t)$ for some sequence s of length k. If there is an $l \in k$ such that $\overline{k} \neq l$ (i.e., l is not the predecessor of k) and s_l is not needed further on in s (i.e., no s_m with $l \in m$ is of the form $s_m = \langle \ulcorner\triangleleft\urcorner, s_n, s_l\rangle$ or $s_m = \langle \ulcorner\triangleleft\urcorner, s_l, s_n\rangle$) then, omitting s_l, we obtain a sequence z of length $\overline{k}$ such that $\mathrm{SeqTerm}(z,\overline{k},t)$. Thus, for the shortest s (smallest k) such that $\mathrm{SeqTerm}(s,k,t)$, we have $\mathrm{NecSeqTerm}(s,k,t)$. ■

6. $\mathrm{VarOccTerm}(v,t)$
Means: v codes a variable, t codes a term and the variable coded by v occurs in the term coded by t.
$\mathrm{Var}(v) \wedge \exists k \exists s[\mathrm{NecSeqTerm}(s,k,t) \wedge \exists(l \in k)(s_l = v)]$.

7. $\mathrm{VarNonOccTerm}(v,t)$
Means: v codes a variable, t codes a term and the variable coded by v does not occur in the term coded by t.
$\mathrm{Var}(v) \wedge \exists k \exists s\{\mathrm{NecSeqTerm}(s,k,t) \wedge$
$\forall(l \in k)[s_l = 0 \vee (\mathrm{Var}(s_l) \wedge (s_l \neq v)) \vee \exists(n,m \in l)[s_l = \langle \ulcorner\triangleleft\urcorner, s_n, s_m\rangle]]\}$.

NOTE. Requiring NecSeqTerm(s,k,t) in **7** instead of SeqTerm(s,k,t) (obviously the latter would do) does not introduce any restrictions. Indeed, the proof of 3.1 shows that if there is a sequence s of some length k which terminates with t and satisfies the requirements of **7** without the "Nec" part, then there is also such a sequence (obtained from the previous one by pruning) where all s_l are "necessary" in the sense of satisfying the requirement NecSeqTerm.

Let us observe that **6**, **7** and 3.1 imply

$$\vdash \mathrm{Term}(t) \wedge \mathrm{Var}(v) \to \mathrm{VarOccTerm}(v,t) \vee \mathrm{VarNonOccTerm}(v,t).$$

8. At(y)
Means: y is the code of an atomic formula.
$\exists u \exists t[\mathrm{Term}(u) \wedge \mathrm{Term}(t) \wedge (y = \langle \ulcorner = \urcorner, u, t\rangle \vee y = \langle \ulcorner \in \urcorner, u, t\rangle)]$.

9. VarOccAt(v,y)
Means: v codes a variable, y codes an atomic formula and the variable coded by v occurs in the atomic formula coded by y.
$\exists u \exists t[\mathrm{Term}(u) \wedge \mathrm{Term}(t) \wedge (y = \langle \ulcorner = \urcorner, u, t\rangle \vee y = \langle \ulcorner \in \urcorner, u, t\rangle) \wedge$
$(\mathrm{VarOccTerm}(v,u) \vee \mathrm{VarOccTerm}(v,t))]$.

10. VarNonOccAt(v,y)
Means: v codes a variable, y codes an atomic formula and the variable coded by v does not occur in the atomic formula coded by y.
$\exists u \exists t[\mathrm{VarNonOccTerm}(v,u) \wedge \mathrm{VarNonOccTerm}(v,t) \wedge$
$(y = \langle \ulcorner = \urcorner, u, t\rangle \vee y = \langle \ulcorner \in \urcorner, u, t\rangle)]$.

11. MakeForm(u,w,v,y)
Means: v codes a variable and if u,w are codes of formulas, then y is the code of a formula created from one or both formulas coded by u,w and possibly the variable coded by v by a single application of $\vee$, $\neg$ or $\exists$.
$\mathrm{Var}(v) \wedge [y = \langle \ulcorner \vee \urcorner, u, w\rangle \vee y = \langle \ulcorner \vee \urcorner, w, u\rangle \vee y = \langle \ulcorner \neg \urcorner, u\rangle \vee y = \langle \ulcorner \exists \urcorner, v, u\rangle]$.

12. SeqForm(s,k,y)
Means: s is a sequence of length k whose last element y is the code of a formula.
$\mathrm{LstSeq}(s,k,y) \wedge \forall (n \in k)[\mathrm{At}(s_n) \vee \exists (m,l \in n) \exists v\, \mathrm{MakeForm}(s_m, s_l, v, s_n)]$.

13. Form(y)
Means: y is the code of a formula.
$\exists k \exists s\, \mathrm{SeqForm}(s,k,y)$.

14. NonAt(y)
Means: y is the code of a non-atomic formula.
$\mathrm{Form}(y) \wedge \exists u \exists w \exists v\, \mathrm{MakeForm}(u,w,v,y)$.

15. VarTopForm(v,y)
Means: y is the code of a formula beginning with $\exists x_i$, where $\ulcorner x_i \urcorner = v$.
$\mathrm{Var}(v) \wedge \exists w[\mathrm{Form}(w) \wedge y = \langle \ulcorner \exists \urcorner, v, w\rangle]$.

16. SeqVarTopForm(v, z, n, y)
Means: v codes a variable, y codes a formula and z is a sequence of length n showing how the formula coded by y is made up of formulas beginning with $\exists x_i$, where $\ulcorner x_i \urcorner = v$, and of atomic formulas, by using $\vee$, $\neg$ and $\exists x_j$, where $j \neq i$.
$\mathrm{Var}(v) \wedge \mathrm{LstSeq}(z, n, y) \wedge \forall(r \in n)[\mathrm{At}(z_r) \vee \mathrm{VarTopForm}(v, z_r) \vee$
$\exists(m, l \in r)\exists u(\mathrm{Var}(u) \wedge u \neq v \wedge \mathrm{MakeForm}(z_m, z_l, u, z_r))]$.

17. NecSeqVarTopForm(v, z, n, y)
Means: As in **15**, but with the additional requirement that all formulas whose codes appear in z (as some z_n) are sub-formulas of the formula coded by y (i.e., each z_n is *necessary* for the formation of the code y).
$\mathrm{SeqVarTopForm}(v, z, n, y) \wedge$
$\forall(r \in n)\{r^+ = n \vee \exists(l, m \in n)\exists u[(r \in m) \wedge \mathrm{MakeForm}(z_r, z_l, u, z_m)]\}$.

LEMMA 3.2. $\vdash \mathrm{Var}(v) \wedge \mathrm{Form}(y) \rightarrow \exists n \exists z\, \mathrm{NecSeqVarTopForm}(v, z, n, y)$.

Proof. Suppose $\mathrm{Var}(v)$ and $\mathrm{Form}(y)$. Then there is a sequence s of length k such that $\mathrm{SeqForm}(s, k, y)$. We now remove from s all s_l for which $\mathrm{VarTopForm}(v, s_l)$ and place these, possibly in the order of their first original appearance in s, at the beginning of a new sequence z. The remainder of z is composed of those s_l which were not removed from s. Let n be the length of z (actually $n = k$ but we do not need this fact). One checks that $\mathrm{SeqVarTopForm}(v, z, n, y)$. If n is the smallest ordinal for which there is a z satisfying this latter formula then, as in the proof of 3.1, we see that $\mathrm{NecSeqVarTopForm}(v, z, n, y)$. ■

18. VarOccFreeForm(v, y)
Means: v codes a variable, y codes a formula and the variable coded by v has a free occurrence in the formula coded by y.
$\exists n \exists z\{\mathrm{NecSeqVarTopForm}(v, z, n, y) \wedge \exists(r \in n)[\mathrm{At}(z_r) \wedge \mathrm{VarOccAt}(v, z_r)]\}$.

19. VarNonOccFreeForm(v, y)
Means: v codes a variable, y codes a formula and the variable coded by v has no free occurrence (is bound) in the formula coded by y.
$\exists n \exists z\{\mathrm{NecSeqVarTopForm}(v, z, n, y) \wedge$
$\forall(r \in n)[(\mathrm{At}(z_r) \wedge \mathrm{VarNonOccAt}(v, z_r)) \vee \mathrm{NonAt}(z_r)]\}$.

NOTE. Choosing NecSeqVarTopForm instead of SeqVarTopForm in **19** (clearly the latter would do) does not introduce any restrictions. (We justify this similarly to the Note following **7**.)

From 3.2 and the implication preceding **8** we see that

$$\vdash \mathrm{Var}(v) \wedge \mathrm{Form}(y) \rightarrow \mathrm{VarOccFreeForm}(v, y) \vee \mathrm{VarNonOccFreeForm}(v, y).$$

20. TermSubsVarForm(t, v, y)
Means: t codes a term, v codes a variable, y codes a formula and the term coded by t is substitutable for the variable coded by v in the formula coded by y.
$\mathrm{Term}(t) \wedge \exists n \exists z\{\mathrm{SeqVarTopForm}(v, z, n, y) \wedge$
$\forall(r \in n)[\mathrm{At}(z_r) \vee \exists(m, l \in r)\exists u\{z_r = \langle \ulcorner \vee \urcorner, z_m, z_l\rangle \vee z_r = \langle \ulcorner \neg \urcorner, z_m\rangle \vee$
$[z_r = \langle \ulcorner \exists \urcorner, u, z_m\rangle \wedge (\mathrm{VarNonOccFreeForm}(v, z_m) \vee$
$(\mathrm{VarOccFree\,Form}(v, z_m) \wedge \mathrm{VarNonOccTerm}(u, t)))]\}]\}$.

21. $\mathrm{SeqRepVarTermTerm}(s, s', k, v, t, u, u')$
Means: v codes a variable, t codes a term, u and u' are coding terms and s and s' are sequences of length k such that s ends with u and s' ends with u' and the term coded by u' is obtained from that coded by u by replacing each occurrence of the variable coded by v by the term coded by t.
$\mathrm{Var}(v) \wedge \mathrm{Term}(t) \wedge \mathrm{LstSeq}(s, k, u) \wedge \mathrm{LstSeq}(s', k, u') \wedge$
$\forall(n \in k)[((s_n = v) \wedge (s'_n = t)) \vee (\mathrm{Var}(s_n) \wedge (s_n \neq v) \wedge (s'_n = s_n)) \vee$
$((s_n = 0) \wedge (s'_n = 0)) \vee \exists(m, l \in n)[s_n = \langle \ulcorner \lhd \urcorner, s_m, s_l\rangle \wedge s'_n = \langle \ulcorner \lhd \urcorner, s'_m, s'_l\rangle]]$.

22. $\mathrm{RepVarTermTerm}(v, t, u, u')$
Means: v codes a variable, t codes a term, u and u' code terms and replacing in the term coded by u the variable coded by v, at all of its occurrences, by the term coded by t one obtains the term coded by u'.
$\exists k \exists s \exists s' \, \mathrm{SeqRepVarTermTerm}(s, s', k, v, t, u, u')$.

23. $\mathrm{RepVarTermAt}(v, t, y, y')$
Means: v codes a variable, t codes a term, y and y' are coding atomic formulas and replacing the variable coded by v by the term coded by t, at all occurrences of this variable in the atomic formula coded by y, one gets the atomic formula coded by y'.
$\exists u \exists u' \exists w \exists w' \{\mathrm{RepVarTermTerm}(v, t, u, u') \wedge \mathrm{RepVarTermTerm}(v, t, w, w') \wedge$
$[(y = \langle \ulcorner = \urcorner, u, w\rangle \wedge y' = \langle \ulcorner = \urcorner, u', w'\rangle) \vee (y = \langle \ulcorner \in \urcorner, u, w\rangle \wedge y' = \langle \ulcorner \in \urcorner, u', w'\rangle)]\}$.

24. $\mathrm{SeqRepVarTermForm}(s, s', k, v, t, y, y')$
Means: v codes a variable, t codes a term, y and y' code formulas and s and s' are sequences of length k such that s ends with y and s' ends with y' and the formula coded by y' is obtained from that coded by y by replacing each free occurrence of the variable coded by v by the term coded by t.
$\mathrm{Var}(v) \wedge \mathrm{LstSeq}(s, k, y) \wedge \mathrm{LstSeq}(s', k, y') \wedge$
$\forall(l \in k)\{[\mathrm{At}(s_l) \wedge \mathrm{RepVarTermAt}(v, t, s_l, s'_l)] \vee \exists(m, n \in l)\exists u$
$[(s_l = \langle \ulcorner \exists \urcorner, v, s_m\rangle \wedge s'_l = s_l) \vee (s_l = \langle \ulcorner \vee \urcorner, s_m, s_n\rangle \wedge s'_l = \langle \ulcorner \vee \urcorner, s'_m, s'_n\rangle) \vee$
$(\mathrm{Var}(u) \wedge u \neq v \wedge s_l = \langle \ulcorner \exists \urcorner, u, s_m\rangle \wedge s'_l = \langle \ulcorner \exists \urcorner, u, s'_m\rangle) \vee$
$(s_l = \langle \ulcorner \neg \urcorner, s_m\rangle \wedge s'_l = \langle \ulcorner \neg \urcorner, s'_m\rangle)]\}$.

25. $\mathrm{RepVarTermForm}(v, t, y, y')$
Means: v codes a variable, t codes a term, y and y' code formulas and replacing in the formula coded by y each free occurrence of the variable coded by v by the term coded by t, one obtains the formula coded by y'.
$\exists k \exists s \exists s' \, \mathrm{SeqRepVarTermForm}(s, s', k, v, t, y, y')$.

4. The formula Pf

A formula φ is a theorem of HF (in symbols: $\mathrm{HF} \vdash \varphi$ or $\vdash \varphi$) if there is a sequence of formulas, terminating with φ, where each formula in the sequence is either an *axiom of* HF or a *logical axiom* or is derived from one or two preceding formulas by a *rule of* *inference* (see 4.1 and 4.2 below). We begin this section by recalling the logical axioms and rules of inference for a first-order theory, as applied to HF. Next, these axioms and rules are expressed by Σ-formulas and codes. This leads to a Σ-formula $\mathrm{Pf}(x)$ such that $\vdash \varphi$ iff $\vdash \mathrm{Pf}(\ulcorner\varphi\urcorner)$ for every formula φ, moreover $\vdash \alpha \to \mathrm{Pf}(\ulcorner\alpha\urcorner)$ for every Σ-sentence α.

It is easily verified that the description of the logic of HF given below is equivalent (in the sense of yielding the same theorems) to the description of the logic of a first-order theory in [Sh] (when applied to HF).

Definition 4.1. The *logical axioms* are the following formulas (where $\varphi \to \psi$ abbreviates $\neg\varphi \vee \psi$):

Sentential (*Boolean*) *Axioms*: For any formulas φ, ψ, μ:

$$\begin{aligned} \varphi &\to \varphi, \\ \varphi &\to \varphi \vee \psi, \\ \varphi \vee \varphi &\to \varphi, \\ \varphi \vee (\psi \vee \mu) &\to (\varphi \vee \psi) \vee \mu, \\ (\varphi \vee \psi) \wedge (\neg\varphi \vee \mu) &\to \psi \vee \mu. \end{aligned}$$

Specialization Axioms: For every formula φ and every x_i:

$$\varphi \to \exists x_i \varphi.$$

Equality Axioms:

$$\begin{gathered} x_1 = x_1, \\ (x_1 = x_2) \wedge (x_3 = x_4) \to [(x_1 = x_3) \to (x_2 = x_4)], \\ (x_1 = x_2) \wedge (x_3 = x_4) \to [(x_1 \in x_3) \to (x_2 \in x_4)], \\ (x_1 = x_2) \wedge (x_3 = x_4) \to [x_1 \lhd x_3 = x_2 \lhd x_4]. \end{gathered}$$

Definition 4.2. The *rules of inference* are as follows:

Modus Ponens: $\dfrac{\varphi,\ \varphi \to \psi}{\psi}$.

Substitution: $\dfrac{\varphi}{\varphi(x_i/\tau)}$ for any term τ that is substitutable for x_i in φ.

$\exists$-*introduction*: $\dfrac{\varphi \to \psi}{\exists x_i \varphi \to \psi}$ provided x_i does not occur free in ψ.

For each of these three rules we say that the formula written below the line is *derived by that rule* from the formula(s) written above the line.

In order to write the formula Pf, it will be convenient to have short suggestive names for certain *terms* (see the coding of formulas in Section 1):

Definition 4.3.

$$\begin{aligned} \mathrm{Neg}(x) &= \langle \ulcorner\neg\urcorner, x\rangle, \\ \mathrm{Disj}(x, y) &= \langle \ulcorner\vee\urcorner, x, y\rangle, \\ \mathrm{Impl}(x, y) &= \mathrm{Disj}(\mathrm{Neg}(x), y), \end{aligned}$$

$$\begin{aligned}\mathrm{Conj}(x,y) &= \mathrm{Neg}(\mathrm{Disj}(\mathrm{Neg}(x),\mathrm{Neg}(y)),\\ \mathrm{Exi}(x,y) &= \langle \ulcorner\exists\urcorner, x, y\rangle,\\ \mathrm{All}(x,y) &= \mathrm{Neg}(\mathrm{Exi}(x,\mathrm{Neg}(y))).\end{aligned}$$

Keeping the style adopted for presenting the 25 Σ-formulas in Section 3, we now list nine Σ-formulas which describe the codes of axioms and the rules of inference. To begin, let us denote by $c_1, \ldots, c_6$ the constant terms which code the axioms HF1, HF2 and the four Equality Axioms in 4.1. The nine Σ-formulas are as follows.

(1) $\mathrm{Ax}(y)$
Means: y is the code of an individual axiom.
$y = c_1 \vee y = c_2 \vee \ldots \vee y = c_6$.

(2) $\mathrm{Ind}(x)$
Means: x is the code of an instance of the Induction Scheme HF3, i.e., of

$$\alpha(0) \wedge \forall x_1 \forall x_2[\alpha(x_1) \wedge \alpha(x_2) \rightarrow \alpha(x_1 \lhd x_2)] \rightarrow \forall x_1 \alpha(x_1),$$

where α is any formula such that x_2 is substitutable for x_1 in α (see Note below).
$\exists y\{\mathrm{Form}(y) \wedge \mathrm{TermSubsVarForm}(\ulcorner x_2\urcorner, \ulcorner x_1\urcorner, y) \wedge$
$\exists y' \exists y'' \exists y'''[\mathrm{RepVarTermForm}(\ulcorner x_1\urcorner, 0, y, y') \wedge$
$\mathrm{RepVarTermForm}(\ulcorner x_1\urcorner, \ulcorner x_2\urcorner, y, y'') \wedge$
$\mathrm{RepVarTermForm}(\ulcorner x_1\urcorner, \langle\ulcorner\lhd\urcorner, \ulcorner x_1\urcorner, \ulcorner x_2\urcorner\rangle, y, y''') \wedge x =$
$\mathrm{Impl}(\mathrm{Conj}(y', \mathrm{All}(\ulcorner x_1\urcorner, \mathrm{All}(\ulcorner x_2\urcorner, \mathrm{Impl}(\mathrm{Conj}(y, y''), y''')))), \mathrm{All}(\ulcorner x_1\urcorner, y))]\}$.

Note. To check the above, imagine that, for some α,

$$y = \ulcorner\alpha(x_1)\urcorner, \quad y' = \ulcorner\alpha(0)\urcorner, \quad y'' = \ulcorner\alpha(x_2)\urcorner, \quad y''' = \ulcorner\alpha(x_1 \lhd x_2)\urcorner.$$

(3) $\mathrm{Sent}(x)$
Means: x codes a formula that is an instance of one of the five Sentential Axioms (see 4.1).
$\exists y \exists z \exists w\{\mathrm{Form}(y) \wedge \mathrm{Form}(z) \wedge \mathrm{Form}(w) \wedge$
$[x = \mathrm{Impl}(y, y) \vee$
$x = \mathrm{Impl}(y, \mathrm{Disj}(y, z)) \vee$
$x = \mathrm{Impl}(\mathrm{Disj}(y, y), y) \vee$
$x = \mathrm{Impl}(\mathrm{Disj}(y, \mathrm{Disj}(z, w)), \mathrm{Disj}(\mathrm{Disj}(y, z), w) \vee$
$x = \mathrm{Impl}(\mathrm{Conj}(\mathrm{Disj}(y, z), \mathrm{Disj}(\mathrm{Neg}(y), w)), \mathrm{Disj}(z, w))]\}$.

(4) $\mathrm{Spec}(x)$
Means: x codes a Specialization Axiom (see 4.1).
$\exists y \exists v[\mathrm{Form}(y) \wedge \mathrm{Var}(v) \wedge x = \mathrm{Impl}(y, \mathrm{Exi}(v, y))]$.

(5) $\mathrm{ModPon}(x, y, y')$
Means: If x, y, y' are codes of formulas then y' codes a formula which follows from the formulas coded by x and y by an application of the Modus Ponens Rule (in 4.2).
$y = \mathrm{Impl}(x, y')$.

(6) $\mathrm{Subst}(y, y')$
Means: y, y' are codes of formulas and the formula coded by y' is obtained from that coded by y by an application of the Substitution Rule (in 4.2).
$\exists v \exists t[\mathrm{TermSubsVarForm}(t, v, y) \wedge \mathrm{RepVarTermForm}(v, t, y, y')]$.

(7) $\exists\mathrm{intro}(y, y')$
Means: y codes a formula and y' codes the formula obtained from that coded by y by an application of the $\exists$-introduction rule (in 4.2).
$\exists u \exists w \exists v[\mathrm{Form}(u) \wedge \mathrm{VarNonOccFreeForm}(v, w) \wedge$
$y = \mathrm{Impl}(u, w) \wedge y' = \mathrm{Impl}(\mathrm{Exi}(v, u), w)]$.

(8) $\mathrm{Prf}(s, k, y)$
Means: s is a sequence of length k of codes of formulas such that the corresponding sequence of formulas is a proof of the formula coded by y.
$\mathrm{LstSeq}(s, k, y) \wedge \forall(n \in k)[\mathrm{Ax}(s_n) \vee \mathrm{Ind}(s_n) \vee \mathrm{Sent}(s_n) \vee \mathrm{Spec}(s_n) \vee$
$\exists(m, l \in n)[\mathrm{ModPon}(s_m, s_l, s_n) \vee \mathrm{Subst}(s_m, s_n) \vee \exists\,\mathrm{intro}(s_m, s_n)]]$.

(9) $\mathrm{Pf}(y)$
Means: y codes a formula which is a theorem of ZF.
$\exists k \exists s\, \mathrm{Prf}(s, k, y)$.

Proposition 4.4 (Proof formalization condition). *For every formula φ,*

$$\vdash \varphi \quad \textit{iff} \quad \vdash \mathrm{Pf}(\ulcorner \varphi \urcorner).$$

Proof. Assume $\vdash \varphi$. Then there is a sequence of formulas $\varphi_0, \ldots, \varphi_p$ that constitutes a proof of φ. If $s = \{\langle l, \ulcorner \varphi_l \urcorner \rangle : l \in p^+\}$, then $s \in \mathbb{S}$ (= the universe of the standard model $\mathfrak{S}$; see Ap.6.6) and $\mathfrak{S} \models \mathrm{Prf}(s, p^+, \ulcorner \varphi \urcorner)$, so that $\mathfrak{S} \models \mathrm{Pf}(\ulcorner \varphi \urcorner)$. Since $\mathrm{Pf}(\ulcorner \varphi \urcorner)$ is a Σ-sentence, we conclude, by 2.5, that $\vdash \mathrm{Pf}(\ulcorner \varphi \urcorner)$.

Now assume $\vdash Pf(\ulcorner \varphi \urcorner)$. Hence $\mathfrak{S} \models \exists k \exists s\, \mathrm{Prf}(s, k, \ulcorner \varphi \urcorner)$. This means that there is a sequence $s \in \mathbb{S}$ of length k, where each s_l codes some formula φ_l, so that $\varphi_{\overline{k}}$ is φ and the sequence of formulas $(\varphi_0, \ldots, \varphi_{\overline{k}})$ constitutes a proof of φ. Thus $\vdash \varphi$. ■

5. Phantom terms, functions and formulas

As commonly defined, we call f a *function* if all elements of f are ordered pairs and

$$\langle u, w \rangle \in f \wedge \langle u, z \rangle \in f \to w = z.$$

Associated to a function are two sets: the domain $\mathrm{dom}(f)$ and the range $\mathrm{rng}(f)$. Functions can be obtained also from the function symbols of the language, in our case, from $\triangleleft$. For example, $x \triangleleft x$ determines a function once we select some set for the domain.

More generally, functions can be obtained from certain formulas. Thus let φ be a formula with exactly $n + 1$ free variables ($n \geq 0$) and let y be one of these variables. Suppose that $\exists! y \varphi$ is a theorem of HF ("$\exists!$" means: "$\exists$ a unique"). Then we shall associate with φ and y a new symbol F^y_φ and we shall call it an n-ary *p-function symbol*, where "p" might stand for "phantom" to stress that we do not wish to consider F^y_φ as part of the language of HF (in [Bo], "p" stands for "pseudo"). We shall call φ the *defining* *formula* for F^y_φ. A term $F^y_\varphi(z_1,\dots,z_n)$ (of the language of HF extended by the p-function symbol F^y_φ) will be called a *p-term*. We would like to write formulas including p-terms, however these will be always regarded as representing certain formulas of HF. Thus let $x_{i_1},\dots,x_{i_n}$ be the other free variables in φ, where $i_1 < \dots < i_n$. Then the representation rule is as follows: If $\beta(z)$ is an atomic formula of HF then

$$\beta(F^y_\varphi(z_1,\dots,z_n)) \textit{ represents any formula } \exists w[\varphi(z_1,\dots,z_n,w) \wedge \beta(w)],$$

where w is a variable substitutable for y in φ, moreover each z_k is substitutable for x_{i_k} in φ. Thus $\alpha(F^y_\varphi(z_1,\dots,z_n))$ represents many formulas, however these are equivalent to each other in HF. The above rule is extended in an obvious way to all formulas containing the symbol F^y_φ so that the representation commutes with the logical connectives and quantifiers (for details see Ap.3).

Only capital letters will be used for p-function symbols and we shall omit the suffix φ and exponent y when circumstances permit. Also, we shall sometimes say *p-function* to mean a p-function symbol F^y_φ, or even its whole definition. Several p-functions will be introduced below. Then by a *p-term* we shall mean any term of the extended language, i.e., a term formed from the variables, 0, the function symbol $\triangleleft$ and the p-function symbols. Formulas containing such terms will be called *p-formulas*. The above abbreviation principle can be naturally extended to any p-formula which then is easily seen to be an abbreviation of a (not unique) formula of HF. By a slight abuse of language we shall occasionally say that a p-formula α *is* a theorem of HF to mean that α *abbreviates* such a theorem.

Our goal in this section is to prepare the definition, for every i, of a p-function K such that for every formula $\gamma(x_i)$, the p-formula $K(\ulcorner\gamma\urcorner) = \ulcorner\gamma(\ulcorner\gamma\urcorner)\urcorner$ is a theorem of HF.

Proposition 5.1. *There exists a ternary p-function, denoted here by* REPL, *such that*

$$\vdash \mathrm{REPL}(\ulcorner x_i\urcorner, \ulcorner\tau\urcorner, \ulcorner\varphi\urcorner) = \ulcorner\varphi(x_i/\tau)\urcorner$$

for every variable x_i, term τ and formula φ [where $\varphi(x_i/\tau)$ results from replacing each free occurrence of x_i in φ by τ].

(We are not concerned at present with the issue of substitutability.) As the defining formula $\varphi(v,t,y,y')$ for REPL we choose the formula:

$$[(\exists! z)\,\mathrm{RepVarTermForm}(v,t,y,z) \wedge \mathrm{RepVarTermForm}(v,t,y,y')] \vee$$
$$[\neg(\exists! z)\,\mathrm{RepVarTermForm}(v,t,y,z) \wedge y' = 0].$$

It is clear that $(\exists! y')\varphi(v,t,y,y')$. So, for the reader who found our description of the meaning of formula **25** (Section 3) wholly convincing, the above proposition requires no further proof. For the more meticulous, we suggest checking through Lemmas 5.2–5.4.

Lemma 5.2. *For every term μ,*

$$\vdash \mathrm{RepVarTermTerm}(\ulcorner x_i\urcorner, \ulcorner\tau\urcorner, \ulcorner\mu\urcorner, u') \leftrightarrow u' = \ulcorner\mu(x_i/\tau)\urcorner.$$

(Here $\mu(x_i/\tau)$ is the term obtained from μ by replacing each occurrence of x_i by τ.)

Proof. It will be enough to show that for any μ,

$$(1)\qquad \vdash \mathrm{RepVarTermTerm}(\ulcorner x_i\urcorner, \ulcorner\tau\urcorner, \ulcorner\mu\urcorner, u') \rightarrow u' = \ulcorner\mu(x_i/\tau)\urcorner,$$

$$(2)\qquad \vdash \mathrm{RepVarTermTerm}(\ulcorner x_i\urcorner, \ulcorner\tau\urcorner, \ulcorner\mu\urcorner, \ulcorner\mu(x_i/\tau)\urcorner).$$

Proof of (1). It will suffice to show (see **22**) that

(3) $$\mathrm{SeqRepVarTermTerm}(s, s', k, \ulcorner x_i \urcorner, \ulcorner \tau \urcorner, \ulcorner \mu \urcorner, u') \to u' = \ulcorner \mu(x_i/\tau) \urcorner$$

is a theorem of HF. We keep the variable x_i and term τ fixed and proceed by induction with respect to the length (number of appearances of $\lhd$) of μ. It is clear from **21** that if μ is shortest possible, i.e., μ is 0 or a variable x_j, then u' is either 0 or $\ulcorner x_j \urcorner$ (if $j \neq i$) or $\ulcorner \tau \urcorner$ (if $j = i$) as required. For the inductive step one assumes that $\mu = \sigma \lhd \nu$ in (3). Then, assuming the left side of (3), we see from **21** that $\ulcorner \sigma \urcorner = s_m$, $\ulcorner \nu \urcorner = s_n$ for some $m, n \in k$ and

$$\ulcorner \mu \urcorner = \langle \ulcorner \lhd \urcorner, s_m, s_n \rangle, \qquad u' = \langle \ulcorner \lhd \urcorner, s'_m, s'_n \rangle.$$

If in (3) we replace s by the initial segment of s terminating with s_m, and s' by the initial segment of s' terminating with s'_m, and μ by σ, then, by the inductive hypothesis (σ being shorter than μ), we get $s'_m = \ulcorner \sigma(x_i/\tau) \urcorner$. Similarly, $s'_n = \ulcorner \nu(x_i/\tau) \urcorner$, so that finally

$$u' = \langle \ulcorner \lhd \urcorner, s'_m, s'_n \rangle = \langle \ulcorner \lhd \urcorner, \ulcorner \sigma(x_i/\tau) \urcorner, \ulcorner \nu(x_i/\tau) \urcorner \rangle = \ulcorner \sigma(x_i/\tau) \lhd \nu(x_i/\tau) \urcorner = \ulcorner \mu(x_i/\tau) \urcorner.$$

Proof of (2). By **22** and (3), it will suffice to find a constant (term) c such that

(4) $$\vdash \exists k \exists s \exists s' \, \mathrm{SeqRepVarTermTerm}(s, s', k, \ulcorner x_i \urcorner, \ulcorner \tau \urcorner, \ulcorner \mu \urcorner, c).$$

Now, for the given term μ, one can explicitly exhibit a sequence s, of some length k, of constant terms such that $\mathfrak{S} \models \mathrm{SeqTerm}(s, k, \ulcorner \mu \urcorner)$. Then formula **21** tells us exactly how to find a sequence s' of constant terms, terminating with some constant c, for which

$$\mathfrak{S} \models \mathrm{SeqRepVarTermTerm}(s, s', k, \ulcorner x_i \urcorner, \ulcorner \tau \urcorner, \ulcorner \mu \urcorner, c).$$

The last formula is a Σ-sentence, hence it is a theorem of HF (see 2.5). Thus (4) follows. ∎

Lemma 5.3. *If φ is an atomic formula, then*

$$\vdash \mathrm{RepVarTermAt}(\ulcorner x_i \urcorner, \ulcorner \tau \urcorner, \ulcorner \varphi \urcorner, y) \leftrightarrow y = \ulcorner \varphi(x_i/\tau) \urcorner.$$

Proof. This is a direct consequence of 5.2 and **23**. ∎

Lemma 5.4. *For every formula φ,*

$$\vdash \mathrm{RepVarTermForm}(\ulcorner x_i \urcorner, \ulcorner \tau \urcorner, \ulcorner \varphi \urcorner, y') \leftrightarrow y' = \ulcorner \varphi(x_i/\tau) \urcorner.$$

(Just now we leave aside the issue of substitutability of τ for x_i.)

Proof. Very similar to the proof of 5.2. Instead of (1), (2), we now have to establish

(1′) $$\vdash \mathrm{RepVarTermForm}(\ulcorner x_i \urcorner, \ulcorner \tau \urcorner, \ulcorner \varphi \urcorner, y') \to y' = \ulcorner \varphi(x_i/\tau) \urcorner,$$

(2′) $$\vdash \mathrm{RepVarTermForm}(\ulcorner x_i \urcorner, \ulcorner \tau \urcorner, \ulcorner \varphi \urcorner, \ulcorner \varphi(x_i/\tau) \urcorner).$$

The proof of (1′) reduces to showing (see **25**)

(3′) $$\vdash \mathrm{SeqRepVarTermForm}(s, s', k, \ulcorner x_1 \urcorner, \ulcorner \tau \urcorner, \ulcorner \varphi \urcorner, y') \to y' = \ulcorner \varphi(x_i/\tau) \urcorner.$$

To show this we use induction on the length of φ. (Here by "length" we mean the number of occurrences of $\vee$, $\neg$ and $\exists$.) If φ is atomic (i.e., of length 0), then (3′) is a direct consequence of **24** and 5.3. For the inductive step, one has to consider four possibilities for φ: $\alpha \vee \beta$, $\neg\alpha$, $\exists x_i \alpha$, and $\exists x_j \alpha$, where $j \neq i$. In each of these cases the inductive hypothesis is applied (to the shorter formula α or β) similarly to the proof of (3) above.

The proof of (2′) is entirely analogous to that of (2) above. ∎

6. Gödel's First Incompleteness Theorem

In this section we wish to prove Gödel's Diagonal Lemma and then his First Incompleteness Theorem. To start with, we introduce two unary p-functions W, H that describe the formation of certain codes. Since a code (of a variable, a term or a formula) is a constant term, that code can be coded again. H will do this, i.e., we shall have $H(\mu) = \ulcorner \mu \urcorner$, whenever μ is already a code. W is just the restriction of H to the codes of variables (i.e., to the ordinals), however we begin with W.

Recall that $\ulcorner x_1 \urcorner$ is $0 \triangleleft 0$, i.e. 0^+, and in general, $\ulcorner x_{k+} \urcorner$ is $\ulcorner x_k \urcorner \triangleleft \ulcorner x_k \urcorner$, i.e. $(\ulcorner x_k \urcorner)^+$, for every ordinal k. The predecessor of a non-zero ordinal x is $\overline{x}$.

Lemma 6.1. *There is a p-function W such that*

$$\vdash W(\ulcorner x_k \urcorner) = \ulcorner (\ulcorner x_k \urcorner) \urcorner$$

for every variable x_k.

Proof. Define W *recursively on ordinals* by (see Ap.3.4):

$$W(x) = \begin{cases} 0 & \text{if } x = 0 \text{ or } x \text{ is not an ordinal,} \\ \langle \ulcorner \triangleleft \urcorner, W(\overline{x}), W(\overline{x}) \rangle & \text{if } x \text{ is a non-zero ordinal.} \end{cases}$$

By this definition,

$$W(\ulcorner x_1 \urcorner) = W(0^+) = \langle \ulcorner \triangleleft \urcorner, W(0), W(0) \rangle = \langle \ulcorner \triangleleft \urcorner, 0, 0 \rangle = \ulcorner 0 \triangleleft 0 \urcorner = \ulcorner (\ulcorner x_1 \urcorner) \urcorner,$$

as required. Assuming that $W(\ulcorner x_k \urcorner) = \ulcorner (\ulcorner x_k \urcorner) \urcorner$, we get

$$\begin{aligned} W(\ulcorner x_{k+} \urcorner) &= W((\ulcorner x_k \urcorner)^+) = \langle \ulcorner \triangleleft \urcorner, W(\ulcorner x_k \urcorner), W(\ulcorner x_k \urcorner) \rangle = \langle \ulcorner \triangleleft \urcorner, \ulcorner (\ulcorner x_k \urcorner) \urcorner, \ulcorner (\ulcorner x_k \urcorner) \urcorner \rangle \\ &= \ulcorner (\ulcorner x_k \urcorner \triangleleft \ulcorner x_k \urcorner) \urcorner = \ulcorner (\ulcorner x_{k+} \urcorner) \urcorner. \ \blacksquare \end{aligned}$$

Recall that all codes (as described in Section 1) belong to the family of terms Γ defined in 1.1.

Lemma 6.2. *There is a p-function H such that $H(\mu) = \ulcorner \mu \urcorner$ for every μ in Γ.*

Proof. According to the definition of an ordered pair,

$$\langle x, y \rangle = \{\{x\}, \{x, y\}\} = \{0 \triangleleft x, (0 \triangleleft x) \triangleleft y\} = (0 \triangleleft (0 \triangleleft x)) \triangleleft ((0 \triangleleft x) \triangleleft y).$$

Thus, for any terms σ, τ,

$$\begin{aligned} \ulcorner \langle \sigma, \tau \rangle \urcorner &= \ulcorner (0 \triangleleft (0 \triangleleft \sigma)) \triangleleft ((0 \triangleleft \sigma) \triangleleft \tau) \urcorner \\ &= \langle \ulcorner \triangleleft \urcorner, \langle \ulcorner \triangleleft \urcorner, 0, \langle \ulcorner \triangleleft \urcorner, 0, \ulcorner \sigma \urcorner \rangle \rangle, \langle \ulcorner \triangleleft \urcorner, \langle \ulcorner \triangleleft \urcorner, 0, \ulcorner \sigma \urcorner \rangle, \ulcorner \tau \urcorner \rangle \rangle. \end{aligned}$$

Let $t(x, y)$ be the term

$$\langle \ulcorner \triangleleft \urcorner, \langle \ulcorner \triangleleft \urcorner, 0, \langle \ulcorner \triangleleft \urcorner, 0, x \rangle \rangle, \langle \ulcorner \triangleleft \urcorner, \langle \ulcorner \triangleleft \urcorner, 0, x \rangle, y \rangle \rangle.$$

Thus $t(\ulcorner \sigma \urcorner, \ulcorner \tau \urcorner)$ is $\ulcorner \langle \sigma, \tau \rangle \urcorner$ for any terms μ, σ. We now define $H(x)$ *recursively* (see Ap.4.9) as

$$H(x) = \begin{cases} W(x) & \text{if } x \text{ is an ordinal,} \\ t(H(u), H(v)) & \text{if } x = \langle u, v \rangle \text{ for some } u, v, \\ 0 & \text{in all other cases.} \end{cases}$$

The definition makes sense because no ordered pair is an ordinal ($0 \in x$ for every non-zero ordinal x, whereas $0 \notin \langle u, v\rangle$ for all u, v). Since $x = \langle u, v\rangle \to u, v \in \mathrm{cl}(x)$, the requirements of Ap.4.9 are satisfied.

Now suppose $\mu \in \Gamma$. If $\mu = 0$ or μ is the code of a variable, then $H(\mu) = W(\mu) = \ulcorner\mu\urcorner$ because the first line of the definition of H applies and we can use 6.1. All other terms μ in Γ are generated from 0 and from the codes of variables by (repeatedly) forming ordered pairs. Since for $\mu = \langle\sigma, \tau\rangle$ the terms σ, τ are shorter than μ, we may proceed by induction and assume that $H(\sigma) = \ulcorner\sigma\urcorner$, $H(\tau) = \ulcorner\tau\urcorner$. Then, by the definition of $t(x, y)$,

$$\ulcorner\mu\urcorner = \ulcorner\langle\sigma, \tau\rangle\urcorner = t(\ulcorner\sigma\urcorner, \ulcorner\tau\urcorner) = t(H(\sigma), H(\tau)) = H(\mu),$$

as we wished to show. ∎

It is clear that no unary p-function H can satisfy $H(\tau) = \ulcorner\tau\urcorner$ for *all* constant terms τ. Indeed, if this were so, the terms $\sigma = (0 \lhd \mu) \lhd \nu$ and $\tau = (0 \lhd \nu) \lhd \mu$, where μ, ν are different constant terms, would have provably equal codes, which is easily seen not to be the case.

Let us agree that when a formula φ is written as $\varphi(x_i)$ then, for any term τ, we shall write $\varphi(\tau)$ to mean $\varphi(x_i/\tau)$.

Lemma 6.3. *For every variable x_i, there exists a p-function K such that, for every formula $\varphi(x_i)$,*

$$\vdash K(\ulcorner\varphi\urcorner) = \ulcorner\varphi(\ulcorner\varphi\urcorner)\urcorner.$$

Proof. Consider the Replacement Function REPL defined in Section 5. Noting that the composite of p-functions is a p-function, we put

$$K(x) = \mathrm{REPL}(\ulcorner x_i\urcorner, H(x), x).$$

By 5.1, for any formula $\varphi(x_i)$ and any term τ one has

$$\vdash \mathrm{REPL}(\ulcorner x_i\urcorner, \ulcorner\tau\urcorner, \ulcorner\varphi\urcorner) = \ulcorner\varphi(\tau)\urcorner.$$

Let $\tau = \ulcorner\varphi\urcorner$ and recall that $\ulcorner\tau\urcorner = H(\tau)$ in this case (by 6.2). Then

$$\vdash \mathrm{REPL}(\ulcorner x_i\urcorner, H(\ulcorner\varphi\urcorner), \ulcorner\varphi\urcorner) = \ulcorner\varphi(\ulcorner\varphi\urcorner)\urcorner,$$

i.e. $\vdash K(\ulcorner\varphi\urcorner) = \ulcorner\varphi(\ulcorner\varphi\urcorner)\urcorner$. ∎

Theorem 6.4 (Gödel's Diagonal Lemma). *For every formula $\alpha(x_i)$, there is a formula δ such that*

$$\vdash \delta \leftrightarrow \alpha(\ulcorner\delta\urcorner).$$

Proof. We replace the variable x_i in α by the p-term $K(x_i)$, and we denote by $\beta(x_i)$ the resulting formula. Thus

$$\vdash \beta(x_i) \leftrightarrow \alpha(K(x_i)).$$

Then, by 6.3,

$$\vdash \beta(\ulcorner\beta\urcorner) \leftrightarrow \alpha(\ulcorner(\ulcorner\beta\urcorner)\urcorner)).$$

Thus a formula δ satisfying the claim of the Diagonal Lemma is $\beta(\ulcorner\beta\urcorner)$. ∎

Theorem 6.5 (Gödel's First Incompleteness Theorem). *If* HF *is a consistent theory, then there is a sentence δ such that neither* HF $\vdash \delta$ *nor* HF $\vdash \neg\delta$. *Moreover*, $\mathfrak{S} \vDash \delta$.

Proof. Consider the p-formula $\neg\,\mathrm{Pf}(x_1)$. By the Diagonal Lemma, there is a sentence δ such that

$$(\triangle) \qquad \vdash \delta \leftrightarrow \neg\,\mathrm{Pf}(\ulcorner\delta\urcorner).$$

(a) Suppose $\vdash \delta$. Then $\vdash \mathrm{Pf}(\ulcorner\delta\urcorner)$, by 4.4. On the other hand, from $(\triangle)$ follows $\vdash \neg\,\mathrm{Pf}(\ulcorner\delta\urcorner)$. This contradicts the assumption of consistency.

(b) Suppose $\vdash \neg\delta$. Then, by $(\triangle)$, $\vdash \mathrm{Pf}(\ulcorner\delta\urcorner)$ and hence $\vdash \delta$, by 4.4. This, together with the assumed $\vdash \neg\delta$, contradicts the consistency assumption.

(c) Suppose $\textsc{non}[\mathfrak{S} \vDash \delta]$. Then $\mathfrak{S} \vDash \neg\delta$, and hence $\mathfrak{S} \vDash \mathrm{Pf}(\ulcorner\delta\urcorner)$, by $(\triangle)$. So $\vdash \delta$, by 4.4, contradicting $\textsc{non}[\mathfrak{S} \vDash \delta]$. ■

7. Pseudo-coding

Our next (and last) goal is to prove Gödel's Second Incompleteness Theorem, i.e., 9.8. A glance at its proof in Section 9 will show that at this moment it hinges only on the Hilbert–Bernays derivability condition

$$\vdash \mathrm{Pf}(\ulcorner\delta\urcorner) \to \mathrm{Pf}(\ulcorner\mathrm{Pf}(\ulcorner\delta\urcorner)\urcorner),$$

albeit only for the sentence δ satisfying $(\triangle)$ in 6.5. In what follows, the above implication will be obtained as a special case of $\alpha \to \mathrm{Pf}(\ulcorner\alpha\urcorner)$, which we show to be a theorem for every Σ-sentence α. (Clearly $\mathrm{Pf}(\ulcorner\delta\urcorner)$ is a Σ-sentence.) Often results are proved by induction on the length of a formula, but in this case induction does not work: $\alpha \to \mathrm{Pf}(\ulcorner\alpha\urcorner)$ is not a theorem for every Σ-formula α. Indeed, we can take α to be $x_1 = x_2$. Then, accepting as a theorem $x_1 = x_2 \to \mathrm{Pf}(\ulcorner x_1 = x_2\urcorner)$, we would quickly reach a contradiction: Since x_2 does not occur in $\mathrm{Pf}(\ulcorner x_1 = x_2\urcorner)$ (i.e., in $\mathrm{Pf}(\langle\ulcorner=\urcorner, 0 \lhd 0, (0 \lhd 0) \lhd (0 \lhd 0)\rangle)$), we would conclude $\vdash x_1 = x_1 \to \mathrm{Pf}(\ulcorner x_1 = x_2\urcorner)$, hence $\vdash \mathrm{Pf}(\ulcorner x_1 = x_2\urcorner)$, and finally $\vdash x_1 = x_2$, by 4.4.

The approach that works is as follows. First, we define a *pseudo-code* of a formula α, denoting it by $[[\alpha]]$. Such a pseudo-code is a term, like the code $\ulcorner\alpha\urcorner$, but whereas $\ulcorner\alpha\urcorner$ is a constant term, $[[\alpha]]$ is in general a term with variables. In fact, the variables of the term $[[\alpha]]$ are exactly those that are free in α. Next, we describe a rather special p-function F (we call it the *special* p-function). Then we replace each variable x_i occurring in $[[\alpha]]$ by the p-term $F(x_i)$. The resulting p-term is denoted by $[[\alpha]](F(x_1), \ldots, F(x_n))$, where n is chosen large enough so that all variables in $[[\alpha]]$ are among $x_1, \ldots, x_n$. Then we show by induction on the length of the Σ-formula α that

$$\alpha \to \mathrm{Pf}([[\alpha]](F(x_1), \ldots, F(x_n)))$$

is a theorem of HF. It is part of the definition of a pseudo-code that if α is sentence then $[[\alpha]]$ is identical with $\ulcorner\alpha\urcorner$. This allows us to conclude $\vdash \alpha \to \mathrm{Pf}(\ulcorner\alpha\urcorner)$ for every Σ-sentence α.

To create $[[\alpha]]$ we follow the same coding procedure as the one leading to the code $\ulcorner\alpha\urcorner$, except that no variable x_i occurring *free* in α is coded. The free occurrences of x_i are just left standing. For example:

$$[[x_1 = x_2]] = \langle\ulcorner=\urcorner, x_1, x_2\rangle \quad \text{and} \quad [[\exists x_1(x_1 \in x_3)]] = \langle\ulcorner\exists\urcorner, \ulcorner x_1\urcorner, \langle\ulcorner\in\urcorner, \ulcorner x_1\urcorner, x_3\rangle\rangle.$$

If we try to define $[[\alpha]]$ by induction on the length of the formula α, i.e., by analogy with defining $\ulcorner\alpha\urcorner$, we hit an obstacle: when an occurrence of x_i is to be coded at a stage when a sub-formula of α containing x_i free is coded, one does not know yet whether this occurrence of x_i will remain free or finally become bound in α. To overcome this obstacle, we introduce the concept of an intermediary code $[[\alpha]]_V$ corresponding to any set V of indices of variables (i.e., non-zero ordinals). This code $[[\alpha]]_V$, which we call a *V-code*, is created just like $[[\alpha]]$, except that for an occurrence of x_i *not* to be coded, we require not only that this occurrence is free in α but also that $i \in V$. Thus $[[\alpha]]_V$ is $[[\alpha]]$ when V contains all the indices of the variables free in α, and $[[\alpha]]_V$ is $\ulcorner\alpha\urcorner$ when V is empty.

V-coding can be defined inductively. We begin with the terms. We put $[[0]]_V = 0$, and for any variable x_i we define

$$[[x_i]]_V = \begin{cases} x_i & \text{if } i \in V, \\ \ulcorner x_i \urcorner & \text{otherwise.} \end{cases}$$

The V-codes of other terms are defined by induction on the length of the term: if σ, τ are V-coded, then $[[\sigma \lhd \tau]]_V = \langle \ulcorner\lhd\urcorner, [[\sigma]]_V, [[\tau]]_V\rangle$.

The V-codes of atomic formulas are defined by $[[\sigma \in \tau]]_V = \langle \ulcorner\in\urcorner, [[\sigma]]_V, [[\tau]]_V\rangle$ and $[[\sigma = \tau]]_V = \langle \ulcorner=\urcorner, [[\sigma]]_V, [[\tau]]_V\rangle$ (for any terms σ, τ).

The V-codes of other formulas are defined by induction on the length of the formula:

$$\begin{aligned} [[\neg\alpha]]_V &= \langle \ulcorner\neg\urcorner, [[\alpha]]_V\rangle, \\ [[\alpha \vee \beta]]_V &= \langle \ulcorner\vee\urcorner, [[\alpha]]_V, [[\beta]]_V\rangle, \\ [[\exists x_k \alpha]]_V &= \langle \ulcorner\exists\urcorner, \ulcorner x_k\urcorner, [[\alpha]]_{V\setminus\{k\}}\rangle. \end{aligned}$$

Definition 7.1. If V contains the indices of *all* variables occurring *free* in α then $[[\alpha]]_V$ (obviously not depending on V in that case) will be denoted by $[[\alpha]]$ and called the *pseudo-code* of α.

It may be worth observing that if we replace each variable x_i occurring in the term $[[\alpha]]$ by $\ulcorner x_i\urcorner$ then we obtain $\ulcorner\alpha\urcorner$.

In this section we describe by means of pseudo-codes some properties of the formula $\mathrm{Pf}(x)$. Our goal is Theorem 7.3. This will be needed for proving Theorem 9.1, from which the Hilbert–Bernays condition and then Gödel's Second Theorem are easily deduced.

Definition 7.2. Given any string $t_1, \ldots, t_n$ of variables and any V-code $[[\alpha]]_V$, we shall denote by $[[\alpha]]_V(t_1, \ldots, t_n)$ the term that results from $[[\alpha]]_V$ when each variable x_i occurring in $[[\alpha]]_V$, with $i \le n$, is replaced by t_i.

Suppose β is a theorem, i.e., $\vdash\beta$. If we replace each of the variables $x_1, \ldots, x_n$ at each of its free occurrences in β by some constant term then the formula so obtained is also a theorem (by the Substitution Rule in 4.2). This just described situation in the meta-theory admits description in HF. Thus, $\vdash\beta$ is equivalent to $\mathrm{Pf}(\ulcorner\beta\urcorner)$ (see 4.4). If $t_1, \ldots, t_n$ are the codes of the constant terms that are substituted for $x_1, \ldots, x_n$ (at the free occurrences of these variables) then this new theorem we get from β has the $\{1, \ldots, n\}$-code $[[\beta]]_{\{1,\ldots,n\}}(t_1, \ldots, t_n)$. Hence the mentioned description in HF is the implication ($\diamond$) in 7.3 below. The reader who accepts the accuracy of our description of the meta-theory of HF by formulas of HF in Sections 3 and 4 might feel no need to see the proof of ($\diamond$) (Lemmas 7.4–7.6). We give it here for the sake of completeness, although it is rather long and tedious (due to the fact that pseudo-codes do not participate in the creation of the formula Pf).

THEOREM 7.3. *For every formula β and every $n = 1, 2, \dots$,*

$$(\diamond) \qquad \mathrm{Const}(t_1) \wedge \ldots \wedge \mathrm{Const}(t_n) \wedge \mathrm{Pf}(\ulcorner\beta\urcorner) \to \mathrm{Pf}([[\beta]]_{\{1,\dots,n\}}(t_1, \dots, t_n))$$

is a theorem of HF.

Proof. We reduce the proof to establishing Lemma 7.6, i.e., the theorem

$$\vdash \mathrm{Const}(t_1) \wedge \ldots \wedge \mathrm{Const}(t_n) \to \\ \mathrm{RepVarTermForm}(\ulcorner x_r\urcorner, t_r, [[\beta]]_{V\setminus\{r\}}(t_1, \dots, t_n), [[\beta]]_V(t_1, \dots, t_n))$$

for every $r \in V \subseteq \{1, \dots, n\}$. Recall that in forming the term $[[\beta]]_{V\setminus\{r\}}(t_1, \dots, t_n)$ all the occurrences of x_r are coded as $\ulcorner x_r\urcorner$, and when these occurrences of $\ulcorner x_r\urcorner$ which correspond to free occurrences of x_r in β are replaced by t_r, we obtain $[[\beta]]_V(t_1, \dots, t_n)$. This procedure of replacing the $\ulcorner x_r\urcorner$ by t_r is described above by RepVarTermForm.

So let us assume that Lemma 7.6 (i.e., the above implication) has been proved. Then, to prove 7.3, we first observe that **4**, **7** and the subsequent Note (Section 3) imply

$$\vdash \mathrm{Const}(t) \to \mathrm{VarNonOccTerm}(u, t).$$

Hence

$$\vdash \mathrm{Var}(v) \wedge \mathrm{Form}(y) \wedge \mathrm{Const}(t) \to \mathrm{TermSubsVarForm}(t, v, y),$$

using **20**, Lemma 3.2 and the Note after **19** (last line). Next, let us observe that

$$\vdash \mathrm{TermSubsVarForm}(t, v, y) \wedge \mathrm{RepVarTermForm}(v, t, y, y') \wedge \mathrm{Pf}(y) \to \mathrm{Pf}(y')$$

by the definitions of Subst and Pf, i.e., **(6)** and **(9)** in Section 4. Combining this with the previous theorem, we get

$$\vdash \mathrm{Const}(t) \wedge \mathrm{RepVarTermForm}(v, t, y, y') \wedge \mathrm{Pf}(y) \to \mathrm{Pf}(y').$$

Consider now this theorem for every $r = 1, \dots, n$ with

$$v = \ulcorner x_r\urcorner, \quad t = t_r, \quad y = [[\beta]]_{\{1,\dots,\bar{r}\}}(t_1, \dots, t_n), \quad y' = [[\beta]]_{\{1,\dots,r\}}(t_1, \dots, t_n),$$

where, for $r = 1$, $[[\beta]]_{\{1,\dots,\bar{r}\}} = [[\beta]]_0 = \ulcorner\beta\urcorner$. Then it follows, by the assumed 7.6 with $V = \{1, \dots, r\}$ and $r = 1, \dots, n$, that

$$\vdash \mathrm{Const}(t_1) \wedge \ldots \wedge \mathrm{Const}(t_r) \wedge \mathrm{Pf}([[\beta]]_{\{1,\dots,\bar{r}\}}(t_1, \dots, t_n)) \to \mathrm{Pf}([[\beta]]_{\{1,\dots,r\}}(t_1, \dots, t_n)).$$

Combining all of these implications, for $r = 1, \dots, n$, we get the assertion of 7.3. ■

The following two lemmas pave the ground for the proof of Lemma 7.6.

LEMMA 7.4. *Suppose τ is a term, $V \subseteq \{1, \dots, n\}$ and $r \in V$. Then*

$$\vdash \mathrm{Const}(t_1) \wedge \ldots \wedge \mathrm{Const}(t_n) \to \\ \mathrm{RepVarTermTerm}(\ulcorner x_r\urcorner, t_r, [[\tau]]_{V\setminus\{r\}}(t_1, \dots, t_n), [[\tau]]_V(t_1, \dots, t_n))$$

(see **21**, **22** in Section 3).

Proof. We use induction on the length of τ. So we have to consider the cases when τ is 0 or a variable, and the case when $\tau = \sigma \triangleleft \mu$.

A. *Case when τ is* 0. We have $[[0]]_V = 0$, for every V, and by **22**,

$$\vdash \mathrm{Var}(v) \wedge \mathrm{Term}(t) \to \mathrm{RepVarTermTerm}(v, t, 0, 0).$$

B. *Case when τ is a variable.* We have three possibilities:

(a): τ is x_i and $i \notin V$. From the definition of V-coding follows $[[x_i]]_{V\setminus\{r\}} = [[x_i]]_V = \ulcorner x_i \urcorner$. Moreover $\ulcorner x_i \urcorner \neq \ulcorner x_r \urcorner$, because $r \in V$. From **21**, **22** one obtains the general theorem

$$\vdash \mathrm{Var}(v) \wedge \mathrm{Var}(u) \wedge \mathrm{Term}(t) \wedge u \neq v \to \mathrm{RepVarTermTerm}(v, t, u, u).$$

It remains to substitute $v = \ulcorner x_r \urcorner$, $t = t_r$, $u = \ulcorner x_i \urcorner$.

(b): τ is x_i, $i \in V$ and $i \neq r$. By the definition of V-coding, we have in this case $[[x_i]]_{V\setminus\{r\}} = [[x_i]]_V = x_i$, and thus

$$[[\tau]]_{V\setminus\{r\}}(t_1, \dots, t_n) = [[\tau]]_V(t_1, \dots, t_n)) = t_i.$$

We have the general theorem

$$\vdash \mathrm{Var}(v) \wedge \mathrm{Term}(t) \wedge \mathrm{Const}(t') \to \mathrm{RepVarTermTerm}(v, t, t', t')$$

(see **4** and **21**, **22** in Section 3). It suffices to substitute here $v = \ulcorner x_r \urcorner$, $t = t_r$, $t' = t_i$.

(c): τ is x_r. We have $[[x_r]]_{V\setminus\{r\}} = \ulcorner x_r \urcorner$, $[[x_r]]_V = x_r$, whence $[[\tau]]_V(t_1, \dots, t_n) = t_r$. In the general theorem

$$\vdash \mathrm{Var}(v) \wedge \mathrm{Term}(t) \to \mathrm{RepVarTermTerm}(v, t, v, t)$$

(following directly from **21**, **22**) we can now substitute $v = \ulcorner x_r \urcorner$, $t = t_r$.

C. *Case when τ is $\sigma \triangleleft \mu$.* From **21**, **22** follows the general theorem

$$\vdash \mathrm{RepVarTermTerm}(v, t, u, u') \wedge \mathrm{RepVarTermTerm}(v, t, w, w') \to \\ \mathrm{RepVarTermTerm}(v, t, \langle \ulcorner \triangleleft \urcorner, u, w\rangle, \langle \ulcorner \triangleleft \urcorner, u', w'\rangle).$$

Let us substitute here $v = \ulcorner x_r \urcorner$, $t = t_r$, and

$$u = [[\sigma]]_{V\setminus\{r\}}(t_1, \dots, t_n), \quad u' = [[\sigma]]_V(t_1, \dots, t_n),$$
$$w = [[\mu]]_{V\setminus\{r\}}(t_1, \dots, t_n), \quad w' = [[\mu]]_V(t_1, \dots, t_n).$$

The inductive hypothesis (σ, μ being shorter than τ) leads us now to the conclusion that the left side of the above implication becomes a theorem when it is preceded by "$\mathrm{Const}(t_1) \wedge \dots \wedge \mathrm{Const}(t_n) \to$". Hence the right side, when so preceded, is a theorem as well. It remains to observe that

$$\langle \ulcorner \triangleleft \urcorner, u, w\rangle = \langle \ulcorner \triangleleft \urcorner, [[\sigma]]_{V\setminus\{r\}}, [[\mu]]_{V\setminus\{r\}}\rangle(t_1, \dots, t_n) = [[\tau]]_{V\setminus\{r\}}(t_1, \dots t_n),$$
$$\langle \ulcorner \triangleleft \urcorner, u', w'\rangle = \langle \ulcorner \triangleleft \urcorner, [[\sigma]]_V, [[\mu]]_V\rangle(t_1, \dots, t_n) = [[\tau]]_V(t_1, \dots t_n)$$

(by 7.2 and the definition of V-coding of terms). ■

LEMMA 7.5. *Suppose α is an atomic formula, $V \subseteq \{1, \dots, n\}$ and $r \in V$. Then*

$$\vdash \mathrm{Const}(t_1) \wedge \dots \wedge \mathrm{Const}(t_n) \to \\ \mathrm{RepVarTermAt}(\ulcorner x_r \urcorner, t_r, [[\alpha]]_{V\setminus\{r\}}(t_1, \dots, t_n), [[\alpha]]_V(t_1, \dots, t_n)).$$

Proof. There are terms σ, μ such that α is either $\sigma \in \mu$ or $\sigma = \mu$. Let us consider only the first case, since the proof for the second is analogous. By **23**, we have the general theorem

$$\vdash \mathrm{RepVarTermTerm}(v,t,u,u') \wedge \mathrm{RepVarTermTerm}(v,t,w,w') \to \\ \mathrm{RepVarTermAt}(v,t,\langle \ulcorner\in\urcorner, u, w\rangle, \langle \ulcorner\in\urcorner, u', w'\rangle).$$

Using the same substitutions for $v,\dots,w'$ as in the proof of Lemma 7.4, we see from Lemma 7.4 that the left side of the above implication becomes a theorem when it is preceded by "$\mathrm{Const}(t_1) \wedge \ldots \wedge \mathrm{Const}(t_n) \to$". So the same applies to the right side. It remains to observe that

$$\langle \ulcorner\in\urcorner, u, w\rangle = \langle \ulcorner\in\urcorner, [[\sigma]]_{V\setminus\{r\}}, [[\mu]]_{V\setminus\{r\}}\rangle(t_1,\dots,t_n) = [[\alpha]]_{V\setminus\{r\}}(t_1,\dots,t_n),$$
$$\langle \ulcorner\in\urcorner, u', w'\rangle = \langle \ulcorner\in\urcorner, [[\sigma]]_V, [[\mu]]_V\rangle(t_1,\dots,t_n) = [[\alpha]]_V(t_1,\dots,t_n). \;\blacksquare$$

Lemma 7.6. *Suppose β is a formula, $V \subseteq \{1,\dots,n\}$ and $r \in V$. Then*

$$\vdash \mathrm{Const}(t_1) \wedge \ldots \wedge \mathrm{Const}(t_n) \to \\ \mathrm{RepVarTermForm}(\ulcorner x_r\urcorner, t_r, [[\beta]]_{V\setminus\{r\}}(t_1,\dots,t_n), [[\beta]]_V(t_1,\dots,t_n)).$$

Proof. We fix $n \geq 1$. When β is atomic, the proof is already there. So we have to consider the cases when β is of the form $\exists x_i\gamma$, or $\gamma \vee \delta$, or $\neg\gamma$, while accepting the inductive hypothesis that the implication in 7.6 has been proved for γ and δ (and for all r, V satisfying $r \in V \subseteq \{1,\dots,n\}$).

A. *Case when β is $\exists x_i\gamma$.* As in the proof of 7.4.B, we consider various possibilities for i, r and V.

(a): $i \notin V$ and hence $\ulcorner x_i\urcorner \neq \ulcorner x_r\urcorner$. We have the general theorem

$$\vdash \mathrm{Var}(u) \wedge u \neq v \wedge \mathrm{RepVarTermForm}(v,t,y,y') \to \\ \mathrm{RepVarTermForm}(v,t,\langle \ulcorner\exists\urcorner, u, y\rangle, \langle \ulcorner\exists\urcorner, u, y'\rangle).$$

(see **24**, **25**). Let us substitute here $u = \ulcorner x_i\urcorner$, $v = \ulcorner x_r\urcorner$, $t = t_r$ and

$$y = [[\gamma]]_{V\setminus\{r\}}(t_1,\dots,t_n), \qquad y' = [[\gamma]]_V(t_1,\dots,t_n).$$

Then, by the inductive hypothesis, the left side of the above implication becomes a theorem when preceded by "$\mathrm{Const}(t_1) \wedge \ldots \wedge \mathrm{Const}(t_n) \to$". So the same applies to the right side, where, by the definition of the V-coding,

$$\langle \ulcorner\exists\urcorner, u, y\rangle = \langle \ulcorner\exists\urcorner, \ulcorner x_i\urcorner, [[\gamma]]_{V\setminus\{r\}}\rangle(t_1,\dots,t_n) \\ = [[\exists x_i\gamma]]_{V\setminus\{r\}}(t_1,\dots,t_n) = [[\beta]]_{V\setminus\{r\}}(t_1,\dots,t_n),$$

and similarly for y' and V in place of y and $V\setminus\{r\}$.

(b): $i \in V$ and $i \neq r$. Let $W = V\setminus\{i\}$. In the general implication used in the proof of (a), we make the same substitutions for u, v, t as in (a) and put

$$y = [[\gamma]]_{W\setminus\{r\}}(t_1,\dots,t_n), \qquad y' = [[\gamma]]_W(t_1,\dots,t_n).$$

Then, by the inductive hypothesis applied to γ and W, the left side of that implication, when preceded by "$\mathrm{Const}(t_1) \wedge \ldots \wedge \mathrm{Const}(t_n) \to$", becomes a theorem. So the same applies to the right side. We observe that now, by the definition of the V-coding, and since $W \setminus \{r\} = (V \setminus \{r\}) \setminus \{i\}$,

$$\begin{aligned}\langle \ulcorner\exists\urcorner, u, y\rangle &= \langle \ulcorner\exists\urcorner, \ulcorner x_i\urcorner, [[\gamma]]_{W\setminus\{r\}}\rangle(t_1,\dots,t_n)\\ &= [[\exists x_i\gamma]]_{V\setminus\{r\}}(t_1,\dots,t_n) = [[\beta]]_{V\setminus\{r\}}(t_1,\dots,t_n).\end{aligned}$$

We have a similar equality for y', W and V in place of y, $W \setminus \{r\}$ and $V \setminus \{r\}$.

(c): $i = r$. We use the general theorem

$$\vdash \mathrm{RepVarTermForm}(v,t,y,y') \to \mathrm{RepVarTermForm}(v,t,\langle\ulcorner\exists\urcorner, v, y\rangle, \langle\ulcorner\exists\urcorner, v, y\rangle).$$

Here we substitute $v = \ulcorner x_r\urcorner = \ulcorner x_i\urcorner$ and let t, y, y' be as in (a). By the definition of V-coding,

$$\begin{aligned}\langle \ulcorner\exists\urcorner, v, y\rangle &= \langle \ulcorner\exists\urcorner, \ulcorner x_r\urcorner, [[\gamma]]_{V\setminus\{r\}}\rangle(t_1,\dots,t_n)\\ &= [[\exists x_r\gamma]]_{V\setminus\{r\}}(t_1,\dots,t_n) = [[\beta]]_{V\setminus\{r\}}(t_1,\dots,t_n)\\ &= [[\exists x_i\gamma]]_V(t_1,\dots,t_n) = [[\beta]]_V(t_1,\dots,t_n).\end{aligned}$$

So we reach the conclusion, arguing similarly to cases (a) and (b).

B. *Case when β is $\gamma \vee \delta$.* We have the general theorem

$$\begin{aligned}&\vdash \mathrm{RepVarTermForm}(v,t,y,y') \wedge \mathrm{RepVarTermForm}(v,t,z,z') \to\\ &\qquad\qquad \mathrm{RepVarTermForm}(v,t,\langle\ulcorner\vee\urcorner, y, z\rangle, \langle\ulcorner\vee\urcorner, y', z'\rangle)\end{aligned}$$

(see **24**, **25**). We substitute here $v = \ulcorner x_r\urcorner$, $t = t_r$ and

$$\begin{aligned}y &= [[\gamma]]_{V\setminus\{r\}}(t_1,\dots,t_n), \quad & y' &= [[\gamma]]_V(t_1,\dots,t_n)),\\ z &= [[\delta]]_{V\setminus\{r\}}(t_1,\dots,t_n), \quad & z' &= [[\delta]]_V(t_1,\dots,t_n)).\end{aligned}$$

Then

$$\begin{aligned}\langle\ulcorner\vee\urcorner, y, z\rangle &= \langle\ulcorner\vee\urcorner, [[\gamma]]_{V\setminus\{r\}}, [[\delta]]_{V\setminus\{r\}}\rangle(t_1,\dots,t_n)\\ &= [[\gamma\vee\delta]]_{V\setminus\{r\}}(t_1,\dots t_n) = [[\beta]]_{V\setminus\{r\}}(t_1,\dots,t_n)\end{aligned}$$

and similarly for $y, z, V \setminus \{r\}$ replaced by y', z', V. We reach the conclusion by using the induction hypothesis for γ and δ.

C. *Case when β is $\neg\gamma$.* We argue as in the preceding case, using the theorem

$$\vdash \mathrm{RepVarTermForm}(v,t,y,y') \to \mathrm{RepVarTermForm}(v,t,\langle\ulcorner\neg\urcorner, y\rangle, \langle\ulcorner\neg\urcorner, y'\rangle). \blacksquare$$

Corollary 7.7. *For any formulas α, β and $n = 1, 2, \dots,$*

$$\begin{aligned}&\vdash \mathrm{Const}(t_1) \wedge \ldots \wedge \mathrm{Const}(t_n) \wedge \mathrm{Pf}(\ulcorner\alpha \to \beta\urcorner) \to\\ &\qquad\qquad \{\mathrm{Pf}([[\alpha]]_{\{1,\dots,n\}}(t_1,\dots,t_n)) \to \mathrm{Pf}([[\beta]]_{\{1,\dots,n\}}(t_1,\dots,t_n))\}.\end{aligned}$$

Proof. Suppose $\mathrm{Const}(t_i)$ for $i = 1,\dots,n$ and $\mathrm{Pf}(\ulcorner\alpha\to\beta\urcorner)$ (where $\alpha \to \beta$ is the abbreviation for $\neg\alpha \vee \beta$). Then, by 7.3,

$$\mathrm{Pf}([[\alpha\to\beta]]_{\{1,\dots,n\}}(t_1,\dots,t_r)),$$

and by the rules of V-coding this is

$$\mathrm{Pf}(\mathrm{Impl}([[\alpha]]_{\{1,\dots,n\}}(t_1,\dots,t_n), [[\beta]]_{\{1,\dots,n\}}(t_1,\dots,t_n))).$$

Since the Modus Ponens Rule of Inference is incorporated in the definition ((**8**) and (**9**)) of the formula Pf, we have the general theorem

$$\vdash \mathrm{Pf}(\mathrm{Impl}(x,y)) \to [\mathrm{Pf}(x) \to \mathrm{Pf}(y)].$$

It remains to substitute $x = [[\alpha]]_{\{1,\dots,n\}}(t_1,\dots,t_n)$ and $y = [[\beta]]_{\{1,\dots,n\}}(t_1,\dots,t_n)$. ■

8. The special p-function F

We shall presently introduce the *special* p-function F. According to the plan outlined at the beginning of the previous section, a key role will be given to the p-term $[[\alpha]]((F(x_1), \dots, F(x_n))$, obtained from the term $[[\alpha]]$ by replacing each x_i occurring in it by the p-term $F(x_i)$. In this section, after defining F, we shall show that the p-formula $\mathrm{Const}(F(x_i))$ is (or rather, abbreviates) a theorem of HF. This will allow us to apply the results of the preceding section to the situation when each t_i is replaced by $F(x_i)$.

To define F, we need to select *definably* from every non-empty set x an element v, so that then canonically $x = u \triangleleft v$, with $u = x \setminus \{v\}$. To select $v \in x$, we use an ordering relation $<$ on the universe of all (hereditarily finite) sets. The description of $<$ starts from the following observation: Any ordering of a (finite) set y determines an ordering of the power set $P(y)$ (= set of subsets of y). This is defined so that for $z_1, z_2 \subseteq y$ and $z_1 \neq z_2$ one puts $z_1 < z_2$ iff the greatest element of the symmetric difference $(z_1 \cup z_2) \setminus (z_1 \cap z_2)$ is in z_2, and otherwise, $z_2 < z_1$. Since all sets can be reached from 0 by repeated application of the power-set operation (and 0 is ordered!), the above principle yields a unique ordering of the universe of all sets. This ordering refines the partial ordering by $\subset$ and it also has the property that $x < y$ when $x \in y$ (see Ap.5 for the details). We can use the same symbol $<$ for this order and for the usual order among ordinals, since on the sub-universe of ordinals, both orders coincide.

Definition 8.1. The *special p-function* F is defined by *recursion on inequality* (see Ap.5.14) as follows:

1° $F(0) = 0$.

2° Suppose $x \neq 0$ and $F(y)$ has been defined for all $y < x$. Then put

$$F(x) = \langle \ulcorner \triangleleft \urcorner, F(u), F(v) \rangle,$$

where v is the $<$-greatest element of x and $u = x \setminus \{v\}$ (i.e., $x = u \triangleleft v$ and $v \notin u$).

The existence of v follows from Ap.5.10. Also, we have $u, v < x$, by 4.5 and 5.9 in Ap. (We leave to the reader the task of describing the p-function G required in 5.14 of Ap.) Let us list some properties of F that may justify why we called F special:

(a) In the standard model $\mathfrak{S}$, the values of F are codes (or rather, named by codes) of constant terms.

(b) The definition of F parallels the coding of terms (compare 2° in 8.1 with $\ulcorner \sigma \triangleleft \tau \urcorner = \langle \ulcorner \triangleleft \urcorner, \ulcorner \sigma \urcorner, \ulcorner \tau \urcorner \rangle$).

(c) F is injective.

(d) For each constant term τ there is a constant term μ such that $\vdash \tau = \mu$ and $F(\mu) = \ulcorner \mu \urcorner$.

(e) $F(k) = \ulcorner (\ulcorner x_k \urcorner) \urcorner$ for every ordinal $k \neq 0$. (Hence F coincides on the sub-universe of ordinals with the two p-functions W, H of Section 6.)

Property (a) follows from 8.2, (b) is obvious, and (c), (d), (e) will not be used below. There is a p-function in [Bo] that plays a similar role to that of F here, however the definition in [Bo] is simpler, because in PA each x is the successor of a unique u, so that $x = Su$ can be used instead of $x = u \lhd v$.

Theorem 8.2. $\vdash \mathrm{Const}(F(x))$.

Proof. Let us prove $\mathrm{Const}(F(x))$ by induction on $<$ (applying the scheme 5.11 of Ap). We have $\mathrm{Const}(F(0))$, due to $F(0) = 0$. Now let $x \neq 0$ and assume $\mathrm{Const}(F(y))$ for all $y < x$. Let $u, v < x$ be such that v is the $<$-largest element of x and $u = x \setminus \{v\}$, so that $F(x) = \langle \ulcorner \lhd \urcorner, F(u), F(v) \rangle$. Since then $\mathrm{Const}(F(u))$ and $\mathrm{Const}(F(v))$, by the inductive assumption, we can use the easily proved theorem

$$\vdash \mathrm{Const}(t_1) \wedge \mathrm{Const}(t_2) \rightarrow \mathrm{Const}(\langle \ulcorner \lhd \urcorner, t_1, t_2 \rangle)$$

to deduce $\mathrm{Const}(\langle \ulcorner \lhd \urcorner, F(u), F(v) \rangle)$, i.e., $\mathrm{Const}(F(x))$. ■

The theorem $\mathrm{Const}(F(x))$ allows us to replace in 7.3 and 7.7 each t_i by $F(x_i)$, thus leading to:

Lemma 8.3. *If β is a theorem and all variables free in β are among $x_1, \ldots, x_n$ then*

$$\vdash \mathrm{Pf}([[\beta]](F(x_1), \ldots, F(x_n))).$$

Proof. Assume $\vdash \beta$. Then $\vdash \mathrm{Pf}(\ulcorner \beta \urcorner)$, by 4.4. Replacing each t_i by $F(x_i)$ in 7.3 and using 8.2, we obtain

$$\vdash \mathrm{Pf}(\ulcorner \beta \urcorner) \rightarrow \mathrm{Pf}([[\beta]]_{\{1,\ldots,n\}}(F(x_1), \ldots, F(x_n))).$$

Moreover the assumption about the variables implies $[[\beta]]_{\{1,\ldots,n\}} = [[\beta]]$. ■

Lemma 8.4. *If $\alpha \rightarrow \beta$ is a theorem and all variables free in α and β are among $x_1, \ldots, x_n$ then*

$$\vdash \mathrm{Pf}([[\alpha]](F(x_1), \ldots, F(x_n))) \rightarrow \mathrm{Pf}([[\beta]](F(x_1), \ldots, F(x_n))).$$

Proof. Assume $\vdash \alpha \rightarrow \beta$. Then $\vdash \mathrm{Pf}(\ulcorner \alpha \rightarrow \beta \urcorner)$, by 4.4. Replacing each t_i by $F(x_i)$ in 7.7 and using 8.2, we obtain

$$\vdash \mathrm{Pf}(\ulcorner \alpha \rightarrow \beta \urcorner) \rightarrow \\ \{\mathrm{Pf}([[\alpha]]_{\{1,\ldots,n\}}(F(x_1), \ldots, F(x_n))) \rightarrow \mathrm{Pf}([[\beta]]_{\{1,\ldots,n\}}(F(x_1), \ldots, F(x_n)))\}.$$

Moreover the assumption about the variables implies $[[\beta]]_{\{1,\ldots,n\}} = [[\beta]]$ and $[[\alpha]]_{\{1,\ldots,n\}} = [[\alpha]]$. ■

In the above lemmas we obtained $[[\alpha]](F(x_1), \ldots, F(x_n))$ by replacing each t_i in $[[\alpha]](t_1, \ldots, t_n)$ by $F(x_i)$. However, $t_1, \ldots, t_n$ were introduced merely for the sake of convenience and clearly we can pass directly from $[[\alpha]]$ to $[[\alpha]](F(x_1), \ldots, F(x_n))$ just by replacing each x_i in $[[\alpha]]$ by $F(x_i)$.

9. Gödel's Second Incompleteness Theorem

The proof of Gödel's Second Incompleteness Theorem is based on the fact that $\alpha \to \mathrm{Pf}(\ulcorner\alpha\urcorner)$ is a theorem of HF for every Σ-sentence α. We obtain this result as a direct consequence of Theorem 9.1 whose proof occupies the major part of this section.

THEOREM 9.1. *If α is a Σ-formula and all variables free in α are among $x_1, \ldots, x_n$ then the p-formula*

$$(\star) \qquad \alpha \to \mathrm{Pf}([[\alpha]](F(x_1), \ldots, F(x_n)))$$

abbreviates a theorem of HF.

DEFINITION 9.2. A formula α will be called a $(\star)$-*formula* iff $(\star)$ abbreviates a theorem of HF for every (or equivalently, some) n such that all free variables of α are among $x_1, \ldots, x_n$.

Due to 8.4, it will suffice to show that *every strict Σ-formula is a $(\star)$-formula.* Recall that (see 2.1) the class $\mathbf{\Sigma}$ of strict Σ-formulas is the least class of formulas containing all atomic formulas of the form $x_i \in x_j$, and such that $\mathbf{\Sigma}$ is closed under conjunctions, disjunctions and the application of existential quantifiers and bounded universal quantifiers $\forall(x_i \in x_j)$, where $i \neq j$. Thus the proof is of 9.1 is reduced to showing Lemmas 9.4–9.7 below. (However we first need the auxiliary result 9.3.)

LEMMA 9.3. *$x_i = x_j$ is a $(\star)$-formula.*

Proof. We have to prove that

$$\vdash x_i = x_j \to \mathrm{Pf}([[x_i = x_j]](F(x_1), \ldots, F(x_n)))$$

provided $i, j \leq n$. Since in this case, by the definition of pseudo-coding,

$$[[x_i = x_j]](F(x_1), \ldots, F(x_n)) = \langle \ulcorner = \urcorner, F(x_i), F(x_j) \rangle,$$

we have to prove

$$(1) \qquad x_i = x_j \to \mathrm{Pf}(\langle \ulcorner = \urcorner, F(x_i), F(x_j) \rangle).$$

We have:

$\vdash x_i = x_i$.

$\vdash \mathrm{Pf}([[x_i = x_i]](F(x_1), \ldots, F(x_n)))$ when $i \leq n$, by 8.3.

$\vdash \mathrm{Pf}(\langle \ulcorner = \urcorner, F(x_i), F(x_i) \rangle)$, by the definition of pseudo-coding.

$\vdash x_i = x_j \to \mathrm{Pf}(\langle \ulcorner = \urcorner, F(x_i), F(x_j) \rangle)$, by the rules of logic. ∎

LEMMA 9.4. *$x_i \in x_j$ is a $(\star)$-formula.*

Proof. We have to show

$$\vdash x_i \in x_j \to \mathrm{Pf}([[x_i \in x_j]](F(x_1), \ldots, F(x_n)))$$

provided $i, j \leq n$. Similarly to the case of the formula $x_i = x_j$, this amounts to proving

$$\vdash x_i \in x_j \to \mathrm{Pf}(\langle \ulcorner \in \urcorner, F(x_i), F(x_j) \rangle).$$

There will be no loss of generality if we take here $i = 1$, $j = 2 = n$, so let $\beta(x_2)$ be the formula

$$x_1 \in x_2 \to \mathrm{Pf}(\langle \ulcorner\in\urcorner, F(x_1), F(x_2)\rangle).$$

We shall prove $\forall x_2 \beta(x_2)$ by induction on $<$ (applying the scheme 5.11 in Ap) with respect to x_2, keeping x_1 fixed. In this proof we shall need the theorem

$$\text{(2)} \quad \vdash F(x_2) = \langle \ulcorner\lhd\urcorner, F(x_3), F(x_4)\rangle \to \{\mathrm{Pf}(\langle \ulcorner\in\urcorner, F(x_1), F(x_3)\rangle) \vee \mathrm{Pf}(\langle \ulcorner=\urcorner, F(x_1), F(x_4)\rangle) \to \mathrm{Pf}(\langle \ulcorner\in\urcorner, F(x_1), F(x_2)\rangle)\}.$$

To prove (2), note first the theorems

$$\vdash \mathrm{Pf}(z_i) \to \mathrm{Pf}(\langle \ulcorner\vee\urcorner, z_1, z_2\rangle) \qquad (i = 1, 2),$$

which follow directly from our definition of the Pf formula (see second Boolean Axiom in 4.1 and the Modus Ponens Rule in 4.2). Hence

$$\vdash \mathrm{Pf}(z_1) \vee \mathrm{Pf}(z_2) \to \mathrm{Pf}(\langle \ulcorner\vee\urcorner, z_1, z_2\rangle).$$

We substitute here $z_1 = \langle \ulcorner\in\urcorner, F(x_1), F(x_3)\rangle$ and $z_2 = \langle \ulcorner=\urcorner, F(x_1), F(x_4)\rangle$ and observe that then

$$\begin{aligned}\langle \ulcorner\vee\urcorner, z_1, z_2\rangle &= \langle \ulcorner\vee\urcorner, \langle \ulcorner\in\urcorner, F(x_1), F(x_3)\rangle, \langle \ulcorner=\urcorner, F(x_1), F(x_4)\rangle\rangle \\ &= [[x_1 \in x_3 \vee x_1 = x_4]](F(x_1), F(x_2), F(x_3), F(x_4)).\end{aligned}$$

This leads to

$$\text{(3)} \quad \vdash \mathrm{Pf}(\langle \ulcorner\in\urcorner, F(x_1), F(x_3)\rangle) \vee \mathrm{Pf}(\langle \ulcorner=\urcorner, F(x_1), F(x_4)\rangle) \to \mathrm{Pf}([[x_1 \in x_3 \vee x_1 = x_4]](F(x_1), F(x_2), F(x_3), F(x_4))).$$

Moreover

$$\text{(4)} \quad \vdash \mathrm{Pf}([[x_1 \in x_3 \vee x_1 = x_4]](F(x_1), F(x_2), F(x_3), F(x_4))) \to \mathrm{Pf}([[x_1 \in x_3 \lhd x_4]](F(x_1), F(x_2), F(x_3), F(x_4)))$$

by the theorem $x_1 \in x_3 \vee x_1 = x_4 \to x_1 \in x_3 \lhd x_4$ and 8.4. Combining (3) and (4), we obtain

$$\vdash \mathrm{Pf}(\langle \ulcorner\in\urcorner, F(x_1), F(x_3)\rangle) \vee \mathrm{Pf}(\langle \ulcorner=\urcorner, F(x_1), F(x_4)\rangle) \to \mathrm{Pf}(\langle \ulcorner\in\urcorner, F(x_1), \langle \ulcorner\lhd\urcorner, F(x_3), F(x_4)\rangle\rangle).$$

So (2) follows.

Returning now to the proof of $\forall x_2 \beta(x_2)$, by induction on $>$, we note that it will suffice to show that

$$\text{(5)} \qquad \forall(x_3 < x_2)\beta(x_3) \to \beta(x_2)$$

is a theorem. Let us assume $\forall(x_3 < x_2)\beta(x_3)$. If $x_2 = 0$ then (5) is a theorem, because $\beta(0)$ is a theorem (being an implication with the premise $x_1 \in 0$). If $x_2 \neq 0$, let $x_3 < x_2$ and x_4 be such that $x_2 = x_3 \lhd x_4$ and $F(x_2) = \langle \ulcorner\lhd\urcorner, F(x_3), F(x_4)\rangle$ (see 8.1). Then

$$x_1 \in x_2 \to x_1 \in x_3 \vee x_1 = x_4 \to \mathrm{Pf}(\langle \ulcorner\in\urcorner, F(x_1), F(x_3)\rangle) \vee \mathrm{Pf}(\langle \ulcorner=\urcorner, F(x_1), F(x_4)\rangle)$$

by the assumption (since $x_3 < x_2$) and (1) in 9.3. Combining this with (2), we get $\beta(x_2)$, so (5) is shown. ∎

Lemma 9.5. *If α and β are $(\star)$-formulas then $\alpha \vee \beta$ and $\alpha \wedge \beta$ are $(\star)$-formulas.*

Proof. Suppose all variables free in $\alpha \vee \beta$ are among $x_1, \dots, x_n$. Then we should show

(a) $\vdash \alpha \vee \beta \to \mathrm{Pf}([[\alpha \vee \beta]](F(x_1), \dots, F(x_n)))$,
(b) $\vdash \alpha \wedge \beta \to \mathrm{Pf}([[\alpha \wedge \beta]](F(x_1), \dots, F(x_n)))$.

To get (a), note that, by assumption,

$$\vdash \alpha \vee \beta \to \mathrm{Pf}([[\alpha]](F(x_1), \dots, F(x_n))) \vee \mathrm{Pf}([[\beta]](F(x_1), \dots, F(x_n))).$$

It now remains to apply the theorem $\mathrm{Pf}(z_1) \vee \mathrm{Pf}(z_2) \to \mathrm{Pf}(\langle \ulcorner \vee \urcorner, z_1, z_2 \rangle)$ with $z_1 = [[\alpha]](F(x_1), \dots, F(x_n))$, $z_2 = [[\beta]](F(x_1), \dots, F(x_n))$, noting that in this case

$$\langle \ulcorner \vee \urcorner, z_1, z_2 \rangle = [[\alpha \vee \beta]](F(x_1), \dots, F(x_n)).$$

The proof of part (b) is analogous. ■

Lemma 9.6. *If α is a $(\star)$-formula then, for every i, $\exists x_i \alpha$ is a $(\star)$-formula. In other words, if all variables free in $\exists x_i \alpha$ are among $x_1, \dots, x_n$ then*

$$\vdash \exists x_i \alpha \to \mathrm{Pf}([[\exists x_i \alpha]](F(x_1), \dots, F(x_n))).$$

Proof. Let $m \geq n$ be such that all variables free in α are among $x_1, \dots, x_m$. Then

$$\vdash \mathrm{Pf}([[\alpha]](F(x_1), \dots, F(x_m))) \to \mathrm{Pf}([[\exists x_i \alpha]](F(x_1), \dots, F(x_m)))$$

by the Specialization Axiom $\vdash \alpha \to \exists x_i \alpha$ and 8.4. Combining this with the assumption $(\star)$, we obtain

$$\vdash \alpha \to \mathrm{Pf}([[\exists x_i \alpha]](F(x_1), \dots, F(x_m))).$$

Since x_i is not free on the right hand side, we can apply the $\exists$-introduction rule (in 4.2), to get

$$\vdash \exists x_i \alpha \to \mathrm{Pf}([[\exists x_i \alpha]](F(x_1), \dots, F(x_m))).$$

All the variables free in $\exists x_i \alpha$ are among $x_1, \dots, x_n$, so we can replace here m by n. ■

Lemma 9.7. *If α is a $(\star)$-formula then, for every $i \neq j$, $\forall (x_i \in x_j) \alpha$ is a $(\star)$-formula. In other words, if all variables free in $\forall (x_i \in x_j) \alpha$ are among $x_1, \dots, x_n$ then*

$$\vdash \forall (x_i \in x_j) \alpha \to \mathrm{Pf}([[\forall (x_i \in x_j) \alpha]](F(x_1), \dots, F(x_n))).$$

Proof. Let α be a $(\star)$-formula. The following two observations will be used in the proof:

(a) If $\beta \leftrightarrow \alpha$ then β is a $(\star)$-formula (in virtue of 8.4).

(b) If a variable x_k is replaced in α by some variable x_l (substitutable for x_k in α) then the formula so obtained is also a $(\star)$-formula.

To simplify matters, we take $i = 1$, $j = 2$ and represent by x_3 the remaining free variables in α. We shall distinguish two cases:

Case 1: x_2 is not free in α. We shall write α as $\alpha(x_1)$. It will be also convenient to take $n = 5$ and assume that x_5 is substitutable for x_1 (otherwise, we can rename the bound variables of α so that x_5 will not be among them and then replace α by the formula β so obtained; see (a) above).

Thus our purpose is to prove $\forall x_2 \gamma(x_2)$ for the formula $\gamma(x_2)$:

$$\forall (x_1 \in x_2) \alpha(x_1) \to \mathrm{Pf}([[\forall (x_1 \in x_2) \alpha(x_1)]](F(x_1), \dots, F(x_5))),$$

under the assumption

$$\vdash \alpha(x_1) \to \mathrm{Pf}([[\alpha(x_1)]](F(x_1), \ldots, F(x_5))). \tag{6}$$

The proof, by induction on $<$, will be somewhat analogous to the proof of $\forall x_2 \beta(x_2)$ in 9.4. We begin by showing

$$\begin{aligned} (7) \qquad & \vdash F(x_2) = \langle \ulcorner \lhd \urcorner, F(x_4), F(x_5) \rangle \to \\ & \{\mathrm{Pf}([[\forall(x_1 \in x_4)\alpha(x_1)]](F(x_1), \ldots, F(x_5))) \wedge \mathrm{Pf}([[\alpha(x_5)]](F(x_1), \ldots, F(x_5))) \to \\ & \mathrm{Pf}([[\forall(x_1 \in x_2)\alpha(x_1)]](F(x_1), \ldots, F(x_5)))\}. \end{aligned}$$

To do this, note first that

$$\vdash \mathrm{Pf}(z_1) \wedge \mathrm{Pf}(z_2) \to \mathrm{Pf}(\mathrm{Conj}(z_1, z_2)), \tag{8}$$

by our definition of the Pf formula in Section 4. Let us substitute here

$$\begin{aligned} z_1 &= [[\forall(x_1 \in x_4)\alpha(x_1)]](F(x_1), \ldots, F(x_5)), \\ z_2 &= [[\alpha(x_5)]](F(x_1), \ldots, F(x_5)) \end{aligned}$$

and note that then

$$\mathrm{Conj}(z_1, z_2) = [[\forall(x_1 \in x_4)\alpha(x_1) \wedge \alpha(x_5)]](F(x_1), \ldots, F(x_5)),$$

by our definition of pseudo-coding. Together with (8) this leads to

$$\begin{aligned} (9) \qquad & \vdash \mathrm{Pf}([[\forall(x_1 \in x_4)\alpha(x_1)]](F(x_1), \ldots, F(x_5))) \wedge \mathrm{Pf}([[\alpha(x_5)]](F(x_1), \ldots, F(x_5))) \to \\ & \mathrm{Pf}([[\forall(x_1 \in x_4)\alpha(x_1) \wedge \alpha(x_5)]](F(x_1), \ldots, F(x_5))). \end{aligned}$$

Since x_5 is substitutable for x_1 in α, one easily proves

$$\vdash \forall(x_1 \in x_4)\alpha(x_1) \wedge \alpha(x_5) \to \forall(x_1 \in x_4 \lhd x_5)\alpha(x_1).$$

Hence

$$\begin{aligned} (10) \qquad & \vdash \mathrm{Pf}([[\forall(x_1 \in x_4)\alpha(x_1) \wedge \alpha(x_5)]](F(x_1), \ldots, F(x_5))) \to \\ & \mathrm{Pf}([[\forall(x_1 \in x_4 \lhd x_5)\alpha(x_1)]](F(x_1), \ldots, F(x_5))), \end{aligned}$$

due to 8.4. Finally, from the definition of pseudo-coding, and under the assumption $F(x_2) = \langle \ulcorner \lhd \urcorner, F(x_4), F(x_5) \rangle$, we obtain

$$\begin{aligned} & [[\forall(x_1 \in x_4 \lhd x_5)\alpha(x_1)]](F(x_1), \ldots, F(x_5)) = \\ & [[\forall x_1(x_1 \in x_4 \lhd x_5 \to \alpha(x_1))]](F(x_1), \ldots, F(x_5)) = \\ & \mathrm{All}(\ulcorner x_1 \urcorner, \mathrm{Impl}(\langle \ulcorner \in \urcorner, \ulcorner x_1 \urcorner, \langle \ulcorner \lhd \urcorner, F(x_4), F(x_5) \rangle \rangle, [[\alpha(x_1)]]_{\{3\}}(F(x_1), \ldots, F(x_5)))) = \\ & \mathrm{All}(\ulcorner x_1 \urcorner, \mathrm{Impl}(\langle \ulcorner \in \urcorner, \ulcorner x_1 \urcorner, F(x_2) \rangle, [[\alpha(x_1)]]_{\{3\}}(F(x_1), \ldots, F(x_5)))) = \\ & [[\forall(x_1 \in x_2)\alpha(x_1)]](F(x_1), \ldots, F(x_5)). \end{aligned}$$

This calculation, together with (9) and (10), yields the desired result (7).

Returning now to the proof of $\forall x_2 \gamma(x_2)$ by induction on $<$, we recall (see 5.1 in Ap) that it will suffice to establish that

$$\forall(x_4 < x_2)\gamma(x_4) \to \gamma(x_2) \tag{11}$$

is a theorem. Thus let us assume $\forall(x_4 < x_2)\gamma(x_4)$.

Suppose $x_2 = 0$. Then the proof of (11) obviously reduces to verifying $\vdash \gamma(0)$. To do this, we note first that $\vdash \forall(x_1 \in 0)\alpha(x_1)$, whence by 8.3,

$$\vdash \mathrm{Pf}([[\forall(x_1 \in 0)\alpha(x_1)]](F(x_1), \dots, F(x_5))).$$

Now, since $\forall(x_1 \in x_2)\alpha(x_1)$ is just an abbreviation of $\forall x_1(x_1 \in x_2 \to \alpha(x_1))$, we easily check that

$$[[\forall(x_1 \in x_2)\alpha(x_1)]](F(x_1), F(0), F(x_3), F(x_4), F(x_5)) =$$
$$[[\forall(x_1 \in 0)\alpha(x_1)]](F(x_1), F(x_2), F(x_3), F(x_4), F(x_5))$$

(as both sides of this equality are equal to

$$\mathrm{All}(\ulcorner x_1 \urcorner, \mathrm{Impl}(\langle \ulcorner \in \urcorner, \ulcorner x_1 \urcorner, 0\rangle, [[\alpha(x_1)]]_{\{3\}}(F(x_1), \dots, F(x_5)))),$$

because $F(0) = 0$). Hence

$$\vdash \mathrm{Pf}([[\forall(x_1 \in x_2)\alpha(x_1)]](F(x_1), F(0), F(x_3), F(x_4), F(x_5))),$$

which shows that $\vdash \gamma(0)$. Thus (11) is a theorem when $x_2 = 0$.

Now suppose $x_2 \neq 0$ and let $x_4, x_5 < x_2$ be such that $x_2 = x_4 \triangleleft x_5$ and $F(x_2) = \langle \ulcorner \triangleleft \urcorner, F(x_4), F(x_5)\rangle$ (see 8.1). Then $\gamma(x_4)$ by assumption, whence

$$\forall(x_1 \in x_4)\alpha(x_1) \to \mathrm{Pf}([[\forall(x_1 \in x_4)\alpha(x_1)]](F(x_1), \dots, F(x_5)))$$

(since x_2 is not free in α). Moreover, replacing x_1 by x_5 in (6) we conclude

$$\vdash \alpha(x_5) \to \mathrm{Pf}([[\alpha(x_5)]](F(x_1), \dots, F(x_5)))$$

(see property (b) above). Hence

$$\forall(x_1 \in x_2)\alpha(x_1) \to \forall(x_1 \in x_4)\alpha(x_1) \wedge \alpha(x_5) \to$$
$$\mathrm{Pf}([[\forall(x_1 \in x_4)\alpha(x_1)]](F(x_1), \dots, F(x_5))) \wedge \mathrm{Pf}([[\alpha(x_5)]](F(x_1), \dots, F(x_5))).$$

Combining this sequence of implications with (7), we obtain $\gamma(x_2)$.

Case 2: x_2 is free in α. The assumption is now that $\alpha(x_1, x_2, x_3)$ is a $(\star)$-formula. Consider a variable that is substitutable for x_2 in α, say x_4. Then, by virtue of (b) above, also $\alpha(x_1, x_4, x_3)$ is a $(\star)$-formula. Treating now the pair x_4, x_3 the same way as x_3 was treated in Case 1 we conclude by the same considerations as in Case 1 that $\forall(x_1 \in x_2)\alpha(x_1, x_4, x_3)$ is a $(\star)$-formula. Hence, applying (b) again, we see that $\forall(x_1 \in x_2)\alpha(x_1, x_2, x_3)$ is a $(\star)$-formula. ■

Let Con(HF) be a sentence stating that HF is a consistent theory. For instance, this could be: "$0 \in 0$ is not provable in HF". To express this in HF, we may take Con(HF) to be $\neg\,\mathrm{Pf}(\ulcorner 0 \in 0 \urcorner)$.

Theorem 9.8 (Gödel's Second Incompleteness Theorem). *If* HF *is a consistent theory, then* Con(HF) *is not a theorem of* HF.

Proof. We take the sentence δ satisfying

$$(\triangle) \qquad \vdash \delta \leftrightarrow \neg\,\mathrm{Pf}(\ulcorner \delta \urcorner)$$

(see 6.5) and we show that

$$(\triangle\triangle) \qquad \vdash \mathrm{Con(HF)} \to \delta.$$

Then, since δ is not a theorem (see 6.5), neither is Con(HF). To show ($\triangle\triangle$), we apply 9.1 to the Σ-sentence $\mathrm{Pf}(\ulcorner\delta\urcorner)$:

$$\vdash \mathrm{Pf}(\ulcorner\delta\urcorner) \rightarrow \mathrm{Pf}(\ulcorner\mathrm{Pf}(\ulcorner\delta\urcorner)\urcorner).$$

On the other hand, applying 8.4 to ($\triangle$), we deduce

$$\vdash \mathrm{Pf}(\ulcorner\delta\urcorner) \rightarrow \mathrm{Pf}(\ulcorner\neg\,\mathrm{Pf}(\ulcorner\delta\urcorner)\urcorner).$$

The last two theorems yield

$$\vdash \mathrm{Pf}(\ulcorner\delta\urcorner) \rightarrow \mathrm{Pf}(\ulcorner\mathrm{Pf}(\ulcorner\delta\urcorner)\urcorner) \wedge \mathrm{Pf}(\ulcorner\neg\,\mathrm{Pf}(\ulcorner\delta\urcorner)\urcorner).$$

By the definition of Pf, we have, for any formulas α, β,

$$\vdash \mathrm{Pf}(\ulcorner\alpha\urcorner) \wedge \mathrm{Pf}(\ulcorner\beta\urcorner) \rightarrow \mathrm{Pf}(\ulcorner\alpha \wedge \beta\urcorner),$$

so we conclude from the above:

$$\vdash \mathrm{Pf}(\ulcorner\delta\urcorner) \rightarrow \mathrm{Pf}(\ulcorner\mathrm{Pf}(\ulcorner\delta\urcorner) \wedge \neg\,\mathrm{Pf}(\ulcorner\delta\urcorner)\urcorner).$$

Since $\vdash \mathrm{Pf}(\ulcorner\delta\urcorner) \wedge \neg\,\mathrm{Pf}(\ulcorner\delta\urcorner) \rightarrow 0 \in 0$, we can use 8.4 to obtain

$$\vdash \mathrm{Pf}(\ulcorner\delta\urcorner) \rightarrow \mathrm{Pf}(\ulcorner 0 \in 0\urcorner),$$

that is,

$$\vdash \mathrm{Pf}(\ulcorner\delta\urcorner) \rightarrow \neg\,\mathrm{Con(HF)}.$$

Hence

$$\vdash \mathrm{Con(HF)} \rightarrow \neg\,\mathrm{Pf}(\ulcorner\delta\urcorner).$$

Combining this with ($\triangle$), we obtain ($\triangle\triangle$). ■

This Incompleteness Phenomenon still attracts very much attention. Apart from the publications listed below (consulted during the preparation of this paper) one can find a more extensive list of references in [L].

10. Finite sets and arithmetic

We conclude this paper by stating a theorem which says that each of the theories HF and PA is an extension by definitions of the other. In detail: If one adds to PA the symbols $\in$ and $\triangleleft$ and one views $x \in y$ and $x \in y \triangleleft z$ as abbreviations of certain formulas of PA (the formulas ($\in$) and ($\triangleleft$) in 10.1) then the axioms of HF can be proved in PA. Conversely, if one adds to HF the function symbols S (successor), + and ×, and one views $Sx = y$, $x + y = z$ and $x \times y = z$ as abbreviations of certain formulas of HF then all axioms of PA become theorems of HF. Most of these facts are discussed in [Be], [My] and [TG] (although in [Be] this mutual relationship between PA and HF is not fully worked out, in [TG] the results are merely stated and in [My] the work concerns the model ω of PA).

This situation would not exclude the possibility that arithmetic, when so defined in HF, is stronger than PA (i.e., that some theorem of HF, after the elimination of $\in$ and $\triangleleft$ in favor of +, × and S, becomes a statement about numbers not provable in PA). The opposite situation (where PA and HF are interchanged) might be possible as well.

That neither of these can happen (at least when ($\in$) in 10.1 below is used for defining the relation $x \in y$ in PA) is a consequence of the following result.

THEOREM 10.1. *Let $\mathcal{T}$ be the theory, in the language* $\{0, =, +, \times, S, \in, \lhd\}$, *axiomatized by the axioms of* PA *and two further axioms* (*where $z < y$ abbreviates $\exists w(z + Sw = y)$*):

($\in$) $x \in y \leftrightarrow \exists(u < y)\exists(v < y)[y = v + 2^x + u \times 2^{Sx} \wedge v < 2^x]$
(*that is, x occurs as an exponent when y is written as a sum of powers of* 2),

($\lhd$) $x \in y \lhd z \leftrightarrow x \in y \vee x = z.$

Then there exist formulas $\alpha(x, y)$, $\beta(x, y, z)$ and $\gamma(x, y, z)$ in the language of HF *such that another axiomatization of $\mathcal{T}$ is provided by the axioms* HF1, HF2, *the axiom scheme* HF3 *and three further axioms*:

(S) $Sx = y \leftrightarrow \alpha(x, y)$,

($+$) $x + y = z \leftrightarrow \beta(x, y, z)$,

($\times$) $x \times y = z \leftrightarrow \gamma(x, y, z).$

Moreover α, β, γ are unique up to equivalence in HF.

Thus PA and HF have the common extension $\mathcal{T}$ by definitions:

$$\mathcal{T} = \mathrm{PA} + (\in) + (\lhd) = \mathrm{FS} + (S) + (+) + (\times).$$

The above definition ($\in$) of the relation $\in$ in arithmetic is due to Ackermann [A]. The proof of 10.1 is too tedious to be included here; let us merely indicate how α, β and γ can be obtained. Suppose $\mathbb{E} : \mathbb{S} \to \omega$ is the "enumeration" by natural numbers of the universe $\mathbb{S}$ of the standard model $\mathfrak{S}$ of HF (see Ap.6), satisfying

1° $\mathbb{E}(0) = 0$,
2° $\mathbb{E}(\{a_0, \ldots, a_n\}) = 2^{\mathbb{E}(a_0)} + \ldots + 2^{\mathbb{E}(a_n)}$

for any distinct $a_0, \ldots, a_n \in \mathbb{S}$. It is not hard to formalize in HF this description of $\mathbb{E}$ so as to get a p-function E (in the sense of Section 5) which maps bijectively the universe of HF onto its sub-universe of ordinal numbers. Then we define

$$\alpha(x, y) \leftrightarrow E(x) \lhd E(x) = E(y),$$
$$\beta(x, y, z) \leftrightarrow E(x) \oplus E(y) = E(z),$$
$$\gamma(x, y, z) \leftrightarrow E(x) \otimes E(y) = E(z),$$

where $\oplus$ and $\otimes$ denote addition and multiplication of ordinals in HF.

11. A theorem of Reinhardt

This section deals with a question raised by J. Mycielski, whether there exists a theorem which is of moderate length (so that it can be written down), whilst each proof of it is so long that it can never be written down. W. Reinhardt has indicated (verbal communication) how to construct a theorem of this kind and we shall work out here some of the details of this construction. In its original form this result concerned PA but we believe that the proof of Theorem 11.1 would become much more cumbersome if we were to replace HF by PA.

Theorem 11.1. *If* HF *is consistent, then there exists a theorem* δ *of* HF *which can be written down, yet such that no proof of* δ *can be written down* (*because of its immense length*).

Proof. The proof is based on Theorem 6.4 (Gödel's Diagonal Lemma). We conclude from Gödel's Lemma that there exists, for any fixed ordinal, a sentence δ asserting about itself that it has no proof whose code is less (in the sense of the inequality $<$ in HF) than that ordinal. It then remains to attend to the details, to ensure that δ is of reasonably moderate length, HF $\vdash \delta$ and there does not exist a proof of δ of moderate length. To be more precise, we start with

Definition 11.2. The *length* of a string of symbols, such as a term, a formula or a proof, is the total number of symbols occurring in the string. Here, in counting the symbols, we disregard the parentheses and commas. Every re-occurring symbol is counted at every occurrence.

Definition 11.3. It will be said that a formula α is of *strictly moderate length*, or simply *strictly moderate*, if α can be practically written down using an imaginably limited amount of paper or time. A formula that is HF-equivalent to a strictly moderate one will be called *moderate.*

This definition is not at all precise, so it will be left to the reader to agree with us when we call a given formula "moderate". Here is an example: Consider the ordinal-valued exponential p-function $\exp_2(x)$ to the base 2, i.e., the p-function satisfying $\exp_2(x) = \exp_2(\overline{x}) + \exp_2(\overline{x})$ for every non-zero ordinal x and $\exp_2(x) = 1$ when $x = 0$ or x is not an ordinal. Then $\exp_2$ is defined recursively on ordinals (Ap.3.3) and thus $y = \exp_2(x)$ may be viewed as an abbreviation of a moderate formula. The same may be said about the 6-fold superposition

$$y = \exp_2(\ldots(\exp_2(x))))))).$$

Let K be the constant term for which $K = \exp_2(\ldots(\exp_2(2)))))))$ is a theorem of HF. Then K denotes a very large ordinal number, in fact $K > 10^{50}$, which supposedly exceeds the number of elementary particles in the Universe. Since only 0, $\lhd$ and parentheses may appear in any constant term, it is impossible to write K explicitly. Thus $y = K$ is a formula of HF that is not strictly moderate. But $y = K$ is evidently moderate since it is HF-equivalent to a strictly moderate formula resulting from a 6-fold application of the moderate formula abbreviated by $y = \exp_2(x)$. We shall need below the formula $z = R(4K + 17)$, and a discussion similar to the above (based on the recursive definition of R in Ap.3.3) shows that this one is moderate as well.

We now claim that for proving Reinhardt's theorem it is enough to find a moderate formula $\alpha(x)$, with one free variable x, such that for every sentence δ:

(a) $\mathfrak{S} \models \alpha(\ulcorner\delta\urcorner) \;\Rightarrow\; \vdash \delta$. [Recall that $\mathfrak{S}$ = standard model of HF.]

(b) *If* $\vdash \delta$ *and* δ *has a proof of length less than* K, *then* $\mathfrak{S} \models \alpha(\ulcorner\delta\urcorner)$.

(c) $\mathfrak{S} \models \neg\alpha(\ulcorner\delta\urcorner) \;\Rightarrow\; \vdash \neg\alpha(\ulcorner\delta\urcorner)$.

Indeed, given such $\alpha(x)$, let δ be the sentence obtained in the proof of the Diagonal Lemma (6.4 above), when applied to $\neg\alpha(x)$, so that

$$\vdash \delta \leftrightarrow \neg\alpha(\ulcorner\delta\urcorner). \tag{1}$$

Then δ is moderate as well. Moreover $\mathfrak{S} \models \alpha(\ulcorner\delta\urcorner)$ is impossible, by (a) and (1). Thus $\mathfrak{S} \models \neg\alpha(\ulcorner\delta\urcorner)$, so (c) and (1) imply $\vdash \delta$. Also, no proof of δ can be of length less than K, by (b).

Let us show that (a), (b), (c) are statisfied when $\alpha(x)$ is

$$\exists(k \in K)\exists(s \in R(4K+17))\mathrm{Prf}(s,k,x). \tag{2}$$

[Note that $\vdash k \in K \leftrightarrow \mathrm{Ord}(k) \wedge k < K$.] Then $\alpha(x)$ is equivalent to

$$\forall y\forall z[(y = K \wedge z = R(4K+17)) \rightarrow \exists(k \in y)\exists(s \in z)\mathrm{Prf}(s,k,x)],$$

so we see from the above discussion of the formulas $y = K$ and $z = R(4K+17)$ that $\alpha(x)$ is moderate. Since condition (a) is now a direct consequence of 4.4, it remains to verify (b) and (c).

Proof of (b). We need to show that if δ has a proof of length less than K then there exists a proof $\delta_0, \ldots, \delta_n$ of δ such that the sequence

$$s = \{\langle 0, \ulcorner\delta_0\urcorner\rangle, \ldots, \langle n, \ulcorner\delta_n\urcorner\rangle\} \tag{3}$$

satisfies $s \in R(4K+17)$ and $n^+ < K$. The next two lemmas will serve this purpose.

Lemma 11.4. *Let w be a term or formula of length l, and suppose that the index j of every variable x_j occurring in w does not exceed m. Then* $\mathrm{rank}(\ulcorner w\urcorner) \leq 2l + m + 12$.

Proof. First, we deduce directly from the definition of rank that $\mathrm{rank}(j) = j$ for all ordinals j (see Ap.4.4 and 4.5). Let $l = 1$. Then w is 0 or x_j and $\ulcorner w\urcorner = 0$ or $\ulcorner w\urcorner = j \leq m$. Since $\mathrm{rank}(j) = j$, the assertion is true. We now proceed by induction on l, noting first that $\langle x,y\rangle = \{\{x\},\{x,y\}\}$ and $\langle x,y,z\rangle = \langle x,\langle y,z\rangle\rangle$ imply

$$\mathrm{rank}(\langle x,y\rangle) \leq \max(\mathrm{rank}(x),\mathrm{rank}(y)) + 2, \tag{4}$$

$$\mathrm{rank}(\langle x,y,z\rangle) \leq \max(\mathrm{rank}(x),\mathrm{rank}(y),\mathrm{rank}(z)) + 4. \tag{5}$$

If $l > 1$, then $\ulcorner w\urcorner$ is $\langle x,y\rangle$ or $\langle x,y,z\rangle$, where x,y,z are codes of certain sub-strings of the string w. If $\ulcorner w\urcorner$ is $\langle x,y\rangle$, then $x = \ulcorner\neg\urcorner$, whence $\mathrm{rank}(x) = 10$ (see Section 1), moreover the length of the sub-string of w whose code in y is $l-1$. By the inductive assumption, $\mathrm{rank}(y) \leq 2(l-1) + m + 12$, and $\mathrm{rank}(\ulcorner w\urcorner) \leq 2l + m + 12$ follows from (4). If $\ulcorner w\urcorner = \langle x,y,z\rangle$, then $l > 2$, x is one of the codes $\ulcorner\in\urcorner, \ulcorner\triangleleft\urcorner, \ldots, \ulcorner\exists\urcorner$ and y,z are codes of sub-strings of w of length at most $l-2$. Hence $\mathrm{rank}(x) \leq \mathrm{rank}(\ulcorner\exists\urcorner) \leq 12$, and using the inductive assumption and (5), we get $\mathrm{rank}(\ulcorner w\urcorner) \leq 2l + m + 12$. ∎

Lemma 11.5. *If each formula δ_i in the sequence s (see* (3)*) is of length at most l and the index j of any variable x_j occurring in any δ_i does not exceed m, then $s < 2l+m+n+16$.*

Proof. One first shows by induction on ordinals n that if $\mathrm{rank}(x) = n$ then $x < n^+$ (this is a consequence of $\mathrm{rank}(n) = n$ and $R(n) < R(n^+) \setminus R(n)$ proved in Ap.5.8 and 5.9). In other words, $x < [\mathrm{rank}(x)]^+$ for all x. Thus it will be enough to show that $\mathrm{rank}(s) \leq 2l + m + n + 15$. Each element of s is a pair $\langle i, \ulcorner\delta_i\urcorner\rangle$, where $i \leq n$ and

$\mathrm{rank}(\ulcorner\delta_i\urcorner) \le 2l + m + 12$, by 11.4. Thus, by (4), each element of s is of rank at most $2l + m + n + 14$, which implies that $\mathrm{rank}(s) \le 2l + m + n + 15$. ■

To conclude the proof of (b), assume that δ has a proof $\delta_0, \dots, \delta_n$ shorter than K. Such proof must employ less than K variables, so replacing these variables by others, if need be, we may assume that the indices j of the variables x_j occurring in the formulas δ_i are less than K. Also, it is clear that $n^+ < K$ and that the length of each formula δ_i is less than K. Thus the assumptions of 11.5 are satisfied with $l < K$, $m < K$ and $n^+ < K$. It now follows from 11.5 that the sequence s (see (3)) satisfies $s < 4K + 16$. The theorems $y < x \to y \in R(\mathrm{rank}(x)^+)$ (see proof of Ap.5.12) and $\mathrm{Ord}(j) \to \mathrm{rank}(j) = j$ now imply that $s \in R(4K + 17)$. Hence $\mathfrak{S} \vDash \alpha(\ulcorner\delta\urcorner)$.

Proof of (c). Clearly, $\neg\alpha(x)$ is equivalent to

$$\forall(k \in K)\forall(s \in R(4K + 17))\neg\,\mathrm{Prf}(s, k, x).$$

It follows from the construction of the standard model $\mathfrak{S}$ (Ap.6) that there are constant terms σ and τ naming K and $R(4K + 17)$ respectively. Thus Lemma 6.2 in the Appendix implies that there are constant terms $\sigma_1, \dots, \sigma_l$ and $\tau_1, \dots, \tau_m$ such that

$$\vdash \neg\alpha(\ulcorner\delta\urcorner) \;\leftrightarrow\; \bigwedge_{\substack{1\le i\le m\\ 1\le j\le l}} \neg\,\mathrm{Prf}(\tau_i, \sigma_j, \ulcorner\delta\urcorner) \tag{6}$$

(see (§§) in the proof of 2.5 above). Consequently $\mathfrak{S} \vDash \neg\alpha(\ulcorner\delta\urcorner)$ yields $\mathfrak{S} \vDash \neg\,\mathrm{Prf}(\tau_i, \sigma_j, \ulcorner\delta\urcorner)$ for all $1 \le i \le m$, $1 \le j \le l$. For any particular τ_i, σ_j, $\mathfrak{S} \vDash \neg\,\mathrm{Prf}(\tau_i, \sigma_j, \ulcorner\delta\urcorner)$ means that σ_j fails to be a sequence of length τ_i coding a proof of δ, that is, σ_j is not a sequence of length τ_i of codes of formulas such that the corresponding sequence of formulas is a proof of δ. Now, there is an effective procedure that, when applied to the term σ_j, leads to verifying that either σ_j does not name a sequence of length (named by) τ_i or, while naming such a sequence, that sequence fails to satisfy at least one of the conditions spelled out in the description of the formula $\mathrm{Prf}(\tau_i, \sigma_j, \ulcorner\delta\urcorner)$ by **(8)** in Sect. 4. Thus the failure of σ_j to code a proof of δ of length τ_i is *provable* in HF. We conclude that

$$\mathfrak{S} \vDash \neg\,\mathrm{Prf}(\tau_i, \sigma_j, \ulcorner\delta\urcorner) \;\Rightarrow\; \vdash \neg\,\mathrm{Prf}(\tau_i, \sigma_j, \ulcorner\delta\urcorner) \tag{7}$$

for all $1 \le i \le m$, $1 \le j \le l$. So $\mathfrak{S} \vDash \neg\alpha(\ulcorner\delta\urcorner)$ implies $\vdash \neg\alpha(\ulcorner\delta\urcorner)$, by (6) and (7). This finishes the proof of (c), and hence of Theorem 11.1. ■

Appendix: The axiomatic theory of finite sets

In this Appendix, totally self-contained, we present a systematic development of the theory of finite sets, also called the theory of *hereditarily finite sets* or *sets of finite rank*. We call this theory HF. A systematic presentation of HF from its axioms has not been available hitherto. Tarski and Givant proposed in [TG] several equivalent axiomatizations, and ours is a modification of one of these. In the exposition of HF given below, we present only as much of the theory as is needed for the proofs of the incompleteness results on the previous pages.

1. Axioms and basic results

The non-logical symbols of HF are 0, $\in$ and $\triangleleft$, where 0 is a constant, $\in$ a binary relation symbol and $\triangleleft$ a binary operation symbol. One can axiomatize HF using $\in$ as the only non-logical symbol (see [TG]). However in that case the theory has no terms, whilst it is precisely the constant terms that we have used for the coding of the meta-theory.

Informally, 0 denotes the empty set and $x \triangleleft y$ is the result of enlarging the set x by a new element, namely y. (Thus $y \in x$ is equivalent to $x \triangleleft y = x$, so that $\triangleleft$ could serve as the only non-logical symbol, while 0 would be defined by HF1.) For the axioms we take the universal closures of the following formulas:

HF1. $z = 0 \leftrightarrow \forall x(x \notin z)$.
HF2. $z = x \triangleleft y \leftrightarrow \forall u(u \in z \leftrightarrow u \in x \vee u = y)$.
HF3. $\alpha(0) \wedge \forall x \forall y[\alpha(x) \wedge \alpha(y) \rightarrow \alpha(x \triangleleft y)] \rightarrow \forall x \alpha(x)$.

The assumption in HF3 is that α is any formula in the first-order language based on 0, $\in$ and $\triangleleft$, moreover α contains a freely occurring variable z such that x and y are substitutable for z, whilst $\alpha(x)$, $\alpha(y)$ and $\alpha(x \triangleleft y)$ denote the effects of substitutions.

Theorem 1.1.

(a) $x \notin 0$.
(b) $u \in x \triangleleft y \leftrightarrow u \in x \vee u = y$.
(c) $z \in 0 \triangleleft y \leftrightarrow z = y$.

Proof.

(a) Substitute 0 for z in HF1.
(b) Substitute $x \triangleleft y$ for z in HF2.
(c) Substitute 0 for x in (b) and use (a). ∎

Theorem 1.2 (Extensionality Property). $x = z \leftrightarrow \forall u(u \in x \leftrightarrow u \in z)$.

Proof. Let $\alpha(x)$ be the asserted formula. To apply HF3, we check that

1° $\vdash \alpha(0)$ [by 1.1(a) and HF1].
2° $\vdash \alpha(x \triangleleft y)$ [by 1.1(b) and HF2]. ∎

Definition 1.3. We denote by $\{x\}$, $\{x, y\}$, $\{x, y, z\}$ and $\langle x, y\rangle$ the following *terms*:

(a) $\{x\} = 0 \triangleleft x$.
(b) $\{x, y\} = \{x\} \triangleleft y$, $\{x, y, z\} = \{x, y\} \triangleleft z$.
(c) $\langle x, y\rangle = \{\{x\}, \{x, y\}\}$.

Theorem 1.4.

(a) $u \in \{x\} \leftrightarrow u = x$.
(b) $u \in \{x, y\} \leftrightarrow u = x \vee u = y$.
(c) $\{x, x\} = \{x\}$.
(d) $\langle x, y\rangle = \langle u, v\rangle \leftrightarrow x = u \wedge y = v$.

Proof. (a), (b) follow directly from the definition. For (c), (d) use Theorem 1.2 and (a), (b). ∎

Theorem 1.5 (Existence of the union of two sets).

$$\exists z \forall u (u \in z \leftrightarrow u \in x \vee u \in y).$$

Proof. Let $\alpha(x)$ be the above formula. If z exists, z is unique by 1.2, so denote z by $x \cup y$. To prove $\forall x \alpha(x)$, we apply HF3:

1° $\vdash \alpha(0)$ [take $z = y$ in $\alpha(0)$].
2° $\vdash [\alpha(x) \rightarrow \alpha(x \lhd w)]$.

To show 2°, assume $\alpha(x)$, i.e., the existence of $x \cup y$. By 1.1(b), $\alpha(x \lhd w)$ is equivalent to $\exists z \forall u (u \in z \leftrightarrow u \in x \vee u = w \vee u \in y)$. We can take $z = (x \cup y) \lhd w$. ■

Theorem 1.6 (Existence of the union of a set of sets).

$$\exists z \forall u (u \in z \leftrightarrow \exists (y \in x)[u \in y]).$$

Proof. Let $\alpha(x)$ be the above formula and let $\bigcup x$ be the above z, if it exists. To prove $\forall x \alpha(x)$, we apply HF3:

1° $\vdash \alpha(0)$ [take $z = 0$ in $\alpha(0)$, i.e., $\bigcup 0 = 0$].
2° $\vdash [\alpha(x) \rightarrow \alpha(x \lhd w)]$.

To show 2°, assume $\alpha(x)$, i.e., that $\bigcup x$ exists. We need to find z such that

$$\begin{aligned} u \in z &\leftrightarrow \exists (y \in x \lhd w)[u \in y] \\ &\leftrightarrow \exists y[(y \in x \vee y = w) \wedge (u \in y)] \\ &\leftrightarrow \exists (y \in x)[u \in y] \vee u \in w \\ &\leftrightarrow u \in \bigcup x \vee u \in w. \end{aligned}$$

So we can take $z = \bigcup x \cup w$ (see previous theorem). ■

Theorem 1.7 (Comprehension Scheme). $\exists z \forall u [u \in z \leftrightarrow (u \in x) \wedge \varphi(u)]$ *for any formula* $\varphi(u)$ *in which* z *is not free.*

Proof. Let $\alpha(x)$ be the above formula. If z exists, then it is unique and we shall denote it by $\{u \in x : \varphi(u)\}$. To prove $\forall x \alpha(x)$, we use HF3:

1° $\vdash \alpha(0)$ [take $z = 0$ in $\alpha(0)$].
2° $\vdash [\alpha(x) \rightarrow \alpha(x \lhd y)]$.

To show 2°, assume $\alpha(x)$, i.e., that $\{u \in x : \varphi(u)\}$ exists. We have to find a $\bar{z}$ such that

$$u \in \bar{z} \leftrightarrow (u \in x \lhd y) \wedge \varphi(u) \leftrightarrow (u \in x \vee u = y) \wedge \varphi(u).$$

Since z is not free in $\varphi(u)$, the required $\bar{z}$ is $\{u \in x : \varphi(u)\} \lhd y$ if $\varphi(y)$ and $\{u \in x : \varphi(u)\}$ if $\neg\varphi(y)$. ■

Definition 1.8 (Intersection).

(a) $x \cap y = \{u \in x : u \in y\}$.
(b) $\bigcap x = \{v \in u_0 : \forall (u \in x)[v \in u]\}$ for some $u_0 \in x$ if $x \neq 0$ and $\bigcap x = 0$ if $x = 0$.

Theorem 1.9 (Replacement Scheme).

$$[(\forall (u \in x) \exists ! v \psi(u, v)] \rightarrow \exists z \forall v [v \in z \leftrightarrow \exists (u \in x) \psi(u, v)]$$

for any formula ψ *in which* z *is not free.* ("∃!" means "there exists a unique".)

Proof. Let $\alpha(x)$ be the above formula. If z exists, then it is unique and we shall denote it by $\{v : \exists(u \in x)\psi\}$. To prove $\vdash \forall x\alpha(x)$, we use HF3:

1° $\vdash \alpha(0)$ [take $z = 0$ in $\alpha(0)$].
2° $\vdash [\alpha(x) \to \alpha(x \lhd y)]$.

To show 2°, assume $\alpha(x)$, that is, assume that $\{v : \exists(u \in x)\psi\}$ exists provided that $\forall(u \in x)\exists!v\psi(u,v)$. To show $\alpha(x \lhd y)$, assume that $\forall(u \in x \lhd y)\exists!v\psi(u,v)$, i.e., that $\forall u[(u \in x \vee u = y) \to \exists!v\psi(u,v)]$. Hence $\exists!v\psi(y,v)$, so let v_y be this unique v. Thus the required z on the right hand side of the implication $\alpha(x \lhd y)$, i.e., $\{v : \exists(u \in x \lhd y)\psi\}$, is $\{v : \exists(u \in x)\psi\} \lhd v_y$. ■

Definition 1.10 (Subset relation).

(a) $y \subseteq x \leftrightarrow \forall v(v \in y \to v \in x)$.
(b) $y \subset x \leftrightarrow y \subseteq x \wedge y \neq x$.

Theorem 1.11 (Existence of the power set). *$\exists z\forall u(u \in z \leftrightarrow u \subseteq x)$.*

Proof. Let $\alpha(x)$ be the above formula. If z exists, it is unique and we shall denote it by $P(x)$. To prove $\forall x\alpha(x)$ we apply HF3. We have

1° $\vdash \alpha(0)$ [take $z = \{0\}$ in $\alpha(0)$].
2° $\vdash [\alpha(x) \to \alpha(x \lhd y)]$.

To show 2°, assume $\alpha(x)$, so that $P(x)$ exists. One checks directly from the definition that

$$u \subseteq x \lhd y \leftrightarrow u \in P(x) \vee \exists(v \in P(x))[u = v \lhd y].$$

Hence, using the existence of union and the Replacement Scheme (Theorems 1.5 and 1.9), we find that $P(x \lhd y) = P(x) \cup \{u : \exists(v \in P(x))[u = v \lhd y]\}$. ■

Definition 1.12. We shall call w an $\in$-*minimal* element of z if $w \in z$ and no element of w belongs to z (i.e., $w \cap z = 0$).

Theorem 1.13 (Foundation (Regularity) Property).

$$z \neq 0 \to \exists(w \in z)[w \cap z = 0].$$

Proof (*by Donald Monk*). We need to show that $\forall(w \in z)[w \cap z \neq 0] \to z = 0$ and this will follow if we prove the implication $\forall(w \in z)[w \cap z \neq 0] \to \forall x\alpha(x)$, where $\alpha(x)$ is the formula $(x \notin z) \wedge (x \cap z = 0)$. By HF3, it will be sufficient to establish:

1° $\vdash \forall(w \in z)[w \cap z \neq 0] \to \alpha(0)$.
2° $\vdash \forall(w \in z)[w \cap z \neq 0] \to [\alpha(x) \wedge \alpha(y) \to \alpha(x \lhd y)]$.

Here 1° is obvious, and to prove 2°, we verify the mutual incompatibility of the following four conditions:

(a) $\forall(w \in z)[w \cap z \neq 0]$.
(b) $\alpha(x)$, that is, $(x \notin z) \wedge (x \cap z = 0)$.
(c) $\alpha(y)$, that is, $(y \notin z) \wedge (y \cap z = 0)$.
(d) $\neg\alpha(x \lhd y)$, that is, $[(x \lhd y) \in z)] \vee [(x \lhd y) \cap z \neq 0]$.

The two cases in (d) are:

Case 1: $(x \lhd y) \cap z \neq 0$. Since $y \in z$ is excluded by (c), we get $x \cap z \neq 0$, and that is excluded by (b).

Case 2: $(x \lhd y) \in z$. Then $(x \lhd y) \cap z \neq 0$, by (a), and we are back to Case 1. ∎

Corollary 1.14. $x \cap \{x\} = 0$ (*i.e.*, $x \notin x$).

Proof. Apply Foundation Property to $z = \{x\} \neq 0$. ∎

2. Ordinals

Definition 2.1. We say that x is *transitive* if every element of x is a subset of x. We call x an *ordinal* if x is transitive and every element of x is transitive. This will be expressed by the formula

$$\forall(y \in x)\{[y \subseteq x] \wedge \forall(z \in y)[z \subseteq y]\}.$$

The above formula will be abbreviated as $\mathrm{Ord}(x)$.

Definition 2.2 (Successor). $x^+ = x \lhd x = x \cup \{x\}$.

Theorem 2.3.

(a) $\mathrm{Ord}(x) \to \mathrm{Ord}(x^+)$.
(b) $\mathrm{Ord}(x) \wedge y \in x \to \mathrm{Ord}(y)$.
(c) $\mathrm{Ord}(0)$.
(d) *If x is a non-zero ordinal*, *then the $\in$-minimal element of x is* 0.

Proofs. Immediate from Definitions 1.12, 2.1 and 2.2. ∎

Ordinals will be mostly denoted by k, l, m, n, p, q, r.

Theorem 2.4 (Comparability of ordinals).

$$\mathrm{Ord}(k) \wedge \mathrm{Ord}(l) \to k \in l \vee k = l \vee l \in k.$$

Proof. Let $\beta(k, l)$ be the formula $k \in l \vee k = l \vee l \in k$ and suppose that

$$\neg\beta(k_0, l_0) \tag{1}$$

for some ordinals k_0, l_0. We claim that then we can select k_0 that satisfies

$$\forall(m \in k_0)\forall l\beta(m, l) \tag{2}$$

($\forall l$ ranging over ordinals). Indeed, if (2) fails then $\{m \in k_0 : \exists l\neg\beta(m, l)\} \neq 0$ and if r_0 is an $\in$-minimal element of this set (see 1.13), then $r_0 \in k_0$ and for each $n \in r_0$, we have $n \in k_0$ (because k_0 is transitive), and $\forall l\beta(n, l)$. In other words, $\forall(n \in r_0)\forall l\beta(n, l)$. So (2) holds if we replace k_0 by r_0—or simply start our considerations with r_0 in place of k_0 (noting that $\exists l\neg\beta(r_0, l)$, by the definition of r_0).

We claim further that, having selected k_0 such that (2) holds, we can select l_0 satisfying

$$\forall(p \in l_0)\beta(k_0, p). \tag{3}$$

Indeed, suppose (3) fails for some p, so that $\{p \in l_0 : \neg\beta(k_0, p)\} \neq 0$. Let q_0 be an $\in$-minimal element of this set. Then $q_0 \in l_0$ and for each $m \in q_0$ we have $m \in l_0$ (because l_0 is transitive), and also $\beta(k_0, m)$. In other words, $\forall(m \in q_0)\beta(k_0, m)$—so we get (3) with q_0 in place of l_0, or else we get (3) by just replacing l_0 at the beginning of our considerations by q_0 (noting that $\neg\beta(k_0, q_0)$, by the definition of q_0).

We now claim that $l_0 \subset k_0$. Indeed, suppose $p \in l_0$. Then $\beta(k_0, p)$, by (3), i.e. $k_0 \in p \vee k_0 = p \vee p \in k_0$. But $k_0 \in p$ contradicts $p \in l_0$ and (1) (due to $p \subseteq l_0$). Also $k_0 = p$ contradicts (1) (since $p \in l_0$ is assumed). Thus $p \in k_0$, and $l_0 \subseteq k_0$ follows. Since $l_0 = k_0$ contradicts (1), we get $l_0 \subset k_0$.

Let $m \in k_0 \setminus l_0$. Then $\beta(m, l_0)$, by (2), i.e.,

$$m \in l_0 \vee m = l_0 \vee l_0 \in m.$$

Here both $m = l_0$ and $l_0 \in m$ contradict (1), because $m \in k_0$. But also $m \in l_0$ is impossible, by the choice of m. ■

Theorem 2.5. *Let k, l be ordinals. Then*

(a) *Exactly one of $k \in l$, $k = l$, $l \in k$ occurs.*
(b) $k \in l \leftrightarrow k \subset l$.
(c) $l \in k \to (l^+ = k \vee l^+ \in k)$.
(d) $k^+ = l^+ \to k = l$.

Proof. (a) Any two of the three mentioned possibilities contradict 1.14. At least one possibility occurs, by 2.4.

(b) Implication $\to$ follows by the transitivity of l and 1.14. Suppose $k \subset l$. Then $k \in l \vee l \in k$, by 2.4. But $l \in k$ yields $l \in l$, contradicting 1.14.

(c) By 2.4, we have to exclude $k \in l^+$, i.e., we have to show that neither $k \in l$ nor $k = l$ can occur. This is indeed so because each of these possibilities, together with $l \in k$, leads to $l \in l$.

(d) Suppose $k \neq l$ and $k^+ = k \cup \{k\} = l \cup \{l\} = l^+$. Then $k \in l$ and $l \in k$, yielding again the contradiction $l \in l$. ■

Definition 2.6. For ordinals k, l, we shall use:

(i) $k < l$ to abbreviate $k \in l$,
(ii) $k \leq l$ to abbreviate $x < y \vee x = y$.

In Section 5 we shall define a binary relation of inequality $<$ on the universe of all sets. On the sub-universe of ordinals, both relations will be seen to coincide.

Theorem 2.7. *Every set x of ordinals is ordered by the binary relation $<$ (or equivalently, by $\subset$). Moreover if $x \neq 0$ then x has a smallest and a largest element.*

Proof. Let x be a set of ordinals. Then x is ordered by $\in$ because of (a), (b) in 2.5. Denote the largest element of x, if existing, by $\max(x)$. Let $\alpha(x)$ be the formula stating the existence of $\max(x)$ for non-empty x:

$$[x \neq 0 \wedge \forall(k \in x)\,\mathrm{Ord}(k)] \to \exists(l \in x)\forall(k \in x)[k \leq l].$$

To prove $\vdash \forall x \alpha(x)$, we use HF3, i.e., we verify:

1° $\vdash \alpha(0)$ [obvious].
2° $\vdash [\alpha(x) \to \alpha(x \triangleleft y)]$.

To show the latter, assume $\alpha(x)$, i.e., the existence of $\max(x)$, provided x is a non-empty set of ordinals. We should deduce $\alpha(x \lhd y)$. Thus assume that $x \lhd y$ is a set of ordinals (evidently non-empty). Then x is a set of ordinals, so if $x \neq 0$, then $\max(x)$ exists by the assumption $\alpha(x)$. Since $x \lhd y$ arises from x by adjoining *one* element, it is clear that $\max(x \lhd y)$ also exists. If, on the other hand, $x = 0$, then $x \lhd y = \{y\}$ and $y = \max(x \lhd y)$.

To deduce the existence of $\min(x)$, replace in the above proof $\leq$ by $\geq$ and "max" by "min" throughout. ■

Theorem 2.8. $\mathrm{Ord}(k) \wedge k \neq 0 \to \exists! l(l^+ = k)$.

Proof. By 2.1, k is a non-empty set of ordinals. Let $l = \max(k)$ (see 2.7). Then $l \in k$, whence $l^+ = k \vee l^+ \in k$, by 2.5. But $l^+ \in k$ contradicts $l = \max(k)$. Thus $l^+ = k$. The uniqueness of l follows from 2.5(d). ■

Definition 2.9. The unique (if existing) l for which $l^+ = k$ will be denoted by $\overline{k}$ and called the *predecessor* of k.

3. p-functions

As commonly defined, x is a *function* if each $y \in x$ is an ordered pair and

$$[\langle u, v_1\rangle \in x \wedge \langle u, v_2\rangle \in x] \to v_1 = v_2.$$

Since $\langle u, v\rangle = \{\{u\}, \{u, v\}\}$, it follows from $\langle u, v\rangle \in x$ that $\{u\} \in \bigcup x$, and hence $u \in \bigcup(\bigcup x)$. Thus the *domain* of a function x is

$$\mathrm{dom}(x) = \Big\{u \in \bigcup\Big(\bigcup x\Big) : \exists v(\langle u, v\rangle \in x)\Big\}$$

(see 1.6 and 1.7).

It will be convenient for us to use also functions in a broader sense; their domains will not be sets and we shall call them p-functions. A symbol denoting a p-function will not belong to the formal language describing HF (if it were added to the language of HF then in discourses like in Part 1 it would have to be coded). Nevertheless, formulas containing (names of) p-functions will be written and understood as representing certain formulas of HF. p-functions will always be introduced by formulas of HF, by a process we shall presently describe.

Definition 3.1. A formula φ of HF will be called *functional* with respect to a free occurring variable y iff $(\exists! y)\varphi$ is a theorem.

If φ has $n+1$ free variables and φ is functional with respect to y, we associate with φ and y a new n-ary function symbol F^y_φ. We shall write formulas of the first order language obtained by adjoining F^y_φ to the language of HF. However this expanded language will NOT be our formal means of describing HF, and formulas containing F^y_φ will be viewed merely as *representing* (mostly abbreviating) certain formulas of HF. F^y_φ will be called a *p-symbol* and formulas containing F^y_φ will be called *p-formulas.* Here we may read "p" as standing for "phantom" (in [Bo], "p" stands for "pseudo")—to symbolize appearance without formal acceptance.

A term $F^y_\varphi(\tau_1, \ldots, \tau_n)$, where $\tau_1, \ldots, \tau_n$ are any terms of HF, will be called a *simple p-term.* In general, terms belonging to the language of HF expanded by F^y_φ will be called *p-terms.* Such p-terms do not represent anything in HF, however formulas containing p-terms (i.e., p-formulas) will, as mentioned above, represent formulas of HF. This representation is not unique: a p-formula represents many, logically equivalent, formulas of HF. To describe this representation, we introduce a process of *reduction* by which a p-formula is reduced to one containing one less simple p-term. In detail, suppose the free variables in φ, other than y, are $x_{i_1}, \ldots, x_{i_n}$ (say, in the increasing order of indices). Then we shall write φ as $\varphi(x_{i_1}, \ldots, x_{i_n}, y)$. Let α be a p-formula. Clearly, α contains a simple p-term and the latter occurs in an atomic sub-formula of α. In other words, there exists an atomic formula $\beta(z)$ and there are HF-terms $\tau_1, \ldots, \tau_n$ such that $\beta(F^y_\varphi(\tau_1, \ldots, \tau_n))$ is a sub-formula of α. The reduction which we now apply to α consists in replacing, inside α, the atomic sub-formula $\beta(F^y_\varphi(\tau_1, \ldots, \tau_n))$ by

$$\exists w[\beta(w) \land \widetilde{\varphi}(\tau_1, \ldots, \tau_n, w)],$$

where w is a new variable, not occurring in α (thus neither in any of the terms $\tau_1, \ldots, \tau_n$) and $\widetilde{\varphi}$ is any variant of φ such that $\tau_1, \ldots, \tau_n, w$ are substitutable for $x_{i_1}, \ldots, x_{i_n}, y$ in $\widetilde{\varphi}$. [A *variant* $\widetilde{\varphi}$ of φ is obtained by renaming the bound variables of φ so that $\varphi \leftrightarrow \widetilde{\varphi}$ is a theorem of logic.]

As a result of such a reduction the number of occurrences of F^y_φ in a formula is reduced by one. Thus, when no further reduction is possible, we end up with a formula of HF and we shall say that the latter is *represented* by α (or, that α *represents* the latter). Let us agree that if α represents γ then α also represents every formula equivalent to γ.

For example, the p-formula $F^y_\varphi(x_{i_1}, \ldots, x_{i_n}) = u$ represents

$$\exists w[w = u \land \varphi(x_{i_1}, \ldots, x_{i_n}, w)],$$

and the latter, in view of $\vdash \exists! y \varphi(x_{i_1}, \ldots, x_{i_n}, y)$, is equivalent to $\varphi(x_{i_1}, \ldots, x_{i_n}, u)$.

Often instead of F^y_φ we shall use only one letter (usually capital); then φ and y will be specified beforehand. By abuse of language, we shall call a p-function symbol a *p-function.*

We shall use several p-functions, each of these introduced by a formula functional with respect to a variable (3.1). A formula made up of symbols of HF and several distinct p-function symbols represents a formula of HF. We obtain the latter by consecutive reductions eliminating simple p-terms as described above.

Occasionally it may be convenient to consider a p-formula $\psi(x, y)$ functional with respect to y. However the p-function F^y_ψ defined by such a formula need not be regarded as a new sort of entity, as we can identify it with the p-function F^y_φ, where φ is an HF-formula represented by ψ.

Examples of p-functions abound; e.g., it is clear that the following formulas are functional with respect to y:

(a) $y = \bigcup x$.

(b) $y = \bigcap x$.

(c) $y = P(x)$ (i.e., $\forall u(u \in y \leftrightarrow u \subseteq x)$).

Other examples are p-functions defined by recursion. We begin with infinite sequences, defined by recursion on ordinals. To define these, we use finite sequences, called here simply sequences.

Definition 3.2 (Sequence). A function whose domain is an ordinal $k \neq 0$ will be called a *sequence of length* k. We denote by $\mathrm{Seq}(s, k)$ any formula stating that s is a sequence of length k (e.g., the formula introduced in Part 1, 2.3(b)). If $\mathrm{Seq}(s, k)$ then, for any formula $\alpha(z)$ and $n < k$, we shall write $\alpha(s_n)$ to abbreviate $\exists z[\langle n, z\rangle \in s \wedge \alpha(z)]$.

Theorem 3.3. *For every constant term τ and p-functions G, H there exists a p-function F such that*

(a) $F(0) = \tau$.
(b) $x \neq 0 \wedge \mathrm{Ord}(x) \rightarrow F(x) = G(F(\overline{x}))$.
(c) $\neg\,\mathrm{Ord}(x) \rightarrow F(x) = H(x)$.

Proof. Let $\psi(x, y)$ be the p-formula

$$\{x \neq 0 \wedge \mathrm{Ord}(x) \wedge \exists s[\mathrm{Seq}(s, x) \wedge y = G(s_{\overline{x}}) \wedge \forall(n < x)[(n = 0 \wedge s_n = \tau) \vee$$
$$(n \neq 0 \wedge s_n = G(s_{\overline{n}}))]]\} \vee [\neg\,\mathrm{Ord}(x) \wedge H(x) = y] \vee [x = 0 \wedge y = \tau].$$

Then it is easy to prove that ψ is functional with respect to y. (The existence of a least ordinal x such that there is no unique y for which $\psi(x, y)$ leads easily to a contradiction.) Next, we check that (a), (b), (c) hold when F is F_ψ^y. ∎

Definition 3.4. We shall say that the p-function F in 3.3 is *defined recursively on ordinals.*

4. The rank function

Definition 4.1. Let $R(x)$ denote the p-function defined recursively on ordinals by

(a) $R(0) = 0$.
(b) $R(x) = P(R(\overline{x}))$ for very non-zero ordinal x.
(c) $R(x) = 0$ if x is not an ordinal.

Theorem 4.2. *For all ordinals m, n:*

(a) $n < m \rightarrow R(n) \subset R(m)$.
(b) $R(m)$ *is transitive.*

Proof. (a) Obviously $R(0) = 0 \subset R(m)$ for $0 < m$. Now, if there is some n such that $n < m$, for some m, but $R(n) \subset R(m)$ fails, then by 2.7 there is the least n with that property. By the above, $n > 0$, whence $\overline{n} < \overline{m}$. Moreover, $R(\overline{n}) \subset R(\overline{m})$. Thus $R(n) \subset R(m)$, by the definition of R, and we get a contradiction.

(b) Clearly $R(0) = 0$ is transitive. If $m \neq 0$, then $\overline{m} < m$, whence

$$x \in R(m) \rightarrow x \subseteq R(\overline{m}) \subset R(m),$$

by (b) in 4.1 and (a) above. ∎

Theorem 4.3. $\exists n[\mathrm{Ord}(n) \wedge x \in R(n)]$ (*i.e., the sets $R(n)$ cover the universe*).

Proof. Let $\alpha(x)$ be the above formula. We apply HF3 to prove $\forall x\alpha(x)$. We have

1° $\alpha(0)$ [because $0 \in R(0^+) = P(R(0)) = P(0) = \{0\}$].
2° $\vdash [\alpha(x) \wedge \alpha(y) \to \alpha(x \triangleleft y)]$.

To prove 2°, assume $\alpha(x) \wedge \alpha(y)$, i.e., $x \in R(n)$, $y \in R(m)$ for some ordinals m, n. Then

$$x \triangleleft y = x \cup \{y\} \subseteq R(n) \cup R(m^+)$$

by 4.1(b) and 4.2(b). So, for $k = \max\{m^+, n\}$, $x \triangleleft y \subseteq R(k)$ (see 4.2(a)), whence $x \triangleleft y \in R(k^+)$. ■

Definition 4.4. For every x, let $\operatorname{rank}(x)$ be the least ordinal n such that $x \in R(n^+)$ (i.e., $x \subseteq R(n)$).

Theorem 4.5. $x \in y \to \operatorname{rank}(x) < \operatorname{rank}(y)$.

Proof. Let $x \in y$ and suppose $\operatorname{rank}(y) = n$. Then $y \subseteq R(n)$ and hence $x \in R(n)$. Thus $n \neq 0$ and $x \subseteq R(\overline{n})$, whence $\operatorname{rank}(x) \leq \overline{n} < n$. ■

Definition 4.6. The *transitive closure* $\operatorname{cl}(x)$ of x is the minimal transitive set y such that $x \subseteq y$.

Concerning the existence of $\operatorname{cl}(x)$, note that $x \subseteq R(\overline{n})$ for some ordinal n (see 4.3), and $R(\overline{n})$ is transitive (see 2.2). Thus $\operatorname{cl}(x)$ is the intersection of all transitive subsets of $R(\overline{n})$ which contain x as a subset.

Theorem 4.7. $\operatorname{rank}(x) = \operatorname{rank}(\operatorname{cl}(x))$.

Proof. One checks without problems that:

(a) $u \subseteq v \to \operatorname{rank}(u) \leq \operatorname{rank}(v)$.
(b) $\operatorname{rank}(u \cup v) = \max\{\operatorname{rank}(u), \operatorname{rank}(v)\}$ (see 4.2(a)).
(c) $\operatorname{rank}(\{w\}) = (\operatorname{rank}(w))^+$.

Now let $\alpha(x)$ be the formula $\operatorname{rank}(x) = \operatorname{rank}(\operatorname{cl}(x))$. We apply HF3 to prove $\forall x\alpha(x)$. We have

1° $\vdash \alpha(0)$ [because $\operatorname{cl}(0) = 0$].
2° $\vdash [\alpha(x) \wedge \alpha(y) \to \alpha(x \triangleleft y)]$.

To prove 2°, assume $\alpha(x) \wedge \alpha(y)$, i.e., that for some m, n, $\operatorname{rank}(x) = \operatorname{rank}(\operatorname{cl}(x)) = n$ and $\operatorname{rank}(y) = \operatorname{rank}(\operatorname{cl}(y)) = m$. Then both sets on the two sides of the inclusion

$$x \triangleleft y = x \cup \{y\} \subseteq \operatorname{cl}(x) \cup \operatorname{cl}(y) \cup \{y\}$$

have the same rank, namely $\max\{n, m^+\}$ (see (b), (c) above). Since the set on the right is transitive,

$$x \triangleleft y \subseteq \operatorname{cl}(x \triangleleft y) \subseteq \operatorname{cl}(x) \cup \operatorname{cl}(y) \cup \{y\}.$$

Thus $x \triangleleft y$ and $\operatorname{cl}(x \triangleleft y)$ have the same rank (see (a)). ■

The name of the following theorem is justified by the fact that $\operatorname{rank}(u) < \operatorname{rank}(x)$ for all $u \in \operatorname{cl}(x)$, by 4.5 and 4.7. To state it, we need the following definition.

Definition 4.8 (Restriction). If f is a function and $z \subseteq \operatorname{dom}(f)$, then the function $\{y \in f : \exists(u \in z)\exists v(y = \langle u, v\rangle)\}$ will be denoted by $f{\restriction}z$ and called the *restriction* of f

to z. A suggestive notation is $f\restriction z = \{\langle u, f(u)\rangle : u \in z\}$. Similarly, if F is a p-function, we denote by $F\restriction z$ the function $\{\langle u, F(u)\rangle : u \in z\}$. (The existence of $F\restriction z$ follows from 1.9.)

Theorem 4.9 (Definition of a p-function by recursion on rank). *For every binary p-function $G(x, y)$ there exists a p-function F such that*

$$\vdash F(x) = G(x, F\restriction \mathrm{cl}(x)).$$

Proof. Let us call a function f *good* if its domain d is transitive and

$$(*) \qquad f(u) = G(u, f\restriction \mathrm{cl}(u))$$

for all $u \in d$. (The right side makes sense because $u \in d \to \mathrm{cl}(u) \subseteq \mathrm{cl}(d) = d$ due to transitivity.) We shall show:

(a) If f is good and d' is a transitive subset of the domain d of f, then $f\restriction d'$ is good.
(b) If f, f' are good functions with domains d, d', then $f\restriction(d \cap d') = f'\restriction(d \cap d')$.
(c) For every x, there exists a good function with domain $\mathrm{cl}(x) \lhd x$.

Assuming for the moment that (a), (b) and (c) have been shown, let us show how to find F. For every x, let f_x be the good function with domain $\mathrm{cl}(x) \lhd x$ (which exists and is unique, by (b) and (c)). Obviously, there is a formula $\varphi(x, y)$ of HF that can be abbreviated as $y = f_x(x)$. And clearly, such φ is functional with respect to y. Let F be F^y_φ. Then $F(x) = f_x(x)$ is a theorem, because according to the reduction procedure in Section 3, this p-formula reduces to $\exists w(w = f_x(x))$. So it remains to prove $f_x(x) = G(x, F\restriction \mathrm{cl}(x))$, and this will follow from $(*)$ (taken with $f = f_x$, $u = x$) provided $F\restriction \mathrm{cl}(x) = f_x\restriction \mathrm{cl}(x)$, i.e. if for all $u \in \mathrm{cl}(x)$ one has $f_u(u) = f_x(u)$. Now, this is indeed so, for if $u \in \mathrm{cl}(x)$ then

$$\mathrm{dom}(f_u) = \mathrm{cl}(u) \cup \{u\} \subseteq \mathrm{cl}(x) \subseteq \mathrm{dom}(f_x),$$

and since both domains contain u, $f_u(u) = f_x(u)$ follows from (b).

Proof of (a). If $(*)$ holds for all $u \in d$, then $(*)$ holds for all $u \in d' \subseteq d$.

Proof of (b). By the Foundation Theorem 1.13, 0 belongs to every transitive set, hence $0 \in d \cap d'$. Also

$$f(0) = G(0, 0) = f'(0).$$

Let $u_0 \in d \cap d'$ be of smallest rank such that $f(u_0) \neq f'(u_0)$, if such u_0 exists. By the above, $u_0 \neq 0$ and we have

$$u \in \mathrm{cl}(u_0) \to \mathrm{rank}(u) < \mathrm{rank}(u_0)$$

(see 4.5 and 4.7), whence $f(u) = f'(u)$ for all such u and thus $f\restriction \mathrm{cl}(u_0) = f'\restriction \mathrm{cl}(u_0)$. Hence

$$f(u_0) = G(u_0, f\restriction \mathrm{cl}(u_0)) = G(u_0, f'\restriction \mathrm{cl}(u_0)) = f'(u_0).$$

This contradicts the choice of u_0. We conclude that $f(u) = f'(u)$ for all $u \in d \cap d'$.

Proof of (c). We have $\mathrm{cl}(0) \lhd 0 = \{0\}$ and a good function f with this domain is defined by $f(0) = G(0, 0)$. Suppose there is an x_0 such that a good function with domain $\mathrm{cl}(x_0) \lhd x_0$ does not exist, and let x_0 be of the least rank possible. Then $x_0 \neq 0$. For each $u \in \mathrm{cl}(x_0)$, let f_u denote the unique (see (b)) good function with domain $\mathrm{cl}(u) \lhd u$. The existence of f_u follows from

$$\mathrm{rank}(u) < \mathrm{rank}(\mathrm{cl}(x_0)) = \mathrm{rank}(x_0)$$

(see 4.5, 4.7). Clearly $\mathrm{cl}(x_0)$ is covered by the domains $\mathrm{cl}(u) \lhd u$ of all f_u, $u \in \mathrm{cl}(x_0)$. So, by (b), the union

$$f = \bigcup \{f_u : u \in \mathrm{cl}(x_0)\}$$

is a good function with domain $\mathrm{cl}(x_0)$. But then the set

$$f \lhd \langle x_0, G(x_0, f)\rangle$$

is evidently a good function with domain $\mathrm{cl}(x_0) \lhd x_0$, contrary to the choice of x_0. The contradiction shows that a good function with domain $\mathrm{cl}(x) \lhd x$ always exists. This finishes the proof of (c), and hence of Theorem 4.9. ■

Certain p-functions of n variables can be defined in a similar way. Let us illustrate this for the case $n = 2$.

Theorem 4.10. *For every ternary p-function $G(x, y, z)$ there exists a binary p-function $F(x, y)$ such that*

$$\vdash F(x, y) = G(x, y, F{\restriction}(\mathrm{cl}(x) \times \mathrm{cl}(y))).$$

Proof. To show this, we proceed by analogy with the preceding proof. For transitive sets d_1, d_2, we say that a function f with domain $d_1 \times d_2$ is *good* if

$$f(u, v) = G(u, v, f{\restriction}(\mathrm{cl}(u) \times \mathrm{cl}(v)))$$

for all $\langle u, v\rangle \in d_1 \times d_2$. One proves that for every x, y there is exactly one good function with domain $(\mathrm{cl}(x) \lhd x) \times (\mathrm{cl}(y) \lhd y)$, and when this function is denoted by f_{xy}, one defines $F(x, y) = f_{xy}(x, y)$. In the part of the proof corresponding to (c) above, one can start again from negating the claim which leads to the existence of an x_0 of least rank for which there is a y_0 such that there does not exist a good function with domain $(\mathrm{cl}(x_0) \lhd x_0) \times (\mathrm{cl}(y_0) \lhd y_0)$. Since also y_0 can then be chosen of least rank, a contradiction results similarly to the proof of (c) above. ■

5. A definable order on the universe

In the ZF set theory there does not seem to exist a formula that defines a relation of global linear order. For HF, such definable order exists, moreover it is a refinement of the partial order by rank and also of the partial order by subsets. The existence of a definable order for HF can be deduced from the definitional equivalence of HF and PA, but we choose here a shorter path: The order will be defined recursively by a suitably chosen formula (the latter can be found in [My]).

In contrast to the ZF set theory, in HF every ordering of a set x is a well-ordering (in spite of the fact that in non-standard models of HF certain sets are infinite).

Theorem 5.1. *Every ordering relation $r \subseteq x \times x$ is a well-ordering.*

Proof. Let $\alpha(x)$ be the formula saying: "if $r \subseteq x \times x$ orders x, then r well-orders" x. To prove $\vdash \forall x \alpha(x)$, we use HF3. We have

1° $\vdash \alpha(0)$ [obvious].

2° $\vdash \alpha(x) \to \alpha(x \lhd y)$.

To prove 2°, assume $\alpha(x)$ and suppose $r \subseteq (x \triangleleft y) \times (x \triangleleft y)$ orders $x \triangleleft y$. Then, by $\alpha(x)$, the restriction of r to x well-orders x. By adding to x the single element y (if it is not in x already), one does not spoil the well-ordering. ■

Corollary 5.2. *For every ordering relation r on a set x, every non-empty $y \subseteq x$ has a minimal element $\min_r(y)$ and a maximal element $\max_r(y)$. [For $\max_r$ consider the inverse relation r^{-1}, which also well-orders x.]*

Definition 5.3. By an n-ary *p-relation* we shall mean a formula φ with n free variables. The characteristic p-function of such a p-relation is the $\{0, 1\}$-valued p-function F such that

$$\vdash (F(x_1, \dots, x_n) = 1) \leftrightarrow \varphi(x_1, \dots, x_n).$$

We say that the p-relation is defined by *recursion on rank* if its characteristic p-function is defined by recursion on rank (as in Theorem 4.9 or its generalization to more variables).

Theorem 5.4. *There exists a binary p-relation $<$ such that*

$$\vdash x < y \leftrightarrow \exists (v \in y \setminus x) \forall (u \in x \setminus y)(u < v).$$

Proof. This reduces to a verification that the above defines $<$ by recursion on rank, i.e., according to 4.10. Obviously $0 < y$ for all $y \neq 0$, and $x \not< 0$ for all x. Moreover, for $x, y \neq 0$ the above condition determines $x < y$ when $u < v$ is known for all $\langle u, v \rangle \in x \times y$, thus $x < y$ is determined by the restriction of $<$ to $\mathrm{cl}(x) \times \mathrm{cl}(y)$. Hence 4.10 can be applied to find the (binary) recursively defined characteristic p-function F of $<$. (We leave checking the details, i.e., finding (the ternary) G, to the reader). ■

Note. It is clear from this definition that $x \subset y \to x < y$. Hence our use hitherto of $<$ between ordinals does not clash with the definition of $<$ in 5.4.

To establish the order properties of $<$ we shall associate with every order r on a set z an order $\{r\}_z$ on the power set $P(z)$ of z.

Definition 5.5. If r is an order relation on z, we shall denote by $\{r\}_z$ the binary relation on the power set $P(z)$ given, for any $x, y \subseteq z$, by

$$x\{r\}_z y \leftrightarrow \exists (v \in y \setminus x) \forall (u \in x \setminus y)(\langle u, v \rangle \in r).$$

(Here $x\{r\}_z y$ abbreviates $\langle x, y \rangle \in \{r\}_z$.)

Note. Each $y \subseteq z$ is described by its characteristic function $\chi_y : z \to \{0, 1\}$, so that $\chi_y(u) = 1$ iff $u \in y$. The functions χ_y can be viewed as z-indexed sequences of 0's and 1's. The order r on z (which is a well-ordering; see 4.1) determines the so-called *anti-lexicographical* ordering of z-indexed sequences of integers. This (when restricted to 0–1 sequences) is an ordering of the set of characteristic functions χ_y, $y \subseteq z$. By the correspondence $\chi_y \Leftrightarrow y$, it induces the order $\{r\}_z$.

It is easily seen that $x\{r\}_z y \leftrightarrow x \neq y \wedge \max_r[(x \setminus y) \cup (y \setminus x)] \in y$. From this we immediately get the proof of:

Theorem 5.6. *$x \subset y \subseteq z \to x\{r\}_z y$.*

Theorem 5.7. *For every order relation r on a set z, $\{r\}_z$ is an order relation on $P(z)$.*

Proof. For any subset u of z, let $\{r\}_u$ denote the restriction of r to u, i.e., $r \cap (u \times u)$. If $z = 0$ then there is nothing to prove. So let us assume that $z \neq 0$ and let us denote $\max_r(z)$ by w (see 4.2). We first show that if $\{r\}_{z\setminus\{w\}}$ is an order on $P(z \setminus \{w\})$, then $\{r\}_z$ is an order on $P(z)$. For this purpose, we first note that for every $x, y \subseteq z$:

(1) $w \in x \cap y \to [x\{r\}_z y \leftrightarrow (x \setminus \{w\})\{r\}_{z\setminus\{w\}}(y \setminus \{w\})]$.
(2) $w \notin x \cup y \to [x\{r\}_z y \leftrightarrow x\{r\}_{z\setminus\{w\}} y]$.
(3) $w \in y \setminus x \to x\{r\}_z y$.

Let $P(z)$ be partitioned in two sets: those subsets of z which contain w and those which do not. Then, assuming that $\{r\}_{z\setminus\{w\}}$ is an order on $P(z \setminus \{w\})$, we see from (1) and (2) that each of the two sets of subsets of z is ordered by $\{r\}_z$, and it follows from (3) that each element of the second set (of those subsets of z which do not contain w) r-precedes each element of the first set. Thus $\{r\}_z$ is an order on $P(z)$.

To complete the proof, let $u_0 = \min_r(z)$ and for each $v \in z$, let $[u_0, v] \subseteq z$ be the closed interval (for the order r) with end-points u_0, v. One now shows by induction with respect to the well-ordering r (see 4.1) that for each $v \in z$, $\{r\}_{[u_0,v]}$ is an order on $P([u_0, v])$. This is evident for $v = u_0$ (because of $[u_0, u_0] = \{u_0\}$, $P(\{u_0\}) = \{0, \{u_0\}\}$ and 5.6). The inductive step is accomplished by applying to $[u_0, v]$ the same observations (1), (2), (3) as were made for z (but with $w = v$ in this case). ■

Theorem 5.8. *Let $<$ be the p-relation introduced by Theorem* 5.4, *let n be a non-zero ordinal and suppose that*:

(a) *$R(n)$ is ordered by $<$, and*
(b) *$R(\overline{n}) < (R(n) \setminus R(\overline{n}))$, i.e., $\forall u \forall v[u \in R(\overline{n}) \wedge v \in (R(n) \setminus R(\overline{n})) \to u < v]$.*

Then:

(a$^+$) *$R(n^+)$ is ordered by $<$, and*
(b$^+$) *$R(n) < (R(n^+) \setminus R(n))$.*

Proof. Assume (a), (b) and $n \neq 0$. By definition, $R(n^+) = P(R(n))$ and on $R(n^+)$ the relation $<$ coincides with $\{<\}_{R(n)}$ (see 5.5). Hence (a$^+$) holds, by 5.7. Now let $y \in R(n^+) \setminus R(n)$ and let $x \in R(n)$, i.e., $x \subseteq R(\overline{n})$. Then $y \subseteq R(n)$ and $y \nsubseteq R(\overline{n})$, so there is a $v \in y$ such that $v \in R(n) \setminus R(\overline{n})$. By (b), this v satisfies $u < v$ for all $u \in R(\overline{n})$, hence also for all $u \in x \subseteq R(\overline{n})$. Thus $x < y$, by 5.4, and (b$^+$) is shown. ■

Theorem 5.9. *$<$ is a relation of order and $<$ refines the partial order by subsets and the partial order by rank.*

Proof. It will suffice to prove that:

(1) $x \not< x \wedge (x < y < z \to x < z) \wedge (x < y \vee y < x \vee x = y)$.
(2) $\mathrm{rank}(x) < \mathrm{rank}(y) \to x < y$.

To begin with, note that (a) and (b) in 5.8 hold when $\overline{n} = 0$, i.e., when $R(\overline{n}) = 0$ and $R(n) = \{0\}$. So it follows from that theorem that (a) and (b) hold for all non-zero ordinals n. In particular, taking n large enough so that $x, y, z \in R(n)$ (see 4.3), we get (1).

Next, assume that $\mathrm{rank}(x) < \mathrm{rank}(y)$ and let $\mathrm{rank}(x) = m$, $\mathrm{rank}(y) = n$. Then

$$m < n \rightarrow m^+ \leq n \rightarrow R(m^+) \subseteq R(n).$$

Since $x \in R(m^+) \subseteq R(n)$ and $y \in R(n^+) \setminus R(n)$, $x < y$ follows from $R(n) < R(n^+) \setminus R(n)$ (which is property (b^+) in 5.8). ■

Corollary 5.10.

(a) *Every set $x \neq 0$ has an $<$-largest and an $<$-smallest element.*
(b) $x \in y \rightarrow x < y$.

Proof. (a) follows from 5.2, and (b) from 4.5 and property (2) in 5.9. ■

Theorem 5.11 (Induction scheme for inequality). *For any formula $\varphi(x)$,*

$$\forall x[\forall(y < x)\varphi(y) \rightarrow \varphi(x))] \rightarrow \forall x\varphi(x).$$

Proof. Suppose the premise $\forall x[\ldots]$ of the above implication holds and $\neg\varphi(x_0)$ is possible, where $x_0 \in R(n_0^+)$. Then the set $z_0 = \{u \in R(n_0^+) : \neg\varphi(u)\}$ is non-empty and $\min_<(z)$ exists (see 5.2). We can assume that $\min_<(z) = x_0$, whence $x_0 \in R(n_0^+)$ and $\mathrm{rank}(x_0) \leq n_0$. Now, for any y,

$$y < x_0 \rightarrow \mathrm{rank}(y) \leq \mathrm{rank}(x_0) \leq n_0,$$

by 5.9(2), whence $y < x_0 \rightarrow y \in R(n_0^+)$. By the definition of z and x_0, it now follows that $y < x_0 \rightarrow \varphi(y)$, i.e., $\forall(y < x_0)\varphi(y)$. Thus $\varphi(x_0)$, by the assumed premise $\forall x[\ldots]$, contrary to the choice of x_0. ■

Theorem 5.12. $\exists z\forall y(y < x \leftrightarrow y \in z)$.

Proof. We see from (2) in 5.9 that $y < x \rightarrow \mathrm{rank}(y) \leq \mathrm{rank}(x)$. Thus

$$\vdash y < x \rightarrow y \in R((\mathrm{rank}(x))^+),$$

by the definition (4.4) of rank, and 4.2(a). By the Comprehension Scheme 1.7, the required z may now be taken as

$$z = \{y \in R((\mathrm{rank}(x))^+) : y < x\}.$$

Definition 5.13. For every x, we shall denote by $[0, x)$ the unique z that satisfies $\forall y(y \in z \leftrightarrow y < x)$.

Theorem 5.14 (Definition of a p-function by recursion on inequality). *For every binary p-function $G(x, y)$ there exists a p-function F such that*

$$\vdash F(x) = G(x, F{\restriction}[0, x)).$$

We omit the straightforward proof (similar to that of 4.9).

6. The standard model $\mathfrak{S}$ of HF

We conclude the Appendix with a description of the standard model of HF. Such a description cannot be given formally in HF; it belongs to the meta-theory and to formalize it one would need (a fragment of) the ZF set theory.

When viewed as an object of the ZF set theory, the standard model of HF is the structure $\mathfrak{S} = \langle \mathbb{S}, 0, \in, \triangleleft \rangle$, where $\mathbb{S}$ is the smallest (infinite) set that contains the empty set 0 and is closed under the operation $\triangleleft$ of adding one element (i.e., $\mathbb{S}$ has the property that for any x, y in $\mathbb{S}$, also $x \cup \{y\}$ is in $\mathbb{S}$). However, it does not seem appropriate to involve the more powerful ZF theory in order to describe a model of HF, so we rather choose for a construction using a part of the meta-theory of HF and only as much of ZF (informally) as is needed. To begin with, we establish some properties of constant terms.

Definition 6.1. We denote by $\mathbb{C}$ the class of constant terms, i.e., $\mathbb{C}$ is the smallest class of terms such that $0 \in \mathbb{C}$ and $\sigma \triangleleft \tau \in \mathbb{C}$ for any $\sigma, \tau \in \mathbb{C}$. For every $\tau \in \mathbb{C}$ the number of appearances of $\triangleleft$ in τ will be called the *length* of τ and denoted by $l(\tau)$.

Lemma 6.2. *For every $0 \neq \tau \in \mathbb{C}$ there are finitely many $\tau_1, \ldots, \tau_m \in \mathbb{C}$, all shorter than τ, such that*

$$\vdash x \in \tau \leftrightarrow x = \tau_1 \vee \ldots \vee x = \tau_m.$$

Proof. The shortest $\tau \neq 0$ is $0 \triangleleft 0$. We have the theorem

$$\vdash x \in 0 \triangleleft 0 \leftrightarrow x \in 0 \vee x = 0 \leftrightarrow x = 0,$$

i.e., $m = 1$ and τ_1 is 0. Now assume that the lemma has been established for all non-zero constant terms shorter than τ, and let τ be $\sigma \triangleleft \mu$. Then σ is shorter than τ, hence there are $\sigma_1, \ldots, \sigma_n \in \mathbb{C}$ such that

$$\vdash x \in \sigma \leftrightarrow x = \sigma_1 \vee \ldots \vee x = \sigma_n.$$

Since $\vdash x \in \tau \leftrightarrow x \in \sigma \vee x = \mu$, we get

$$\vdash x \in \tau \leftrightarrow x = \sigma_1 \vee \ldots \vee x = \sigma_n \vee x = \mu.$$

By inductive assumption, each σ_i is shorter than σ, thus also shorter than τ. Also μ is shorter than τ. ■

Lemma 6.3. *Let $\sigma, \tau \in \mathbb{C}$ be such that* NON $\vdash \sigma = \tau$ *(i.e., $\sigma = \tau$ is not a theorem of* HF*). Then, for some $\nu \in \mathbb{C}$, one of the two possibilities occurs*:

(A) $\vdash \nu \in \sigma$ *and* $\vdash \nu \notin \tau$.

(B) $\vdash \nu \notin \sigma$ *and* $\vdash \nu \in \tau$.

Proof. To begin with, let us establish (A) and (B) when one of the terms σ, τ is 0. Suppose σ is 0, and assume that $\sigma = \tau$ is not a theorem. Then obviously τ is not the term 0, so τ is of the form $\mu \triangleleft \nu$ and we have $\vdash \nu \in \tau$. Clearly $\vdash \nu \notin 0$. Thus (B) follows. Analogously, if τ is 0 and $\sigma = \tau$ is not a theorem, we get (A).

We now prove the lemma by induction on the length $l(\sigma)$. The case $l(\sigma) = 0$ has been dealt with already. Suppose that $l(\sigma) > 0$ and that the lemma has been shown in every case (i.e., for every τ) when σ is replaced by a term shorter than σ. Suppose further that $\sigma = \tau$ is not a theorem. We should show that then, for every τ, there is a ν such that either (A) or (B). This was already shown when τ is 0. So it remains to deal with the situation when neither σ nor τ is 0. Then, by 6.2, there are $\sigma_1, \ldots, \sigma_n \in \mathbb{C}$, all shorter than σ, and $\tau_1, \ldots, \tau_m \in \mathbb{C}$, all shorter than τ, such that

(1) $\vdash x \in \sigma \leftrightarrow x = \sigma_1 \vee \ldots \vee x = \sigma_n$.

(2) $\vdash x \in \tau \leftrightarrow x = \tau_1 \vee \ldots \vee x = \tau_m$.

Suppose for a moment that for every σ_i there is a τ_j such that $\vdash \sigma_i = \tau_j$, and also for every τ_j there is a σ_i such that $\vdash \sigma_i = \tau_j$. It is clear that then (1) and (2) above imply $\vdash (x \in \sigma \leftrightarrow x \in \tau)$, i.e., $\vdash \sigma = \tau$. As this contradicts our assumption, at least one of the two possibilities must occur.

(a) There is a σ_i such that NON $\vdash \sigma_i = \tau_j$ for all τ_j. Since σ_i is shorter than σ, we conclude by the inductive assumption that either (A) or (B) holds for σ_i (in place of σ) and every τ_j (in place of τ), for suitable ν_j. Hence $\vdash \sigma_i \neq \tau_j$ for all j. We now replace x by σ_i in

$$\vdash x \notin \tau \leftrightarrow x \neq \tau_1 \wedge \ldots \wedge x \neq \tau_m$$

(see (2)), concluding $\vdash \sigma_i \notin \tau$. On the other hand, $\vdash \sigma_i \in \sigma$, by (1). So (A) holds with σ_i in place of ν.

(b) There is a τ_j such that NON $\vdash \sigma_i = \tau_j$ for all σ_i. This case is quite analogous to the previous; so, proceeding as in (a), we show that (B) holds with τ_j in place of ν. ∎

Corollary 6.4. *For every $\sigma, \tau \in \mathbb{C}$,* NON $\vdash \sigma = \tau$ *implies* $\vdash \sigma \neq \tau$.

Corollary 6.5. *If $\sigma, \tau \in \mathbb{C}$ are such that $\vdash \nu \in \sigma \Leftrightarrow \vdash \nu \in \tau$ for every $\nu \in \mathbb{C}$, then $\vdash \sigma = \tau$.*

Definition 6.6. To define the structure $\mathfrak{S} = \langle \mathbb{S}, 0, \in, \triangleleft \rangle$ we introduce the following:

(a) An equivalence relation $\equiv$ on $\mathbb{C}$ given by

$$\sigma \equiv \tau \Leftrightarrow \vdash \sigma = \tau.$$

(b) For every $\tau \in \mathbb{C}$, the $\equiv$-equivalence class $[\tau]$ containing τ.

(c) The set $\mathbb{S}$ of all equivalence classes $[\tau]$, i.e., $\mathbb{S} = \{[\tau] : \tau \in \mathbb{C}\}$.

(d) The binary relation $\in$ on $\mathbb{S}$ defined by

$$[\tau] \in [\mu] \Leftrightarrow \vdash \tau \in \mu.$$

(e) The binary operation $\triangleleft$ on $\mathbb{S}$ defined by

$$[\tau] \triangleleft [\mu] = [\tau \triangleleft \mu].$$

(f) The abbreviation 0 for [0].

It is easy to see that this structure is well defined, i.e., $[\tau] \in [\mu]$ and $[\tau] \triangleleft [\mu]$ depend only on the equivalence classes $[\tau]$, $[\mu]$.

Proposition 6.7. *If* HF *is consistent, then $\mathfrak{S}$ is a model of* HF.

Proof. Let us check that HF1, HF2 and HF3 (see Section 1) hold in $\mathfrak{S}$.

HF1. Suppose $[\sigma] \in \mathbb{S}$ is such that $\mathfrak{S} \models [\tau] \notin [\sigma]$ for every τ. Then σ cannot be of the form $\mu \triangleleft \nu$, because $[\nu] \in [\mu \triangleleft \nu]$ holds in $\mathfrak{S}$ (see 6.6(d)). Thus σ is 0. Conversely, we have $\mathfrak{S} \models [\tau] \notin [0]$ for all τ, because $\mathfrak{S} \models [\tau] \in [0]$ is equivalent to $\vdash \sigma \in 0$ and this contradicts Theorem 1.1 and the assumption of consistency.

HF2. By 6.5, the Extensionality Property 1.2 holds for $\mathfrak{S}$. Thus HF2 will be true in $\mathfrak{S}$ if we show that for every $\tau, \mu, \nu \in \mathbb{C}$,

$$\vdash \tau \in \mu \triangleleft \nu \ \Leftrightarrow\ (\vdash \tau \in \mu \text{ OR } \vdash \tau = \nu).$$

To verify $\Rightarrow$, assume $\vdash \tau \in \mu \triangleleft \nu$. If $\vdash \tau = \nu$, then we are done. If NON $\vdash \tau = \nu$, then by 6.4, $\vdash \tau \neq \nu$. So, in this case,

$$\vdash \tau \in \mu \triangleleft \nu \ \Rightarrow\ \vdash (\tau \in \mu \vee \tau = \nu) \wedge (\tau \neq \nu) \ \Rightarrow\ \vdash \tau \in \mu.$$

The verification of $\Leftarrow$ is immediate:

$$(\vdash \tau \in \mu \text{ OR } \vdash \tau = \nu) \ \Rightarrow\ \vdash (\tau \in \mu \vee \tau = \nu) \ \Rightarrow\ \vdash \tau \in \mu \triangleleft \nu.$$

HF3. For each equivalence class $a \in \mathbb{S}$, let $h(a)$, the *height* of a, be the smallest $l(\tau)$ (see 6.1) such that $\tau \in a$. It is easy to see that if $a \neq 0$, then there are $b, c \in \mathbb{S}$ such that $\mathbb{S} \models (a = b \triangleleft c)$ and $h(b), h(c) < h(a)$. Given a formula $\alpha(x)$ with one free variable x, let $A_\alpha \subseteq \mathbb{S}$ be the truth set of α, i.e.,

$$A_\alpha = \{a \in \mathbb{S} : \mathbb{S} \models \alpha(a)\}.$$

Assuming that $0 \in A_\alpha$ and moreover that $a, b \in A_\alpha$ implies $a \triangleleft b \in A_\alpha$ for all $a, b \in \mathbb{S}$, we deduce by induction on the height $h(c)$ that $c \in A_\alpha$ for every $c \in \mathbb{S}$. Thus $A_\alpha = \mathbb{S}$, and HF3 is verified. ■

References

[A] W. Ackermann, *Die Wiederspruchsfreiheit der allgemeinen Mengenlehre*, Math. Ann. 114 (1937), 305–315.

[A-B] Z. Adamowicz and T. Bigorajska, *Existentially closed structures and Gödel's second incompleteness theorem*, J. Symbolic Logic 66 (2001), 349–356.

[Be] E. W. Beth, *Axiomatique de la théorie des ensembles sans axiome de l'infini*, Bull. Soc. Math. Belg. 16 (1964), 127–136.

[Bo] G. Boolos, *The Logic of Provability*, Cambridge Univ. Press, Cambridge, 1993.

[F] S. Feferman, *Arithmetization of metamathematics in a general setting*, Fund. Math. 49 (1960), 35–92.

[GT] S. Givant and A. Tarski, *Peano arithmetic and the Zermelo-like theory of sets with finite ranks*, Notices Amer. Math. Soc. 24 (1977), A-437.

[H-B] D. Hilbert and P. Bernays, *Grundlagen der Mathematik*, Vol. 2, Springer, Berlin, 1939.

[J] T. Jech, *On Gödel's second incompleteness theorem*, Proc. Amer. Math. Soc. 121 (1994), 311–313.

[Ki] M. Kikuchi, *Kolmogorov complexity and the second incompleteness theorem*, Arch. Math. Logic 36 (1997), 437–443.

[Kl] S. C. Kleene, *Introduction to Metamathematics*, Van Nostrand, 1952; Russian translation by A. S. Yesenin-Vol'pin, Izdat. Inostr. Liter., Moscow, 1957.

[Ko] H. Kotlarski, *Other proofs of old results*, Math. Logic Quart. 44 (1998), 474–480.

[L] P. Lindström, *Aspects of Incompleteness*, Lecture Notes in Logic 10, Springer, Berlin, 1997.

[Mo] A. Mostowski, *Sentences Undecidable in Formalized Arithmetic*, North-Holland, Amsterdam, 1964.

[My] J. Mycielski, *The definition of arithmetic operations in the Ackermann model*, Algebra i Logika Sem. 3 (1964), no. 5–6, 64–65 (in Russian).

[Sh] J. R. Shoenfield, *Mathematical Logic*, Addison-Wesley, Reading, MA, 1967.

[Sm] C. A. Smorynski, *The incompleteness theorems*, in: Handbook of Mathematical Logic, J. Barwise (ed.), North-Holland, Amsterdam, 1977, 821–865.

[TG] A. Tarski and S. Givant, *A Formalization of Set Theory without Variables*, Amer. Math. Soc. Colloq. Publ. 41, Amer. Math. Soc., Providence, RI, 1987.

[TMR] A. Tarski with A. Mostowski and R. M. Robinson, *Undecidable Theories*, North-Holland, Amsterdam, 1953.

[V] P. Vopěnka, *A new proof of the Gödel's result of non-provability of consistency*, Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys. 14 (1966), 111–116.