

Arithmetization of metamathematics in a general setting

by

S. Feferman* (Stanford, Calif.)

1. Introduction

The method of arithmetization, as developed by Gödel [10], exploits the possibility of defining within a formal theory \mathcal{T} , or in arithmetical theories closely related to \mathcal{T} , various syntactical and logical notions concerning \mathcal{T} . In broad terms, the applications of the method can be classified as being *extensional* if essentially only numerically correct definitions are needed, or *intensional* if the definitions must more fully express the notions involved, so that various of the general properties of these notions can be formally derived.

The following are some results of extensional type: incompleteness theorems (Gödel's first underivability theorem [10] Satz VI, Rosser [29] Theorem II); non-definability of predicates in formal theories (Tarski [31], Kleene [15] Theorem XIII); undecidability of various theories (Rosser [29] Theorem III, Tarski, Mostowski and Robinson [32]); and degrees of unsolvability of various theories (Myhill [25], our [7]). Among the intensional results we have the following: unprovability of consistency statements (Gödel's second underivability theorem [10] Satz XI, comparison of theories by relative consistency proofs (Novak [26], Wang [36], [37], Shoenfield [30]); and ordinal logics (Turing [33], our [8]). A result of mixed character is the arithmetization of Gödel's completeness theorem

* The results reported in this paper were obtained while the author was a student of Professor Alfred Tarski at the University of California, Berkeley. A more complete presentation of them has been given in the author's thesis [4]; announcement of the results has also been made in [5] and [6].

We are indebted to Professor Tarski for a number of helpful suggestions regarding this research; as well as to Professor Leon Henkin for his kind guidance during the period 1955-56 when Professor Tarski was on leave. We wish also to thank Professors John Myhill and Georg Kreisel, both for a number of stimulating conversations, and also the latter for his helpful comments on a draft of this paper.

Finally, thanks are due to Professor Steven Orey for his interest in widening the range of application of our work, as will be evidenced at various points in the text.

This paper was prepared under Contract DA-04-200-ORD-997 for the Office of Ordnance Research, U.S.A.

for first-order logic (Hilbert-Bernays [14] v. II, Wang [34]) whereby every (recursively axiomatizable) theory \mathcal{T} is interpretable in arithmetic when the formal statement of the consistency of \mathcal{T} has been added. Finally, the proofs of non-finite axiomatizability of various theories \mathcal{T} obtained by showing \mathcal{T} to be reflexive, i. e. that the consistency of every finite subtheory of \mathcal{T} is provable in \mathcal{T} (Mostowski [23], Kreisel and Wang [19], Montague [21], [22]), are instances where intensional methods are used to deduce purely extensional results.

The relative freedom with which one can choose the formal definitions in extensional applications, providing only that they are numerically correct, has had two consequences. The first is that it has been possible to obtain results thereby of very great generality (cf. Mostowski's treatment of incompleteness and non-definability theorems [24], and the book *Undecidable theories* [32]). The second is the possibility of using intensionally incorrect definitions to obtain improved extensional results. Such is the case with the Rosser construction [29], starting with a given numerical definition $\text{Prf}_{\mathcal{T}}(x, y)$ of the relation "y is a proof in \mathcal{T} of x", of a new definition $\text{Prf}'_{\mathcal{T}}(x, y)$ as

$$\text{Prf}'_{\mathcal{T}}(x, y) \wedge \bigwedge_{z \leq y} \sim \text{Prf}_{\mathcal{T}}(\text{ng}(x), z).$$

Of course, $\text{Prf}'_{\mathcal{T}}(x, y)$ does not express the proof relation in \mathcal{T} but rather the relation "y is a proof in \mathcal{T} of x and for all $z \leq y$, z is not a proof in \mathcal{T} of the negation of x". Similar constructions are used by Kleene (symmetric form of Gödel's theorem, [15], Theorem XV) and by the author in [7].

Corresponding developments, of full generality and of useful variability of definitions, have so far been lacking in the treatment of intensional problems. A formal arithmetical statement $\text{Con}_{\mathcal{T}}$ expressing the consistency of a theory \mathcal{T} is only determined when we have been presented with a description of the set of axioms of \mathcal{T} from which we will be in a position to see that the given statement $\text{Con}_{\mathcal{T}}$ is intensionally correct. But when the set of axioms is regarded as being extensionally given, as simply being "there", no general criterion seems to be available for seeing whether one among many possible sentences $\text{Con}_{\mathcal{T}}$ correctly expresses the consistency of \mathcal{T} . Hence the generality of the following statement of Gödel's second undervivability theorem—if \mathcal{T} is sufficiently strong and consistent then $\text{Con}_{\mathcal{T}}$ is not provable in \mathcal{T} —is only apparent when regarded from the extensional point of view.

That the ability to deal with consistency statements in some generality would be useful can be seen in connection with Turing's treatment of ordinal logics [33]. For example, one might want to discuss a series of recursively axiomatizable theories \mathcal{T}_{ξ} , ξ a constructive ordinal

notation, where $\mathcal{T}_{\xi+1}$ is obtained from \mathcal{T}_{ξ} by adding $\text{Con}_{\mathcal{T}_{\xi}}$ as a new axiom. But just how these statements $\text{Con}_{\mathcal{T}_{\xi}}$ are to be found is not at all clear, for we do not see how to determine them without first having been presented with an intensional description of \mathcal{T}_{ξ} , and we do not know how to obtain such a description without first knowing how to determine the associated consistency statements. Turing, though he realized the difficulty, simply left the matter up in the air. In general, such difficulties have mostly been avoided in the literature, with results that are often ambiguous⁽¹⁾.

The situation is even less satisfying regarding manipulation of intensional definitions, in analogy to the Rosser construction. For example, if \mathcal{T} is consistent, we can construct a numerically correct definition of proof $\text{Prf}'_{\mathcal{T}}(x, y)$ as

$$\text{Prf}'_{\mathcal{T}}(x, y) \wedge \bigwedge_{z_1 \leq y} \bigwedge_{z_2 \leq y} \bigwedge_{w \leq y} \sim [\text{Prf}_{\mathcal{T}}(w, z_1) \wedge \text{Prf}_{\mathcal{T}}(\text{ng}(w), z_2)];$$

then we can prove in arithmetic the statement $\text{Con}'_{\mathcal{T}}$, taken to be

$$\bigwedge_x \bigwedge_{y_1} \bigwedge_{y_2} \sim [\text{Prf}'_{\mathcal{T}}(x, y_1) \wedge \text{Prf}'_{\mathcal{T}}(\text{ng}(x), y_2)].$$

But $\text{Prf}'_{\mathcal{T}}(x, y)$ is intensionally incorrect, so we can ascribe no clear intensional meaning to the result. On the other hand, since we cannot formally derive other properties of provability in terms of the definition $\text{Prf}'_{\mathcal{T}}(x, y)$, we see no results of extensional interest which follow from the proof of $\text{Con}'_{\mathcal{T}}$, in contrast to Rosser's use of such definitions.

One approach to the treatment of intensional problems in a general setting could be by means of the three derivability conditions set down for a proof definition $\text{Prf}_{\mathcal{T}}(x, y)$ by Hilbert and Bernays in [14], Vol. II, p. 285-286. These are shown to be sufficient for the proof of Gödel's second undervivability theorem (loc. cit. p. 286-288). On the other hand, we see how to verify for familiar theories that the intensionally correct proof definitions satisfy the given conditions (for example, for the arithmetics Z_{μ} and Z , cf. loc. cit. p. 289-328). Then it is sufficient, in a variety of intensional problems, to assume that one is dealing with such a proof definition, without needing to give it explicitly. The elegant possibilities in this direction are well demonstrated by Löb's solution

⁽¹⁾ We shall not try to list examples of such; the reader will not have to go far into the literature in this area to discover many instances of careless or vague formulations of intensional statements. We do not, by any means, contend to be the first to have questioned such treatment. As just one example, in a review (*Journal of Symbolic Logic* 16 (1951), p. 142) of one paper, Mostowski has written: "Full presentation of [the] results should be based on more exact definitions... of the proposition $\text{Con}(S)$there are many non-equivalent propositions of arithmetic which might be regarded as expressing the consistency of S ".

in [20] of a problem of Henkin's⁽²⁾. A similar assumption is used by Kreisel and Wang in [19].

This approach suffers several limitations. First of all, there is no simple method of seeing whether a given proof definition satisfies the derivability conditions. Further, the third derivability condition is peculiar to theories which contain a sufficient amount of arithmetic; it is thus not relevant to other problems of arithmetization, for example the treatment of Gödel's completeness theorem. On the other hand, the first two derivability conditions are rather too weak to be used by themselves with such other problems; it simply happens that they are sufficient (when accompanied by the third condition) to be used for the proof of the second underivability theorem. Finally, one does not see clearly from the conditions what are the possibilities of manipulating those proof definitions which satisfy them.

The purpose of this paper is to present a systematic approach to intensional problems in a general setting which does not suffer these limitations. This is done for axiomatic theories framed within a fixed system of logic which, for purposes of definiteness, we take here to be first-order predicate logic with identity. Then the proof relation for a theory is completely determined by the set of (non-logical) axioms of the theory. With a given theory \mathcal{T} we can associate the class of all formulas $\tau(x)$ which numerically define the set of axioms of \mathcal{T} . Then, by formally copying our notion of logical proof, we can associate with each such formula τ , in a uniform way, a formula $\text{Prf}_\tau(x, y)$ and, in turn, a sentence Con_τ . It will be evident from the construction (§ 4) that $\text{Prf}_\tau(x, y)$ correctly expresses that y is a proof of x , via the first-order logic, from the set of axioms defined by τ . In other words, whenever a formula $\tau(x)$ can be recognized to express correctly that x is an axiom of \mathcal{T} , the associated sentence Con_τ will be recognized to be a correct expression of the proposition that \mathcal{T} is consistent. In this way all intensionally correct statements of consistency for familiar theories can be obtained as special cases.

It will be shown in § 4 that a large body of propositions from "general" metamathematics can be formally derived when one uses the

(²) It may happen that a problem which appears at first sight intensional, can be seen to have purely extensional significance. For example, if one considers the sentence φ which expresses that it is not provable in \mathcal{T} , the question of provability of φ in \mathcal{T} turns out to be independent of which definition $\text{Prf}_{\mathcal{T}}(x, y)$ is used, as long as it is numerically correct (Gödel's first underivability theorem). Henkin's question concerns the sentence ψ which expresses that it is provable in \mathcal{T} ; in contrast to the preceding, the question can be answered in two ways by suitable choice of proof definition. Thus, the latter problem remains intensional. This point has been discussed in more detail by Kreisel in [17].

formulas $\text{Prf}_\tau(x, y)$ no matter what the formula τ is; the first two derivability conditions of Hilbert-Bernays are just very special instances of these propositions. The investigation beyond this basis then proceeds in two ways. Generality is achieved for various metamathematical propositions by finding those classes of formulas $\tau(x)$ for which they are valid. Useful variations are obtained by changing the proof definitions from the "inside", by taking different choices of formulas $\tau(x)$ for a given theory \mathcal{T} ; this is in contrast to changes from the "outside" as in the Rosser construction, which are, in effect, changes of the notion of logical derivation.

Following these lines, we find a large class of formulas, the so-called *RE*-formulas, for which Gödel's second underivability theorem can be obtained in quite general form, roughly as follows: *if \mathcal{T} is a sufficiently strong consistent theory and $\tau(x)$ is an RE-formula which defines the axioms of \mathcal{T} then $\text{not } \vdash_{\mathcal{T}} \text{Con}_\tau$* (5.6). (A version of the third derivability theorem used in this proof is found in 5.5.) On the other hand, we show some limitation on $\tau(x)$ is necessary for sufficiently strong reflexive theories; for example, we construct a definition $\pi^*(x)$ of the set axioms of Peano's arithmetic \mathcal{P} , for which $\vdash_{\mathcal{P}} \text{Con}_{\pi^*}$ (5.9, 5.10).

Because the formulas $\text{Prf}_\tau(x, y)$ accurately reflect logical derivability they prove to be especially useful in connection with problems of interpretability. Let us write $\mathcal{T} \preceq \mathcal{S}$ if \mathcal{T} is relatively interpretable in \mathcal{S} (in the sense of Tarski [32]). Then the arithmetized version of Gödel's completeness theorem takes the following form (roughly): *if $\tau(x)$ is any formula which defines the set of axioms of a theory \mathcal{T} then $\mathcal{T} \preceq \mathcal{P} + \{\text{Con}_\tau\}$* (6.2). In contrast to the second underivability theorem, there are essentially no restrictions on τ here; under certain simple restrictions, it is possible to replace \mathcal{P} by some finitely axiomatizable theory \mathcal{F} in the above (6.3). Because of the freedom to choose τ , this theorem is very close to being extensional; however, the fact that the structure of the statement Con_τ is intensionally correct with respect to logical derivability is essential in its proof.

By a systematic exploitation of statements of the sort for which $\vdash_{\mathcal{P}} \text{Con}_{\pi^*}$, it is possible to obtain clear extensional results of the following sort:

- (1) *if every finite subtheory of an axiomatizable theory \mathcal{T} is $\preceq \mathcal{P}$ then $\mathcal{T} \preceq \mathcal{P}$* (6.9; communicated to us by S. Orey);
- (2) *\mathcal{P} is not only not finitely axiomatizable, but it is also not interpretable in any of its finite subtheories* (6.8).

We further obtain the following results of mixed (extensional and intensional) type;

- (3) *for any RE-formula π defining the axioms of \mathcal{P} , $\mathcal{P} + \{\sim \text{Con}_\pi\} \preceq \mathcal{P}$* (6.6);

(4) for any RE-formula π defining the axioms of \mathcal{P} , $\mathcal{P} + \{\text{Con}_\pi\} \not\leq \mathcal{P}$ (6.5).

If \mathcal{T} is a theory with a finite set of axioms, there is a natural definition $\tau(x)$ of this set, obtained by listing the axioms; we denote this formula by $[\mathcal{T}](x)$. It has the further property that if τ' is any other formula which defines the axioms of \mathcal{T} then

$$\vdash_{\mathcal{P}} \text{Con}_{\tau'} \rightarrow \text{Con}_{[\mathcal{T}]}$$

Hence if we define $\tau \leq \tau'$ by $\vdash_{\mathcal{P}} \text{Con}_{\tau'} \rightarrow \text{Con}_{\tau}$, the natural description of \mathcal{T} is a minimum in the partial ordering \leq . It should be intuitively clear that there is no natural or favored description τ of an extensionally given infinite set of axioms. However, it may still be conceived that there is one favored way of expressing the consistency of \mathcal{T} , i. e. a choice of definition τ which is a minimum in the above ordering. We show in 7.4 that this is in general not the case, for \mathcal{T} a reflexive theory. Thus a thoroughgoing intensional approach to problems for arbitrary theories \mathcal{T} is not possible. § 7 concludes with a discussion of similar difficulties which are connected with any attempts to make general statements about formal relative consistency proofs. § 8 shows how the results of §§ 5-7 can be extended to theories which do not contain the symbols of arithmetic directly.

The reader will find that in general the material presented in §§ 2-4 is quite familiar; we have thus omitted most proofs there. However, we have stated all notions and results precisely, as a necessary prerequisite to dealing with their arithmetized versions. The explicit introduction and treatment of the classes of PR- and RE-formulas in § 3 is apparently new. So also is the system of notation for arithmetized concepts presented in § 4; we hope it will be found more manageable than the systems presently available in the literature. Proofs are given for all essentially new results in §§ 5-8.

We have tried to pitch our presentation somewhere between the level of an informal exposition and of a treatise in which all supporting details would be worked out⁽³⁾. It should be found quite adequate as a basis for further investigations. In particular, a detailed study of ordinal logics based in the present methods (along the lines described in our abstracts [8]) will follow in the near future.

⁽³⁾ Work is underway on a monograph, being written in collaboration with Richard Montague, on the method of arithmetization and some of its applications. It is intended that this monograph should contain a completely detailed presentation of the work discussed here, as well as of the research of Montague in [21] and [22].

2. Some preparations

2(a) Mathematical notions and notation. We shall primarily be concerned here with certain functions on and relations in the set ω of natural numbers $0, 1, 2, \dots$. All unspecified mathematical variables will be presumed here to range over ω . We shall occasionally use a few logical symbols in our informal discussion, for convenience of abbreviation. These are

$$=, (\exists x), \text{ and } (\exists x)_{x < a}$$

(if and only if, there exists an x , and there exists an x with $x < a$), with the quantified variables ranging over natural numbers.

We denote by ω^n the set of all ordered n -tuples $\langle a_0, \dots, a_{n-1} \rangle$ where for each $i < n$, $a_i \in \omega$. We take $\langle a \rangle = a$, so that $\omega^1 = \omega$. We call R an n -ary relation if $R \subseteq \omega^n$ and also write $R[a_0, \dots, a_{n-1}]$ if $\langle a_0, \dots, a_{n-1} \rangle \in R$. For n -ary relations R and S , $R \cup S$, $R \cap S$ and $R - S$ are taken in their usual meaning. The empty set (and n -ary relation for any n) is here denoted by 0 . The set consisting of just a_0, \dots, a_k is denoted by $\{a_0, \dots, a_k\}$. By an n -ary function or operation we understand an $(n+1)$ -ary relation F such that for any a_0, \dots, a_{n-1} there is a unique y , for which $F[a_0, \dots, a_{n-1}, y]$; the unique such y is denoted by $F(a_0, \dots, a_{n-1})$. If R is an $(n+1)$ -ary relation the equation

$$F(a_0, \dots, a_{n-1}, b) = (\mu y)_{y < b} R[a_0, \dots, a_{n-1}, y]$$

defines the function which has as value the least $y < b$ such that $R[a_0, \dots, a_{n-1}, y]$ if $(\exists y)_{y < b} R[a_0, \dots, a_{n-1}, y]$, and otherwise has value b .

We take as known the theory of primitive and general recursive functions and relations; Kleene's book [15] will serve as a standard reference here, as well as on other general questions of metamathematics. We shall also refer to certain relations as being recursively enumerable, namely those which can be defined in the form

$$(\exists y) S[a_0, \dots, a_{n-1}, y]$$

where S is recursive (or, equivalently, primitive recursive).

The function P which enumerates the primes, starting with 2, in order of increasing size, is primitive recursive; we shall write p_i for $P(i)$. We can then introduce a primitive recursive function $(a)_i$ of two variables a and i such that for any $a \neq 0$, $(a)_i$ is the unique k for which $p_i^k | a$ but $p_i^{k+1} \nmid a$. We shall say that a represents a sequence, and write $a \in Sq$ if $a \neq 0$ and whenever $p_i | a$ and $j < i$, then $p_j | a$. Let $L(a)$ denote the highest n such that $p_{n-1} | a$, if $a > 1$, and let $L(1) = 0$. Then any $a \in Sq$ corresponds in a natural way to the sequence of non-zero numbers $\langle (a)_0, \dots, (a)_{L(a)-1} \rangle$ (the empty sequence if $L(a) = 0$). The set Sq and function L are primitive recursive.

2(b) Syntax. We restrict our considerations here to theories formalized within first-order classical logic with identity. It will be clear to the reader how to extend the methods developed to classical theories with variables for higher types, either directly or by reduction. We start by describing a general language from which all our theories can be constructed. This is a language only in a generalized sense of the word. Its basic elements are taken to be numbers, rather than symbols or expressions, and the logical connectives will correspond to certain operations on numbers. By this device, every expression will be directly identified with what is usually called its Gödel-number, and we thus avoid the constant passage back and forth between expressions and numbers.

To be definite, we can set out, as a basic list of symbols:

$$v_n = 8n + 25, \quad c_n = 8n + 3, \quad r_{n,m} = 8 \left[\frac{(n+m)(n+m+1)}{2} + m \right] + 5,$$

$$f_{n,m} = 8 \left[\frac{(n+m)(n+m+1)}{2} + m \right] + 7,$$

referred to respectively as the $(n+1)$ -st variable, $(n+1)$ -st constant, $(m+1)$ -st $(n+1)$ -placed relation and the $(m+1)$ -st, $(n+1)$ -placed function, symbol. In the following we shall also fix x, y, z to be v_0, v_1, v_2 respectively. The set of all v_n is denoted by Vr . The set of constant, relation and function symbols is referred to as the set of *non-logical constants*, denoted by *Const*. We then define three operations on numbers, $\sim a = 2^1 \cdot 3^a$, $a \rightarrow b = 2^0 \cdot 3^a \cdot 5^b$ and $\bigwedge_a b = 2^{17} \cdot 3^a \cdot 5^b$. In terms of these, operations $a \vee b$, $a \wedge b$, $a \leftrightarrow b$, $\bigvee_a b$ can be defined in the usual manner. As a way of conveniently describing the construction of terms and atomic formulas, we make the abbreviation

$$a[b_0, \dots, b_n] = p_0^{a_0} \cdot p_1^{b_0} \cdot \dots \cdot p_{n+1}^{b_n}.$$

We further abbreviate

$$r_{1,0}[b_0, b_1] \quad \text{by} \quad b_0 \approx b_1.$$

In the following, unless otherwise specified, K will represent an arbitrary finite set such that

$$r_{1,0} \in K \subseteq \text{Const}.$$

By the set of *terms in K*, Tm_K , we understand the smallest set which contains all variables, all $c_n \in K$ and such that whenever it contains ξ_0, \dots, ξ_n and whenever $f_{n,m} \in K$ it also contains $f_{n,m}[\xi_0, \dots, \xi_n]$. The set of *atomic formulas in K*, At_K , is the set of all numbers $r_{n,m}[\xi_0, \dots, \xi_n]$ where $\xi_0, \dots, \xi_n \in Tm_K$ and $r_{n,m} \in K$. The set of *formulas of K*, Fm_K , is

the smallest set which contains At_K and such that whenever it contains φ, ψ also contains $\sim\varphi, \varphi \rightarrow \psi$ and $\bigwedge_{v_n} \varphi$. It is easily seen then that the sets Tm_K, At_K and Fm_K are primitive recursive. Furthermore, it results from our particular construction (as it would from any similar construction) of the basic symbols and operations that every term and formula of K is uniquely "readable", i. e. has a unique grammatical structure.

In this way we can continue to develop syntax in the usual fashion but at the same time strictly as a branch of number theory. Among other concepts needed are that of a variable being *free* in a formula, that of the *set of free variables of φ* , $Fv(\varphi)$, that of the *set of sentences in K*, St_K , and that of one formula γ being obtained by *generalization* (preceding by universal quantifiers) from another φ . We also assume to be defined an operation $Sb_{\xi}^a \gamma$, to be read the substitution of (the term) ξ for (the variable or constant) a at all free occurrences of a in (the term or formula) γ . It shall be understood that this operation is always performable, with *bound variables being renamed in γ , if necessary, to avoid collisions* (cf. [14], Vol. II, p. 378). More generally, for $\xi = p_0^{\xi_0} \cdot \dots \cdot p_n^{\xi_n}$, $a = p_0^{a_0} \cdot \dots \cdot p_n^{a_n}$, we assume available a *simultaneous substitution operation* $SSb(\xi, a, \gamma)$, ordinarily to be written

$$Sb \begin{pmatrix} a_0 & a_n \\ \dots & \dots \\ \xi_0 & \xi_n \end{pmatrix} \gamma,$$

which reduces to the preceding operation in case $n = 0$. In particular, we shall write

$$\gamma(\xi_0, \dots, \xi_n) \quad \text{for} \quad Sb \begin{pmatrix} v_0 & v_n \\ \dots & \dots \\ \xi_0 & \xi_n \end{pmatrix} \gamma.$$

Thus writing " $\varphi(v_0)$ " or " $\varphi(x, y)$ ", for example, denotes the same as writing " φ ", and does not imply any commitment regarding the set of free variables of φ . All of the relations and operations described here are primitive recursive and their basic properties can be deduced by ordinary number-theoretical arguments, including mathematical induction. Incidentally, it is useful to observe that the ordering $<$ of natural numbers automatically provides us with an ordering of all our syntactic entities.

2(c) Logic⁽⁴⁾. As the set Ax_K of *logical axioms in K* we take the set of *all generalizations of all formulas $\varphi \in Fm_K$ where φ satisfies one*

(4) The systematization of logic presented in this section follows closely one used by Professor Tarski in his course in metamathematics given in Berkeley 1951-1952.

of the following conditions, for some $\theta, \psi, \chi \in Fm_K$, $\xi \in Tm_K$, and $u \in \mathcal{V}_r$:

- (i) $\varphi = [\theta \rightarrow (\psi \rightarrow \chi)] \rightarrow [(\theta \rightarrow \psi) \rightarrow (\theta \rightarrow \chi)]$;
- (ii) $\varphi = \theta \rightarrow (\psi \rightarrow \theta)$;
- (iii) $\varphi = \theta \rightarrow (\sim \theta \rightarrow \psi)$;
- (iv) $\varphi = (\theta \rightarrow \psi) \rightarrow [(\sim \theta \rightarrow \psi) \rightarrow \psi]$;
- (v) $\varphi = \bigwedge_u (\theta \rightarrow \psi) \rightarrow (\bigwedge_u \theta \rightarrow \bigwedge_u \psi)$;
- (vi) $\varphi = \bigwedge_u \psi \rightarrow Sb_{\xi}^u \psi$;
- (vii) $\varphi = \psi \rightarrow \bigwedge_u \psi$ (u not free in ψ);
- (viii) $\varphi = \bigwedge_x (x \approx x)$;
- (ix) $\varphi = \bigwedge_{v_0} \dots \bigwedge_{v_{n+1}} (\forall_i \approx v_{n+1} \rightarrow \xi \approx Sb_{v_{n+1}}^{v_i} \xi)$, where $i \leq n$ and
 $\xi = f_{n,m}[v_0, \dots, v_n]$;
- (x) $\varphi = \bigwedge_{v_0} \dots \bigwedge_{v_{n+1}} (\forall_i \approx v_{n+1} \rightarrow (\psi \leftrightarrow Sb_{v_{n+1}}^{v_i} \psi))$, where $i \leq n$ and
 $\psi = r_{n,m}[v_0, \dots, v_n]$.

Clearly Ax_K is a primitive recursive set.

We wish next to describe the basic proof-theoretic notions relative to a set of non-logical axioms A in a language with symbols from K . We shall be primarily interested in the case that A is a set of sentences in K . We shall call a system $\mathcal{A} = \langle A, K \rangle$ an *axiom system* or *axiomatic theory* if $A \subseteq St_K$. It is technically convenient to deal with a more general notion; we call any system $\mathcal{A} = \langle A, K \rangle$ a *pseudo-system* or *pseudo-theory*, if $A \subseteq Fm_K$. The system $\mathcal{L}_K = \langle 0, K \rangle$ will correspond to the pure logic in K . Given any n , let $A \upharpoonright n$ be the set of $\varphi \in A$ with $\varphi \leq n$ and let $\mathcal{A} \upharpoonright n = \langle A \upharpoonright n, K \rangle$; $\mathcal{A} \upharpoonright n$ is thus a certain finite subsystem of \mathcal{A} . For any $\mathcal{A} = \langle A, K \rangle$ and $\varphi_0, \dots, \varphi_n \in Fm_K$, we define $\mathcal{A} + \{\varphi_0, \dots, \varphi_n\} = \langle A \cup \{\varphi_0, \dots, \varphi_n\}, K \rangle$. We assume that $\mathcal{A} = \langle A, K \rangle$ is an arbitrary pseudo-theory in the following.

By $Prf_{\mathcal{A}}$ we understand the binary relation such that: for any φ, Ψ , $Prf_{\mathcal{A}}[\varphi, \Psi]$ if and only if $\Psi \in Sq$ and $\varphi = (\Psi)_{L(\Psi)-1}$ and for each $i < L(\Psi)$, $(\Psi)_i \in Fm_K$ and either

- (i) $(\Psi)_i \in Ax_K$,
- or (ii) $(\Psi)_i \in A$,
- or (iii) for some $j, k < i$, $(\Psi)_k = (\Psi)_{j \rightarrow} (\Psi)_i$.

In other words $Prf_{\mathcal{A}}[\varphi, \Psi]$ holds if Ψ represents a sequence of formulas giving a proof of φ from the logical axioms, together with the formulas of A , using *modus ponens* as the only rule of inference. The set

of formulas provable in A , $Pr_{\mathcal{A}}$, is taken to be the set of all φ such that for some Ψ , $Prf_{\mathcal{A}}[\varphi, \Psi]$ ⁽⁵⁾. We shall also write

$$\vdash_{\mathcal{A}} \varphi \quad \text{for} \quad \varphi \in Pr_{\mathcal{A}}.$$

It is easily seen that the relation $Prf_{\mathcal{A}}$ is primitive recursive in A and that, moreover, it is recursively enumerable whenever A is; the latter also holds true of $Pr_{\mathcal{A}}$.

The development of logic under this notion of provability is standard, with one exception. While we have allowed formulas to be used as axioms in arbitrary pseudo-theories \mathcal{A} , we have not provided a rule of generalization. In fact it is easy to give examples of a system \mathcal{A} , formula φ and variable u such that $\vdash_{\mathcal{A}} \varphi$ but not $\vdash_{\mathcal{A}} \bigwedge_u \varphi$. However, we can say the following for any pseudo-theory \mathcal{A} , and any given variable u :

2.1. If for each $\varphi \in A$, u is not a free variable of φ , then for any $\varphi \in Fm_K$,

$$\vdash_{\mathcal{A}} \bigwedge_u \varphi \quad \text{if and only if} \quad \vdash_{\mathcal{A}} \varphi.$$

Thus variables which are free in some formula of A act as constant symbols with respect to provability from \mathcal{A} . This special technical feature of our notion of provability will be found useful in later developments (2.5, 2.8, 2.9). It is seen that when \mathcal{A} is an axiom system, i. e. A is a set of sentences, the present notion of logical derivability accords with the usual one.

Among the familiar results concerning provability, we mention here but two. First, we have the Deduction Theorem.

2.2. For any $\varphi_0, \dots, \varphi_k, \psi \in Fm_K$,

$$\psi \in Pr_{\mathcal{A} + \{\varphi_0, \dots, \varphi_k\}} \quad \text{if and only if} \quad (\varphi_0 \wedge \dots \wedge \varphi_k) \rightarrow \psi \in Pr_{\mathcal{A}}.$$

Secondly, we express (in part) the usual result about deducibility from finite subsystems by the statement:

2.3. For any $\varphi \in Fm_K$,

$$\varphi \in Pr_{\mathcal{A}} \quad \text{if and only if, for some } n, \quad \varphi \in Pr_{\mathcal{A} \upharpoonright n}.$$

The notions of being *closed* (under logical deduction), *consistent*, and *complete*, as applied to a pseudo-theory \mathcal{A} , are explained respectively in the following definitions:

⁽⁵⁾ In extensional problems it is often unnecessary to distinguish a theory from its set of provable formulas (or sentences); such are, for example, the *theories in standard formalization* of [32]. However we may not be able to recognize by given mathematical means that two axiom systems $\mathcal{A} = \langle A, K \rangle$ and $\mathcal{A}' = \langle A', K \rangle$ determine the same theory in this sense, i. e. that $Pr_{\mathcal{A}} = Pr_{\mathcal{A}'}$. Hence it is essential, for intensional problems, to distinguish between different axiom systems at the start.

- 2.4. $\mathcal{A} \in CL$ if and only if $Pr_{\mathcal{A}} \subseteq A$;
 $\mathcal{A} \in CON$ if and only if for each $\varphi \in Fm_K$, either not $\vdash_{\mathcal{A}} \varphi$
or not $\vdash_{\mathcal{A}} \sim \varphi$;
 $\mathcal{A} \in CM$ if and only if $\mathcal{A} \in CL \cap CON$ and for each $\varphi \in Fm_K$,
either $\vdash_{\mathcal{A}} \varphi$ or $\vdash_{\mathcal{A}} \sim \varphi$.

It is easily seen that for any fixed $\varphi_0 \in Fm_K$,

$$\mathcal{A} \in CON \quad \text{if and only if not} \quad \vdash_{\mathcal{A}} \varphi_0 \wedge \sim \varphi_0.$$

Thus, this definition of consistency accords with any usual one. On the other hand, the definition of completeness given here is much more restrictive than the usual one; for example, if \mathcal{A} is an axiom system and $\mathcal{A} \in CM$, it follows that for any $\varphi \in Fm_K$, either $\vdash_{\mathcal{A}} \varphi$ or $\vdash_{\mathcal{A}} \bigwedge_n \sim \varphi$.

(From this it is seen that any axiom system which is complete in the sense of 2.4 would be the theory of a one element domain.) Actually, we shall only be interested in applying this notion to the case of certain pseudo-systems (in which case we cannot apply generalization). In this connection, we introduce one further type of completeness, which we refer to as *quantificational completeness*; we put

- 2.5. $\mathcal{A} \in QCM$ if and only if $\mathcal{A} \in CM$ and for each $\varphi \in Fm_K$ and $u \in Vr$,
 $\vdash_{\mathcal{A}} \bigwedge_u \varphi$ if and only if, for all $w \in Vr$, $\vdash_{\mathcal{A}} Sb_w^u \varphi$.

Since $CM \subseteq CL$, it makes no difference for $\mathcal{A} \in CM$ whether we write, for any given φ , $\vdash_{\mathcal{A}} \varphi$, $\varphi \in Pr_{\mathcal{A}}$, or $\varphi \in A$.

Concerning these notions, we obtain first, easily,

- 2.6. $\mathcal{A} \in CON$ if and only if for each n , $\mathcal{A} \upharpoonright n \in CON$.

Next, though the notion CM of completeness does not coincide with the usual one, we see that Lindenbaum's extension theorem still holds, namely:

- 2.7. If $\mathcal{A} \in CON$ then there exists $\mathcal{B} = \langle B, K \rangle$ such that $\mathcal{B} \in CM$ and $A \subseteq B$.

Indeed, it is useful to observe that such a set B can be constructed which is also arithmetically definable in terms of A , as follows. Put $\theta \in B$ if and only if there exists $\Phi \in Sq$ and n such that

- (i) $L(\Phi) = n+1$ and $\theta = (\Phi)_n$;
- (ii) for each $i \leq n$, $(\Phi)_i \in Fm_K$ and $(\Phi)_0 < (\Phi)_1 < \dots < (\Phi)_n$;
- (iii) $\sim((\Phi)_0 \wedge \dots \wedge (\Phi)_n) \notin Pr_{\mathcal{A}}$;
- (iv) if $\varphi \in Fm_K$ and $\varphi < (\Phi)_0$ then $\sim \varphi \in Pr_{\mathcal{A}}$;
- (v) if $\varphi \in Fm_K$, $i < n$ and $(\Phi)_i < \varphi < (\Phi)_{i+1}$ then $(\Phi)_0 \wedge \dots \wedge (\Phi)_i \rightarrow \sim \varphi \in Pr_{\mathcal{A}}$.

The proof then proceeds in the standard way (cf. [32], p. 15). Another extension theorem which will be of use to us is essentially due to Henkin

and Hasenjaeger (cf. [12] and [13]), and concerns quantificationally complete systems. We do not state it in its most general form here.

- 2.8. If $\mathcal{A} \in CON$ and $A \subseteq St_K$ then there exists $C = \langle C, K \rangle$ such that $C \in QCM$ and $A \subseteq C$.

The proof depends on the observation that if a pseudo-theory $\mathcal{D} = \langle D, K \rangle$ is consistent, $\varphi \in Fm_K$, u, w are variables such that $w \neq u$, and w is not free in any formula of $D \cup \{\varphi\}$, then $\mathcal{D} + \{Sb_w^u \varphi \rightarrow \bigwedge_u \varphi\}$ is consistent. Thus if F is a function which enumerates all formulas of the form $\bigwedge_u \varphi$, $F(n) = \bigwedge_{u_n} \varphi_n$, we easily construct a sequence of variables w_n such that the system $A' = \langle A', K \rangle$, where

$$A' = A \cup \{Sb_{w_0}^{u_0} \varphi_0 \rightarrow \bigwedge_{u_0} \varphi_0, \dots, Sb_{w_n}^{u_n} \varphi_n \rightarrow \bigwedge_{u_n} \varphi_n, \dots\},$$

is consistent. Then by 2.7 we can extend \mathcal{A}' to a complete system C which is then of necessity, quantificationally complete. We see that A' is recursive in A and hence, again, C is arithmetically definable in terms of A .

We say that a system \mathcal{A} is *finitely axiomatizable*, or *recursively axiomatizable*, if there is some system $C = \langle C, K \rangle$ such that $Pr_C = Pr_{\mathcal{A}}$ and C is finite, or recursive, respectively ⁽⁶⁾.

Given two systems $\mathcal{A} = \langle A, K \rangle$, $\mathcal{B} = \langle B, L \rangle$, \mathcal{B} is said to be an *extension* of \mathcal{A} , and \mathcal{A} a *subsystem* of \mathcal{B} , in symbols $\mathcal{A} \subseteq \mathcal{B}$, if $K \subseteq L$ and $Pr_{\mathcal{A}} \subseteq Pr_{\mathcal{B}}$; we shall say that \mathcal{A} is a finite subsystem of \mathcal{B} if, in addition, A is finite. We shall say that \mathcal{B} is an extension of \mathcal{A} with the same constants (or, without new constants) if $\mathcal{A} \subseteq \mathcal{B}$ but $K = L$. Finally, we shall in general put $\mathcal{A} \cup \mathcal{B} = \langle A \cup B, K \cup L \rangle$.

2(d). Models and interpretations. We shall make use of semantical notions as a guide to certain parts of our discussion and as a simplifying device in some proofs. However, it will be seen that all non-constructive methods can be eliminated from our treatment of problems of arithmetization, if desired.

We shall illustrate our use of the semantical notions by reference to a simple special case, namely when $K = \{r_{1,0}, r_{1,1}\}$. If we set $s = r_{1,1}$, we would say, in usual terms, that we are dealing with a system with basic symbols \approx, s , where s is a binary relation symbol. By a K -model we understand a system $\mathfrak{M} = \langle M, G \rangle$ where $M \neq 0$ and G is a function

⁽⁶⁾ By the result of Craig [1], \mathcal{A} is recursively axiomatizable if and only if there is a $C = \langle C, K \rangle$ such that $Pr_C = Pr_{\mathcal{A}}$ and C is primitive recursive (recursively enumerable). This will not, of course, keep us from distinguishing whether a given axiom system A is primitive recursive, recursive, or recursively enumerable (cf. ⁽⁶⁾). We shall prove a stronger result than Craig's in 4.13.

on K which assigns to each n -ary relation or function symbol of K an n -ary relation or function, resp., on M ; in our case, $G_s \subseteq M^2$ —moreover always $\langle a, b \rangle \in G_{r_1, 0}$ if and only if $a = b$. By an assignment W in M , we mean simply a sequence $W \in M^\omega$. W is said to *satisfy* a formula in \mathfrak{M} under the usual conditions (cf. [31]), stated for our case as follows:

- (i) W satisfies $r[\nabla_n, \nabla_m] \equiv \langle W(n), W(m) \rangle \in G_r$, for each $r \in K$;
- (ii) if $\varphi, \psi \in Fm_K$ then W satisfies $\sim\varphi \equiv W$ does not satisfy φ , and W satisfies $\varphi \rightarrow \psi \equiv W$ does not satisfy φ or W satisfies ψ ;
- (iii) if $\varphi \in Fm_K$ then

$$W \text{ satisfies } \bigwedge_{v_i} \varphi \equiv \text{for every } W' \text{ such that } W'(j) = W(j) \\ \text{for all } j \neq i, \text{ we have } W' \text{ satisfies } \varphi.$$

A useful transformation of condition (iii) can be made when W is an assignment whose range is M . In this case we have

(iii)' if $\varphi \in Fm_K$ then

$$W \text{ satisfies } \bigwedge_{v_i} \varphi \equiv W \text{ satisfies } Sb_{v_k}^{v_i} \varphi \text{ for every } k.$$

If $\varphi \in St_K$, φ is said to be *true* in \mathfrak{M} , and \mathfrak{M} is called a *model* of φ , if some assignment W satisfies φ in \mathfrak{M} .

The main theorem connecting semantical notions with logical notions is the following:

2.9. If $\mathcal{A} \in CON$ and $A \subseteq St_K$ then there exists a model $\mathfrak{M} = \langle M, G \rangle$ (with $M \subseteq \omega$) such that every sentence of A is true in \mathfrak{M} .

The proof makes use of the strongly complete extension \mathcal{C} of \mathcal{A} obtained in 2.8. We can define an equivalence relation between natural numbers n, p by the condition

$$(1) \quad \nabla_n \approx \nabla_p \in \mathcal{C}.$$

Then M can be taken to be the set of all equivalence classes. Alternatively, we can take M to be a set of natural numbers, with

$$p \in M \equiv \text{for all } n < p, \nabla_n \approx \nabla_p \notin \mathcal{C}.$$

To construct the model for our particular K we then take

$$(2) \quad \langle m, n \rangle \in G_s \equiv s[\nabla_m, \nabla_n] \in \mathcal{C} \quad (\text{for } m, n \in M).$$

If W is defined by

$$(3) \quad W(n) = (\mu p)(\nabla_n \approx \nabla_p \in \mathcal{C})$$

we see that W is an assignment with range M . It is then easy to verify, comparing conditions (i), (ii), (iii)' above with the fact that $\mathcal{C} \in QCM$, that \mathfrak{M} is a model for \mathcal{A} . In fact, it is seen that for any $\varphi \in Fm_K$,

$$(4) \quad W \text{ satisfies } \varphi \text{ in } \mathfrak{M} \equiv \varphi \in \mathcal{C}.$$

As we know, the Theorem 2.9 has as an immediate consequence the completeness, for the deduction of all sentences true in all models, of our formalization of logic (⁷).

The formal analog of the notion of model is that of interpretation. We take the notion, \mathcal{A} is *relatively interpretable* in \mathcal{B} , which we shall write $\mathcal{A} \preceq \mathcal{B}$, as explained by Tarski in [32], p. 20-21, 29. It is easy to see that this relation is the same as that to be described in the following; we again illustrate by considering the particular $K = \{r_{1,0}, r_{1,1}\}$. Consider $\mathcal{A} = \langle A, K \rangle$, $\mathcal{B} = \langle B, L \rangle$. Let ψ_s be a fixed formula of L with $Fv(\psi_s) = \{x, y\}$; let δ be a fixed formula of L with $Fv(\delta) = \{x\}$. Let I be the primitive recursive function uniquely determined on Fm_K by the following conditions:

(i) For any n, m

$$I(\nabla_n \approx \nabla_m) = (\nabla_n \approx \nabla_m)$$

and

$$I(s[\nabla_n, \nabla_m]) = \psi_s(\nabla_n, \nabla_m);$$

(ii) for any $\varphi, \psi \in Fm_K$, $I(\sim\varphi) = \sim I(\varphi)$ and

$$I(\varphi \rightarrow \psi) = I(\varphi) \rightarrow I(\psi);$$

(iii) for any $\varphi \in Fm_K$ and any variable u

$$I(\bigwedge_u \varphi) = \bigwedge_u \{ \delta(u) \rightarrow I(\varphi) \}.$$

We then write $\mathcal{A} \preceq_{\mathcal{I}} \mathcal{B}$ if $\vdash_{\mathcal{B}} \bigvee_x \delta(x)$ and for each $\varphi \in A$, $\vdash_{\mathcal{B}} I(\varphi)$.

Then, if \mathcal{A} is an axiom system, $\mathcal{A} \preceq \mathcal{B}$ if and only if there is an I such that $\mathcal{A} \preceq_{\mathcal{I}} \mathcal{B}$. In such an interpretation we can say that we are “modeling”

\mathcal{A} in \mathcal{B} with “domain” obtained by relativizing quantifiers to δ . (In case K contains function symbols, the formulation needed above becomes slightly more complicated. For example if K contains a single unary function symbol g , we find $I(\varphi)$ by first constructing φ' logically equivalent to φ , such that every atomic formula in φ' involving g has the form $g[\nabla_n] \approx \nabla_m$, and we then set $I(\varphi) = I(\varphi')$. In particular, $I(g[\nabla_n] \approx \nabla_m) = \psi_g(\nabla_n, \nabla_m)$, where

$$\vdash_{\mathcal{B}} \bigwedge_x \{ \delta(x) \rightarrow \bigvee_y [\delta(y) \wedge \psi_g(x, y) \wedge \bigwedge_z \{ \delta(z) \wedge \psi_g(x, z) \rightarrow y \approx z \}] \}.$$

(⁷) The completeness theorem is due to Gödel [9]. The present proof is based on a simplified version, found by Hasenjaeger [12] and independently by Henkin (cf. [12], footnote 3), of the proof given by Henkin in [13].

It is a direct matter to see that:

2.10. If $\mathcal{A}, \mathcal{B}, \mathcal{C}$ are axiom systems then:

- (i) if $\mathcal{A} \subseteq \mathcal{B}$ then $\mathcal{A} \preceq \mathcal{B}$;
- (ii) if $\mathcal{A} \preceq \mathcal{B}$ and $\mathcal{B} \preceq \mathcal{C}$ then $\mathcal{A} \preceq \mathcal{C}$;
- (iii) if $\mathcal{A} \preceq \mathcal{B}$ and $\mathcal{B} \in \text{CON}$ then $\mathcal{A} \in \text{CON}$.

Several other relations comparing theories which have been discussed in the literature are closely related to \preceq . Among these we may mention the notions of *translatability* introduced in [35], [36]; these also satisfy the result 2.10, in place of \preceq . However, we have found that these other notions are much less satisfactory to deal with in connection with the method of arithmetization (further remarks on this will be found in § 7).

3. Number theory and certain of its extensions

Let $K_0 = \{c_0, f_{0,0}, f_{1,0}, f_{1,1}\}$. For any ξ, η we shall set $\xi' = f_{0,0}[\xi]$, $\xi + \eta = f_{1,0}[\xi, \eta]$, and $\xi \cdot \eta = f_{1,1}[\xi, \eta]$. In other words, K_0 provides the basic symbols of arithmetic. We shall speak of K_0 -systems and K_0 -pseudo-systems, $\mathcal{A} = \langle A, K_0 \rangle$. The syntactic notion of *numeral* is introduced for such systems by the recursive condition $\bar{0} = c_0, (n+1) = (\bar{n})'$. Further $\neq, \leq, <$ are counted as abbreviations in the following sense; $\xi \neq \eta = \sim(\xi \approx \eta)$, $\xi \leq \eta = \bigvee_{v_n} (\xi + v_n \approx \eta)$, where v_n is the least variable to occur neither in ξ nor in η , and $\xi < \eta = \xi \leq \eta \wedge \xi \neq \eta$.

Two particular K_0 -theories in which we are interested are $\mathcal{Q} = \langle Q, K_0 \rangle$ ("Robinson's system") and $\mathcal{P} = \langle P, K_0 \rangle$ ("Peano's" or "classical" arithmetic). The sets Q, P both contain the following six sentences:

$$\bigwedge_x \bigwedge_y (x' \approx y' \rightarrow x \approx y), \quad \bigwedge_x (x' \neq \bar{0}), \quad \bigwedge_x (x + \bar{0} \approx x),$$

$$\bigwedge_x \bigwedge_y (x + y' \approx (x + y)'), \quad \bigwedge_x (x \cdot \bar{0} \approx \bar{0}), \quad \bigwedge_x \bigwedge_y (x \cdot y' \approx (x \cdot y) + x).$$

In addition, Q has the single axiom

$$\bigwedge_x (x \neq \bar{0} \rightarrow \bigvee_y x \approx y'),$$

while P has, instead, all sentences which are generalizations of the following, for some $\varphi \in \mathcal{F}m_{K_0}$:

$$\varphi(\bar{0}) \wedge \bigwedge_x (\varphi(x) \rightarrow \varphi(x')) \rightarrow \bigwedge_x \varphi(x).$$

Clearly $\mathcal{Q} \subseteq \mathcal{P}$.

In \mathcal{Q} we can obtain the following elementary arithmetical results:

- 3.1. (i) $\vdash_{\mathcal{Q}} \bar{n} + \bar{m} \approx \overline{n+m}$ and $\vdash_{\mathcal{Q}} \bar{n} \cdot \bar{m} \approx \overline{n \cdot m}$, for all n, m ;
- (ii) $\vdash_{\mathcal{Q}} \bar{n} \neq \bar{m}$ whenever $n \neq m$;
- (iii) $\vdash_{\mathcal{Q}} x \leq \bar{n} \leftrightarrow x \approx \bar{0} \vee x \approx \bar{1} \vee \dots \vee x \approx \bar{n}$;
- (iv) $\vdash_{\mathcal{Q}} x \leq \bar{n} \vee \bar{n} \leq x$.

As is well known, all recursive functions and relations are definable in \mathcal{Q} , and in all its extensions⁽⁸⁾. The precise meaning of this statement will be formulated in the following terminology:

3.2. DEFINITION. Let \mathcal{A} be a K -system, where $K_0 \subseteq K$. Let R be an $(n+1)$ -ary relation (in ω) and φ a formula of K with $Fv(\varphi) = \{v_0, \dots, v_n\}$.

(i) φ is said to *numerate* R in \mathcal{A} if for any k_0, \dots, k_n

$$R[k_0, \dots, k_n] \equiv \vdash_{\mathcal{A}} \varphi(k_0, \dots, k_n);$$

(ii) φ is said to *bi-numerate* R in \mathcal{A} if φ numerates R in \mathcal{A} and $\sim\varphi$ numerates $\omega^n - R$ in \mathcal{A} ;

R is said to be *numerable* or *bi-numerable* in \mathcal{A} if there is a formula φ which numerates it or bi-numerates it in \mathcal{A} , respectively. Each formula φ of \mathcal{A} with free variables v_0, \dots, v_n is said to be a *numeration* of the relation R defined by the equivalence (i); φ is said to be a *bi-numeration* if, in addition it bi-numerates that R . The notion "bi-numerate" coincides with what has been expressed elsewhere by "define" (cf. [4], [24], [32]), "strongly define" or "strongly represent" (cf. [11]) and "numeralwise express" (cf. [15]), in the case that \mathcal{A} is a consistent theory. The use of numerate coincides with the use of "represent" (cf. [3], [4], [25]), "weakly represent", or "weakly define"⁽⁹⁾.

The notions introduced above will also be used to connect functions and formulas, by regarding each $(n+1)$ -ary function F as an $(n+2)$ -ary relation. For example, we will say that a formula φ with free variables v_0, \dots, v_n, v_{n+1} numerates F in \mathcal{A} if φ numerates F as a relation, i. e. if for each k_0, \dots, k_n, p

$$F(k_0, \dots, k_n) = p \equiv \vdash_{\mathcal{A}} \varphi(\bar{k}_0, \dots, \bar{k}_n, \bar{p}).$$

If φ has the form $\xi(v_0, \dots, v_n) \approx v_{n+1}$, this also implies that φ bi-numerates F in \mathcal{A} . Stronger notions of definability of functions have been found useful in the literature, but we shall not need them here.

⁽⁸⁾ It is also well known that a weaker theory than \mathcal{Q} (the theory R defined in [32], p. 52-53) also has these properties. However, the finiteness and simplicity of \mathcal{Q} give it a very useful role in our investigations.

⁽⁹⁾ We believe the new terminology, as against the weight of past usage, recommends itself on the grounds of uniformity, relative simplicity, and adaptability to a variety of means of expression ("numerate", "-ates", "-able", "-ation"); moreover, it immediately suggests the notions involved.

The construction of numerations of primitive recursive functions and relations can be carried out most directly in certain extensions of \mathcal{P} involving new logical constants, and new axioms which correspond to the defining equations for primitive recursive functions. This is presented in precise terms as follows.

3.3. DEFINITION. $\mathcal{P}' = \langle P', K \rangle$ is said to be a p. r. extension of \mathcal{P} if there is a sequence of distinct function symbols g_0, \dots, g_m , of $n_0 + 1, \dots, n_m + 1$ places, respectively, and a sequence of sentences ψ_0, \dots, ψ_m such that:

(i) For each $i \leq m$, there are $\xi, \eta \in \mathcal{Fm}_{K_i}$, where $K_i = K_0 \cup \{g_0, \dots, g_{i-1}\}$, for which $\mathcal{Fv}(\xi) \subseteq \{v_0, \dots, v_{n_i-1}\}$, $\mathcal{Fv}(\eta) \subseteq \{v_0, \dots, v_{n_i+1}\}$ and

$$\psi_i = \bigwedge_{v_0} \dots \bigwedge_{v_{n_i-1}} [\mathcal{G}_i[v_0, \dots, v_{n_i-1}, \bar{0}] \approx \xi(v_0, \dots, v_{n_i-1}) \\ \bigwedge_{v_{n_i}} \mathcal{G}_i[v_0, \dots, v_{n_i-1}, v_{n_i}] \approx \eta(v_0, \dots, v_{n_i}, \mathcal{G}_i[v_0, \dots, v_{n_i-1}, v_{n_i}])],$$

and

(ii) $K = K_n$ and P' consists of all sentences of P , together with ψ_0, \dots, ψ_n , as well as with all sentences which are generalizations of the following, for some $\varphi \in \mathcal{Fm}_K$:

$$\varphi(\bar{0}) \wedge \bigwedge_{\mathbf{x}} (\varphi(\mathbf{x}) \rightarrow \varphi(\mathbf{x}')) \rightarrow \bigwedge_{\mathbf{x}} \varphi(\mathbf{x}).$$

The use of terms ξ, η in the description of the ψ_i permits us to formally express in one step the construction of new functions from given ones by means of composition and by permutation and substitution of variables. Thus we see that the following holds.

3.4. THEOREM. (i) If F is an $(n+1)$ -ary primitive recursive function then there is a p. r. extension \mathcal{P}' of \mathcal{P} and a term ξ with $\mathcal{Fv}(\xi) = \{v_0, \dots, v_n\}$ such that the formula $\xi(v_0, \dots, v_n) \approx_{v_{n+1}}$ numerates F in \mathcal{P}' (if \mathcal{P}' is consistent); further, under these conditions, every such formula numerates a primitive recursive function.

(ii) If R is an $(n+1)$ -ary primitive recursive relation then there are \mathcal{P}' , ξ as in (i) such that the formula $\xi(v_0, \dots, v_n) \approx \bar{0}$ bi-numerates R in \mathcal{P}' (if \mathcal{P}' is consistent); further, under these conditions, every such formula bi-numerates a primitive recursive relation.

To obtain numerations of these functions and relations in \mathcal{P} , we must make use of the elimination procedure discovered by Gödel and presented in [10]. The details of this have been described elsewhere (cf. [15], p. 243-245 and [13], Vol. I, p. 401-419). What we need to know of it is as follows. With each p. r. extension \mathcal{P}' of \mathcal{P} and each formula φ of \mathcal{P}' is associated a formula $\varphi^{(\mathcal{P}')}$ of \mathcal{P} with exactly the same free variables. The association is such that $(\sim \varphi)^{(\mathcal{P}')} = \sim \varphi^{(\mathcal{P}')}$, $(\varphi \rightarrow \psi)^{(\mathcal{P}')} = \varphi^{(\mathcal{P}')} \rightarrow \psi^{(\mathcal{P}')}$, and

$(\bigwedge_u \varphi)^{(\mathcal{P}')} = \bigwedge_u \varphi^{(\mathcal{P}')}$ for all formulas φ, ψ and variables u . Also, for each formula φ of \mathcal{P} , $\varphi^{(\mathcal{P}')} = \varphi$.

Moreover, we have the following.

3.5. (i) If φ is a formula of \mathcal{P}' then $\vdash_{\mathcal{P}'} \varphi \leftrightarrow \varphi^{(\mathcal{P}')}$.

(ii) If φ is a sentence of \mathcal{P}' then

$$\vdash_{\mathcal{P}'} \varphi \text{ if and only if } \vdash_{\mathcal{P}} \varphi^{(\mathcal{P}')}.$$

(iii) $\mathcal{P}' \in \text{CON}$ if and only if $\mathcal{P} \in \text{CON}$.

Since the hypothesis that \mathcal{P} is consistent will be needed in a number of theorems in this paper, we will for simplicity now assume it outright. The strictly finitistically inclined reader can easily return it to its role as a hypothesis wherever necessary. On this count we can drop the conditions on \mathcal{P}' written in parentheses in 3.4.

Specific to K -theories for which $K_0 \subseteq K$ is the notion of ω -consistency. We say that \mathcal{A} is ω -consistent if there is no formula φ with $\mathcal{Fv}(\varphi) = \{x\}$ such that $\vdash_{\mathcal{A}} \bigvee_n \varphi(x)$ and $\vdash_{\mathcal{A}} \sim \varphi(\bar{n})$ for each n ; we say that \mathcal{A} is weakly ω -consistent if this is assumed to be true only for PR-formulas φ (cf. 3.6 (i)). We shall assume directly that \mathcal{Q} is weakly ω -consistent. This follows from known finitary consistency proofs of \mathcal{Q} and related systems (cf. [15], p. 470, Theorem 53)⁽¹⁰⁾. On the other hand, the hypothesis of ω -consistency of \mathcal{P} will be explicitly introduced whenever needed.

The elimination procedure and the result 3.4 suggests the introduction of two classes of formulas in \mathcal{P} .

3.6. DEFINITION. Let $\varphi \in \mathcal{Fm}_{K_0}$.

(i) φ is said to be a PR-formula if for some p. r. extension \mathcal{P}' of \mathcal{P} and some term ξ of \mathcal{P}' , $\varphi = (\xi \approx \bar{0})^{(\mathcal{P}')}$.

(ii) φ is said to be an RE-formula if for some PR-formula ψ and variables u_0, \dots, u_{k-1} , $\varphi = \bigvee_{u_0} \dots \bigvee_{u_{k-1}} \psi$.

The class of PR-formulas is thus seen to be a subclass of the class of RE-formulas. Furthermore, it is not hard to see by inspection of the elimination procedure that each of these classes is primitive recursive.

3.7. LEMMA. (i) If φ_1 is a PR-formula we can effectively find a PR-formula φ such that

$$\vdash_{\mathcal{P}} \varphi \leftrightarrow \sim \varphi_1.$$

(ii) If φ_1, φ_2 are PR-(RE-) formulas we can effectively find PR-(RE-) formulas φ and ψ such that

$$\vdash_{\mathcal{P}} \varphi \leftrightarrow \varphi_1 \vee \varphi_2, \quad \text{and} \quad \vdash_{\mathcal{P}} \psi \leftrightarrow \varphi_1 \wedge \varphi_2.$$

⁽¹⁰⁾ Kreisel has pointed out to us that in contrast, full ω -consistency of \mathcal{Q} finitistically implies the consistency of \mathcal{P} . Cf. his paper [16], especially p. 48-49.

(iii) If φ_1 is a PR-(RE-) formula, u, w are variables, $u \neq w$, then we can effectively find PR-(RE-) formulas φ and ψ such that

$$\vdash_{\mathcal{P}} \varphi \leftrightarrow \bigwedge_u (u < w \rightarrow \varphi_1), \quad \text{and} \quad \vdash_{\mathcal{P}} \psi \leftrightarrow \bigvee_u (u < w \wedge \varphi_1).$$

Closer inspection of the elimination procedure also suggests the introduction of the following classes of formulas.

3.8. DEFINITION. Let $\varphi \in \mathcal{Fm}_{K_0}$.

(i) φ is said to be an elementary formula, $\varphi \in \mathcal{EF}$, if it belongs to the least class containing each of the formulas $u \approx \bar{0}$, $u' \approx v$, $u + v \approx w$, $u \cdot v \approx w$, for u, v, w variables, and which is closed under \sim and \rightarrow ;

(ii) φ is said to be a bounded prenex formula, $\varphi \in \mathcal{BPF}$, if φ belongs to the least class containing \mathcal{EF} and which contains with every formula φ_1 also $\bigvee_u \varphi_1$ and $\bigwedge_u (u < w \rightarrow \varphi_1)$, for u, w variables.

Then it can be shown that every RE-formula is logically equivalent to a formula in bounded prenex form:

3.9. LEMMA. With every RE-formula φ we can effectively associate a $\psi \in \mathcal{BPF}$ such that

$$\vdash_{\mathcal{L}_{K_0}} \varphi \leftrightarrow \psi \quad (11).$$

An easy inductive proof, using the properties 3.1 of \mathcal{Q} then serves to show the following.

3.10. LEMMA. If $\psi \in \mathcal{BPF}$, $Fv(\psi) \subseteq \{v_0, \dots, v_{n-1}\}$ then for any k_0, \dots, k_{n-1}

$$\psi(\bar{k}_0, \dots, \bar{k}_{n-1}) \text{ is true implies } \vdash_{\mathcal{Q}} \psi(\bar{k}_0, \dots, \bar{k}_{n-1}) \quad (12).$$

We can now combine 3.4, 3.5, 3.7, 3.9 and 3.10 to reach the following conclusion.

3.11. THEOREM. (i) If φ is a PR-formula and $Fv(\varphi) = \{v_0, \dots, v_n\}$ then φ is a bi-numeration in \mathcal{Q} of an $(n+1)$ -ary primitive recursive relation R ; further to each such R corresponds a PR-formula φ which bi-numerates it in \mathcal{Q} ;

(ii) If φ is an RE-formula and $Fv(\varphi) = \{v_0, \dots, v_n\}$ then φ is a numeration in \mathcal{Q} of an $(n+1)$ -ary recursively enumerable relation R ; further to each such R corresponds an RE-formula φ which numerates it in \mathcal{Q} .

(iii) The same results (i), (ii) apply to any recursively axiomatizable consistent extension \mathcal{A} of \mathcal{Q} .

(*) If we loosen the requirement on provability here from \mathcal{L}_{K_0} to \mathcal{P} , then we can obtain ψ in somewhat more restricted form, according to the results of Davis [2] and Robinson [28]. However, the fact of logical provability will be quite useful in our work.

(**) This is closely related to Satz V of [10].

3.11 (i) is proved by considering the bounded prenex formulas φ_1, φ_2 associated with φ and $\sim\varphi$.

Some remarks on this theorem: To avoid the concept of truth used in 3.10, one must establish a restricted version of 3.5 (ii), namely: If φ is a sentence of \mathcal{P}' of the form $\xi(\bar{k}_0, \dots, \bar{k}_n) \approx \bar{p}$ and $\vdash_{\mathcal{P}} \varphi$ then $\vdash_{\mathcal{Q}} \varphi^{(\mathcal{P}'})$ (cf. [15], p. 244). In any case, 3.10 can be avoided if we only apply 3.11 to \mathcal{P} and its extensions; it is the arithmetization of 3.10 which will mainly interest us. Regarding part (iii) of the above theorem, observe that the same formula which bi-numerates a primitive recursive relation in \mathcal{Q} can be used to bi-numerate it in \mathcal{A} . On the other hand, it is by no means evident that (ii) extends to any but the (weakly) ω -consistent extensions of \mathcal{Q} . The weakening of this condition to ordinary consistency has only been made possible by some recent work [3]. However, it should be kept in mind that in general the same RE-formula will numerate different relations in \mathcal{A} and \mathcal{Q} unless \mathcal{A} is (weakly) ω -consistent.

We can also prove, using the representation of functions by terms in 3.4 (i), that every primitive recursive function is definable in the stronger sense (for functions, described earlier) in \mathcal{Q} . Hence, using Kleene's normal form theorem for general recursive functions (cf. [15], p. 288) we can obtain the following.

3.12. THEOREM. A relation R is recursive if and only if it is bi-numerable in \mathcal{Q} . The same holds true of any consistent recursively axiomatizable extension \mathcal{A} of \mathcal{Q} .

A simple diagonal argument shows that, in contrast with the classes of PR- and RE-formulas, the class of bi-numerations in \mathcal{A} is not even recursively enumerable, nor is any subclass of it which contains at least one bi-numeration of every recursive set.

4. The arithmetization of general metamathematics

We wish to describe in this section the construction of a particular p. r. extension \mathcal{M} of \mathcal{P} in which can be formalized various metamathematical results about arbitrary K -pseudo-systems, for any finite set K of constants. We shall not attempt to specify all the symbols and axioms of \mathcal{M} . Rather, our effort will be directed to showing how, by a simple notational device, we can grasp the mathematical content of \mathcal{M} .

Let us begin with a simple example. Consider the primitive recursive function $E(n, m) = n^m$. To this function corresponds in \mathcal{M} a certain two-placed function symbol g_0 , and the axiom:

$$\bigwedge_x (g_0[x, \bar{0}] \approx \bar{1} \wedge \bigwedge_y g_0[x, y'] \approx (g_0[x, y]) \cdot x).$$

We shall write

$$u^w \quad \text{for} \quad g_0[u, w],$$



so that we can equally well describe the axiom as being:

$$\bigwedge_x \{x^{\bar{0}} \approx \bar{1} \wedge \bigwedge_y x^{(y^x)} \approx (x^y) \cdot x\}.$$

Then we see that for any n, m , $\vdash_{\mathcal{M}} \bar{n}^{\bar{m}} \approx (\bar{n}^{\bar{m}})$. Consider now the primitive recursive function $I(n, m) = n \rightarrow m = 2^0 \cdot 3^n \cdot 5^m$. To this function corresponds in \mathcal{M} a certain term ξ with two free variables x, y , namely

$$((\bar{2}^{\bar{x}}) \cdot (\bar{3}^{\bar{y}})) \cdot (\bar{5}^{\bar{y}})$$

we agree to write

$$u \rightarrow w \quad \text{for} \quad \xi(u, w).$$

Thus we see that, for example, for any sentences φ, ψ ,

$$\vdash_{\mathcal{M}} (\overline{\varphi \rightarrow \psi}) \approx (\overline{\varphi} \rightarrow \overline{\psi}).$$

Similarly, we would introduce the “ $(u+1)$ -st variable” by writing

$$\overline{\forall} u \quad \text{for} \quad \bar{8} \cdot u + \bar{25}$$

and then the set of variables would be defined by the formula $\overline{\forall} r$ with one free variable x ,

$$\bigvee (x \approx \overline{\forall} r_u).$$

Another, slightly more complicated example arises when we try to formalize the notion of the set of terms of K , Tm_K . Take, for example, the case that $K = \{r_{1,0}, c_0, f_{1,0}\}$. We might first introduce the term $\overline{Fn}_{1,0}$ by writing

$$\overline{Fn}_{1,0} \quad \text{for} \quad (\bar{2}^{\bar{f}_{1,0}} \cdot \bar{3}^{\bar{x}}) \cdot \bar{5}^{\bar{y}}.$$

We will have a certain one-placed function symbol g_1 which will be used to represent the characteristic function of Tm_K . Then Tm_K will be a formula, with one free variable x , namely $g_1[x] \approx \bar{0}$. Further the following statement involving Tm_K should be provable in \mathcal{M} :

$$\vdash_{\mathcal{M}} Tm_K(x) \leftrightarrow \overline{\forall} r(x) \vee x \approx \bar{c}_0 \vee \bigvee_y \bigvee_z [y < x \wedge z < x \wedge Tm_K(y) \wedge Tm_K(z) \wedge x \approx \overline{Fn}_{1,0}(y, z)].$$

This corresponds directly to this statement:

$$\xi \in Tm_K \equiv \xi \text{ is a variable or } \xi = c_0 \text{ or}$$

$$\text{there exist } \eta, \zeta \in Tm_K \text{ such that } \xi = f_{1,0}[\eta, \zeta].$$

Of course, g_1 is not introduced directly by an ordinary recursive definition, but rather by a course-of-values recursion; but this, as we know, can be reduced to a series of ordinary primitive recursive definitions.

The general method should now be clear. With each $(n+1)$ -ary primitive recursive function is associated a certain $(n+1)$ -placed function symbol g of \mathcal{M} . With our ordinary mathematical notation for the function, say “ F ” is associated a metamathematical notation for the term $g[\overline{v}_0, \dots, \overline{v}_n]$, obtained by adding a dot, say “ \overline{F} ”. Then for any terms ξ_0, \dots, ξ_n of \mathcal{M} , according to our convention regarding substitution,

$$\overline{F}(\xi_0, \dots, \xi_n) \quad \text{is} \quad g[\xi_0, \dots, \xi_n].$$

(Of course, due to the variety of our mathematical notations, this procedure cannot be completely uniform.) Similarly, with each $(n+1)$ -ary primitive recursive relation we associate a certain $(n+1)$ -placed function symbol h in \mathcal{M} , to correspond to the characteristic function of the relation. Then with our mathematical notation for the relation, say R , we associate a metamathematical notation for the formula $h[\overline{v}_0, \dots, \overline{v}_n] \approx \bar{0}$, say “ \overline{R} ”. Then for any terms ξ_0, \dots, ξ_n , $R(\xi_0, \dots, \xi_n)$ “formally expresses” that the relation holds between the objects “designated” by the terms. Since we are dealing with a p.r. extension \mathcal{M} , we will also have available axioms or theorems corresponding to the defining equations or conditions of the functions and relations involved.

With this procedure in mind, we shall now describe a partial list of terms and formulas in \mathcal{M} . Some of these will be followed by their “mathematical interpretation”, if it is not already clear. These are: $u \dot{-} w$, u^w , $u|w$ (“ u divides w ”), p_u (“the $(u+1)$ -st prime”), $(u)_w$ (“the exponent of p_w in u ”), $Sq(u)$ (“ u represents a sequence”), $L(u)$ (“the length of the sequence represented by u ”), $\overline{\forall} r_u$ (“the $(u+1)$ -st variable”), $\overline{\forall} r(u)$ (“ u is a variable”), $\sim u$, $u \rightarrow w$, $u \wedge w$, $u \vee w$, $u \leftrightarrow w$, \bigwedge_w , \bigvee_w , $\overline{Fn}_{0,m}(u)$ (“the term formed from the $(m+1)$ -st 1-placed function symbol followed by u ”), $\overline{Fn}_{1,m}(u, w)$ (“the term formed from the $(m+1)$ -st 2-placed function symbol followed by u and then by w ”), \dots , $Tm_K(u)$, $Rl_{0,m}(u)$ (“the atomic formula formed from the $(m+1)$ -st 1-placed relation symbol followed by u ”), $Rl_{1,m}(u, w)$, \dots , $u \approx w (Rl_{1,0}(u, w))$, $At_K(u)$ (“ u is an atomic formula in K ”), $Fm_K(u)$ (“ u is a formula of K ”), $\overline{\forall} r(u, w)$ (“ u is a free variable of w ”), $St_K(u)$ (“ u is a sentence of K ”), $Gen(u, w)$ (“ u is obtained by generalization from w ”), $Sb_t^u w$ (“the substitution of t for u in w ”), $SSb(t, u, w)$ (“the simultaneous substitution of each term of the sequence t for the corresponding term of u in w ”)—also written

$$Sb \left(\begin{matrix} (u)_{\bar{0}} & \dots & (u)_{L(t)-\bar{1}} \\ (t)_{\bar{0}} & \dots & (t)_{L(t)-\bar{1}} \end{matrix} w \right),$$

and Ax_K (“the set of logical axioms of K ”). Examples of some formulas provable in \mathcal{M} are the following.

$$\begin{aligned} & \vdash_{\mathcal{M}} \forall x \forall y \forall z (\text{Fm}_K(y) \wedge \text{Tm}_K(z) \rightarrow \text{Ax}_K(\bigwedge_x y \rightarrow \text{Sb}_z^x y)); \\ & \vdash_{\mathcal{M}} \text{Ax}_K(\bigwedge_{\substack{v_1 \\ v_0}} \text{vr}_0 \approx \text{vr}_0); \\ & \vdash_{\mathcal{M}} \forall x \forall y \forall z (\text{Fm}_K(y) \wedge \text{Tm}_K(z) \rightarrow \text{Sb}_z^x(\sim y) \approx \sim \text{Sb}_z^x y); \\ & \vdash_{\mathcal{M}} \forall x \forall y \forall z (\text{Tm}_K(y) \wedge \text{Tm}_K(u) \wedge \text{Tm}_K(z) \rightarrow \text{Sb}_z^x(y \approx u) \approx (\text{Sb}_z^x y \approx \text{Sb}_z^x u), \end{aligned}$$

(where u is a variable distinct from x, y, z).

In addition to the symbols listed above we shall want three symbols specific to number theory: u^{\wedge} ($\text{Fn}_{0,0}(u)$), nm_u ("the $(u+1)$ -st numeral"), and $u \leq w$. Thus, for example,

$$\begin{aligned} & \vdash_{\mathcal{M}} \text{nm}_0 \approx c_0, \\ & \vdash_{\mathcal{M}} \text{nm}_{(x)} \approx (\text{nm}_x)^{\wedge}, \\ & \vdash_{\mathcal{M}} \text{St}_{K_0}(\text{nm}_x \leq \text{nm}_y). \end{aligned}$$

After a little work with such examples it becomes evident how one can express and prove in \mathcal{M} statements corresponding to every syntactic proposition (referred to an arbitrary fixed finite set of constants K) which has been proved mathematically within elementary number theory.

What is not as evident is how to express the logical notions associated with that of provability from an arbitrary pseudo-system $\mathcal{A} = \langle A, K \rangle$. However, if we look back at the definition of $\text{Prf}_{\mathcal{A}}$, we see that the only concept which enters in it which we may not be able to express in \mathcal{M} is that of membership in A . Even if we know that A is recursive, or primitive recursive, we are still faced with the further problem of choosing one from among the many numerations of A in \mathcal{P} . In order not to prejudice the investigations, we therefore take the following initially non-committal approach. Let a be a formula with one free variable x . We shall define a formula Prf_a with two free variables x, y which will express, when $a(x)$ is read as expressing that x belongs to A , that y is a proof from $\mathcal{A} = \langle A, K \rangle$ of x .

4.1. DEFINITION. Let a be a formula of \mathcal{M} , and let u, v, w be the first three variables not free in a and distinct from x, y, z . We take

$$\text{Prf}_a$$

to be the following formula of K_0 ,

$$\begin{aligned} & (\text{Sq}(y) \wedge \text{L}(y) \neq \bar{0} \wedge \bigwedge_u \{u < \text{L}(y) \rightarrow \text{Fm}_K((y).u) \\ & \wedge [\text{Ax}_K((y).u) \vee a((y).u)] \vee \bigvee_v \bigvee_w \{v < u \wedge w < u \\ & \wedge (y).v \approx (y).w \rightarrow (y).u\}) \wedge x \approx (y). \text{L}(y) - 1 \}^{(\mathcal{M})}. \end{aligned}$$

Properly speaking, $\text{Prf}_a(x, y)$ should be written $\text{Prf}_{\langle a, K \rangle}(x, y)$; however, it will always be clear which set K is intended.

We have allowed in this definition for the possibility that a may contain some variables other than x free. This will permit us to discuss provability from $\mathcal{A} \upharpoonright n$, for arbitrary n , by the following device:

4.2. DEFINITION. Let a be a formula of \mathcal{M} with $Fv(a) = \{x\}$ and a' the formula $a(x) \wedge x \leq z$. We put

$$\text{Prf}_{a' \upharpoonright z}$$

to be the formula

$$\text{Prf}_{a'}.$$

We shall be most interested, in the following, in dealing with Prf_a and $\text{Prf}_{a \upharpoonright z}$ in the case that $Fv(a) = \{x\}$; in this case we have $Fv(\text{Prf}_a) = \{x, y\}$ and $Fv(\text{Prf}_{a \upharpoonright z}) = \{x, y, z\}$. We shall often be able to deal with both cases by assuming that $Fv(a) \subseteq \{x, z\}$.

In case A is finite we see one clear way of defining A formally. This is expressed in our symbolism as follows.

4.2. DEFINITION. Let A be a finite set, $A = \{k_0, \dots, k_{n-1}\}$. Suppose that $k_0 < \dots < k_{n-1}$. By $[A]$ we mean the formula $x \neq x$, in case $n = 0$, and the formula

$$x \approx k_0 \vee \dots \vee x \approx k_{n-1},$$

if $n > 0$.

Thus $\text{Prf}_{[A]}(x, y)$ will express the proof relation with respect to the finite set A ; in particular $\text{Prf}_{[0]}(x, y)$ corresponds to logical provability. Clearly every formula $[A]$, for A finite, is logically equivalent to a bounded prenex formula, and by 3.7 (ii) is equivalent (in \mathcal{P}) to a PR -formula.

The formal expression of provability is now a simple matter.

4.3. DEFINITION. For any formula a of \mathcal{M} , Pr_a and $\text{Pr}_{a \upharpoonright z}$ are the formulas of K_0 defined as follows:

- (i) $\text{Pr}_a(x) = \bigvee_y \text{Prf}_a(x, y)$,
- (ii) $\text{Pr}_{a \upharpoonright z}(x) = \bigvee_y \text{Prf}_{a \upharpoonright z}(x, y)$.

That the formal definitions of proof and provability are extensionally correct is guaranteed by the following theorem.

4.4. THEOREM. Let $a \in \text{Fm}_{K_0}$, $Fv(a) = \{x\}$. Let $\mathcal{A} = \langle A, K \rangle$ be an arbitrary pseudo-system and $\mathcal{S} = \langle S, K' \rangle$ a theory with $\mathcal{Q} \subseteq \mathcal{S}$. Then we have the following.

- (i) If a numerates (bi-numerates) A in \mathcal{S} then Pr_a numerates (bi-numerates) $\text{Prf}_{\mathcal{A}}$ in \mathcal{S} .
- (ii) If a numerates A in \mathcal{S} and \mathcal{S} is ω -consistent then Pr_a numerates $\text{Pr}_{\mathcal{A}}$ in \mathcal{S} .

(iii) The results (i) and (ii) hold if we take instead of Pr_a and Pr_a the formulas $\text{Pr}_{a \uparrow z}$, $\text{Pr}_{a \uparrow y}$, respectively, and instead of $\text{Pr}_{\mathcal{M}}$ and $\text{Pr}_{\mathcal{M}}$ the relations $R[a, b, n]$ and $S[a, n]$ equivalent, respectively, to $\text{Pr}_{\mathcal{M} \uparrow n}[a, b]$ and $\text{Pr}_{\mathcal{M} \uparrow n}[a]$.

(i) is proved most simply by distributing the superscript (\mathcal{M}) in the definition of Pr_a in 4.1 over the connectives and quantifiers, and then examining each part. (ii) then follows immediately. Because of the assumption about ω -consistency in (ii), we shall generally attempt to limit our arguments to the use of (i) (or the corresponding part of (iii)). In one respect, more can be said in case a is a PR - or RE -formula; the next theorem is obtained from 3.7-3.9.

4.5. THEOREM. Let a be a formula of K_0 , $Fv(a) \subseteq \{x, z\}$.

(i) If a is a PR - (RE -) formula then we can effectively find a PR - (RE -) formula ψ such that

$$\vdash_{\mathcal{M}} \text{Pr}_a \leftrightarrow \psi$$

and an RE -formula θ such that

$$\vdash_{\mathcal{M}} \text{Pr}_a \leftrightarrow \theta.$$

(ii) If β is a formula in bounded prenex form, $\beta \in BPF$, for which

$$\vdash_{\mathcal{L}_{K_0}} \alpha \leftrightarrow \beta$$

then we can effectively find a $\psi \in BPF$ such that

$$\vdash_{\mathcal{L}_{K_0}} \text{Pr}_a \leftrightarrow \psi.$$

The second part of this theorem, though slightly less informative, is occasionally more useful, since it gives us logical provability.

Let us turn now to the intensional questions. It is immediately seen that to any metamathematical statement of the form "for any pseudo-system \mathcal{A}, \dots " should correspond a class of sentences of \mathcal{M} dependent upon a choice of formula a —or, as we might put it more briefly, a sentence schema in a . Moreover, we should expect that for the elementary propositions about provability, each instance of such a schema should be provable in \mathcal{M} . Thus, for example, the basic properties of provability obtain the following arithmetization.

4.6. THEOREM. Let a be a formula of \mathcal{M} , $Fv(a) \subseteq \{x, z\}$.

$$(i) \vdash_{\mathcal{M}} \text{Pr}_a(x) \rightarrow \text{Fm}_{\mathcal{K}}(x).$$

$$(ii) \vdash_{\mathcal{M}} \text{Ax}_{\mathcal{K}}(x) \rightarrow \text{Pr}_a(x).$$

$$(iii) \vdash_{\mathcal{M}} a(x) \wedge \text{Fm}_{\mathcal{K}}(x) \rightarrow \text{Pr}_a(x).$$

$$(iv) \vdash_{\mathcal{M}} \text{Pr}_a(x) \wedge \text{Pr}_a(x \rightarrow y) \rightarrow \text{Pr}_a(y).$$

(v) If φ is any formula of \mathcal{M} , $y \in Fv(\varphi)$ then

$$\vdash_{\mathcal{M}} \bigwedge_x [(\text{Ax}_{\mathcal{K}}(x) \rightarrow \varphi(x)) \wedge (a(x) \wedge \text{Fm}_{\mathcal{K}}(x) \rightarrow \varphi(x))]$$

$$\wedge \bigwedge_x \bigwedge_y [\text{Fm}_{\mathcal{K}}(x) \wedge \text{Fm}_{\mathcal{K}}(y) \wedge \varphi(x) \wedge \varphi(x \rightarrow y) \rightarrow \varphi(y)] \rightarrow \bigwedge_y (\text{Pr}_a(x) \rightarrow \varphi(y)).$$

The first two derivability conditions of Hilbert-Bernays [14], vol. II, p. 285-286 are easily seen to follow from 4.4, 4.6 for any choice of a . However, we would gain no advantage in our work by stating these two special conditions. Continuing along these lines we obtain:

4.7. THEOREM. Let a, β be formulas of \mathcal{M} , $Fv(a) \cup Fv(\beta) \subseteq \{x, z\}$.

$$(i) \vdash_{\mathcal{M}} \bigwedge_x (\beta(x) \wedge \text{Fm}_{\mathcal{K}}(x) \rightarrow a(x)) \rightarrow \bigwedge_x (\text{Pr}_{\beta}(x) \rightarrow \text{Pr}_a(x)).$$

$$(ii) \vdash_{\mathcal{M}} \text{Pr}_{\text{Pr}_a}(x) \leftrightarrow \text{Pr}_a(x).$$

$$(iii) \text{ If } Fv(a) = \{x\}, \vdash_{\mathcal{M}} \text{Pr}_a(x) \leftrightarrow \bigvee_z \text{Pr}_{a \uparrow z}(x).$$

$$(iv) \vdash_{\mathcal{M}} \text{Fm}_{\mathcal{K}}(x) \wedge \forall u (u \wedge \bigwedge_y (a(y) \wedge \text{Fm}_{\mathcal{K}}(y) \rightarrow \sim \text{Fv}(u, y)) \rightarrow$$

$$[\text{Pr}_a(\bigwedge_u x) \leftrightarrow \text{Pr}_a(x)], \quad (u \text{ is a variable } \neq x, y, z).$$

$$(v) \vdash_{\mathcal{M}} \text{Pr}_{\{0\}}(x) \rightarrow \text{Pr}_a(x).$$

We see that 4.7 (iii) corresponds to the statement 2.3 and 4.7 (iv) to 2.1. Also the deduction theorem 2.2 can be arithmetized; for example, for the case $k = 0$, this can be stated as follows.

4.8. THEOREM. Let a be a formula of \mathcal{M} , $Fv(a) \subseteq \{x, z\}$.

(i) Let $\beta = a \vee (x \approx u)$, where u is a variable distinct from x, y, z . Then $\vdash_{\mathcal{M}} \text{Fm}_{\mathcal{K}}(y) \rightarrow (\text{Pr}_{\beta}(x) \leftrightarrow \text{Pr}_a(u \rightarrow x))$.

(ii) Let $\beta' = a \vee (x \approx \bar{q}_0)$, for any particular $q_0 \in \text{Fm}_{\mathcal{K}}$. Then $\vdash_{\mathcal{M}} \text{Pr}_{\beta'}(x) \leftrightarrow \text{Pr}_a(\bar{q}_0 \rightarrow x)$.

We are now in a position to discuss the arithmetized versions of the notions of closure, consistency, completeness and quantificational completeness.

4.9. DEFINITION. Let a be a formula of \mathcal{M} , $Fv(a) \subseteq \{x, z\}$. We set:

$$(i) \text{Cl}_a = \bigwedge_x (\text{Pr}_a(x) \leftrightarrow a(x))^{(\mathcal{M})};$$

$$(ii) \text{Con}_a = \bigwedge_x [\text{Fm}_{\mathcal{K}}(x) \rightarrow \sim \text{Pr}_a(x) \vee \sim \text{Pr}_a(\sim x)]^{(\mathcal{M})};$$

$$(iii) \text{Cm}_a = \text{Cl}_a \wedge \text{Con}_a \wedge \bigwedge_x [\text{Fm}_{\mathcal{K}}(x) \rightarrow \text{Pr}_a(x) \vee \text{Pr}_a(\sim x)]^{(\mathcal{M})};$$

$$(iv) \text{QCm}_a = \text{Cm}_a \wedge \bigwedge_x \bigwedge_u \{ \text{Fm}_{\mathcal{K}}(x) \wedge \forall u (u) \rightarrow \\ \rightarrow [\text{Pr}_a(\bigwedge_u x) \leftrightarrow \bigwedge_w (\forall u (u) \rightarrow \text{Pr}_a(\text{Sb}_w^u x))] \}^{(\mathcal{M})},$$

(where u, w are distinct variables $\neq x, z$).

We see that if $Fv(a) = \{x\}$ then Cl_a , Con_a , Cm_a , QCm_a are sentences in K_0 , and that if $Fv(a) \subseteq \{x, z\}$ they are formulas in K_0 with at most z free.

One easily obtains the following as a formalization of simple statements about consistency. It is sufficient to our purposes, from now on, to consider only the case $Fv(a) = \{x\}$.

4.10. THEOREM. *Suppose that a, β are formulas of \mathcal{M} , $Fv(a) = Fv(\beta) = \{x\}$.*

(i) *For any particular $\varphi_0 \in Fm_K$,*

$$\vdash_{\mathcal{M}} Con_a \leftrightarrow \sim Pr_a(\bar{\varphi}_0 \wedge \sim \bar{\varphi}_0).$$

(ii) $\vdash_{\mathcal{M}} Con_a \leftrightarrow \bigwedge_z Con_{a \uparrow z}$.

(iii) $\vdash_{\mathcal{M}} \bigwedge_x (\beta(x) \wedge Fm_K(x) \rightarrow \alpha(x)) \rightarrow (Con_a \rightarrow Con_\beta)$.

The arithmetization of the extension Theorems 2.7 and 2.8 is less immediate, but also not very difficult. It is possible, as in such cases, to formalize a statement that for each set A of a certain type there exists another set B of a certain type whenever we are able to show that in the cases under consideration, the set B exhibited is (uniformly) arithmetically definable in terms of A .

4.11. THEOREM. *Let a be a formula of \mathcal{M} , $Fv(a) = \{x\}$. We can effectively construct a formula β of \mathcal{M} , $Fv(\beta) = \{x\}$, such that*

$$\vdash_{\mathcal{M}} Con_a \rightarrow \bigwedge_x (\alpha(x) \wedge Fm_K(x) \rightarrow \beta(x)) \wedge Cm_\beta.$$

The construction of β here can be formally described as follows. Let $Cnj(y, u)$ correspond in \mathcal{M} to the primitive recursive function $Cnj(\Phi, i) = (\Phi)_0 \wedge \dots \wedge (\Phi)_i$. Then we set β equal to

$$\begin{aligned} & \bigvee_y \{Sq(y) \wedge L(y) \neq \bar{0} \wedge \bigwedge_t [t < L(y) - \bar{1} \rightarrow ((y)_t < (y)_{t-1}) \wedge Fm_K((y)_t)] \\ & \wedge \sim Pr_a(\sim Cnj(y, L(y) - \bar{1})) \wedge \bigwedge_z [Fm_K(z) \wedge z < (y)_{\bar{0}} \rightarrow Pr_a(\sim z)] \\ & \wedge \bigwedge_t \bigwedge_z [t < L(y) - \bar{1} \wedge Fm_K(z) \wedge (y)_t < z \wedge z < (y)_{t-1} \rightarrow \\ & \rightarrow Pr_a(Cnj(y, t) \rightarrow \sim z)] \wedge x \approx (y)_{L(y) - \bar{1}} \}. \end{aligned}$$

Similarly, by following the proof of 2.8 we obtain:

4.12. THEOREM. *Let a be a formula of \mathcal{M} , $Fv(a) = \{x\}$. We can effectively construct a formula γ of \mathcal{M} , $Fv(\gamma) = \{x\}$, such that*

$$\vdash_{\mathcal{M}} Con_a \rightarrow \bigwedge_x (\alpha(x) \wedge St_K(x) \rightarrow \gamma(x)) \wedge QCm_\gamma. \quad (13)$$

(13) It is hard to see how the Theorems 4.11 and 4.12 could be simply formulated in the previous treatments of the problems of arithmetization. These already indicate the simplicity and flexibility which can be obtained by using formulas Con_a , Cm_a , etc.

Note that the first part of the conclusion here corresponds to the hypothesis $A \subseteq St_K$ of 2.8.

We wish to conclude this section with two special observations which are needed upon occasion in the following.

Though many times a result about formulas Pr_a can be obtained when a is an RE -formula, thus including the case when a is a PR -formula, it is occasionally necessary to carry through a proof under the more restrictive conditions. This is made possible by a device used by Craig in [1], which we modify slightly in the following.

4.13. THEOREM. *Let $\mathcal{A} = \langle A, K \rangle$ be an axiom system with A recursively enumerable. If a is an RE -formula which numerates A in \mathcal{Q} then we can construct a PR -formula a_1 and an axiom system $\mathcal{A}_1 = \langle A_1, K \rangle$ such that:*

(i) a_1 bi-numerates A_1 in \mathcal{Q} ;

(ii) $Pr_{\mathcal{A}} = Pr_{\mathcal{A}_1}$;

(iii) $\vdash_{\mathcal{A}} Pr_a(x) \leftrightarrow Pr_{a_1}(x)$;

(iv) *if \mathcal{S} is any ω -consistent axiomatic theory with $\mathcal{P} \subseteq \mathcal{S}$ and a_1 any (RE -) formula which numerates A_1 in \mathcal{S} then we can find an (RE -) formula a' which numerates A in \mathcal{S} for which*

$$\vdash_{\mathcal{S}} Pr_{a'}(x) \leftrightarrow Pr_{a_1}(x).$$

Proof. It should be recalled that when \mathcal{S} is assumed ω -consistent, an RE -formula a numerates A in \mathcal{S} just in case it numerates it in \mathcal{Q} . The only effect of the use of \mathcal{S} in (iv) is that we can consider arbitrary numerations α'_i, α' which are not necessarily RE -formulas. It is sufficient to consider $\alpha(x) = \bigvee_y \theta(x, y)$ where θ is a PR -formula, and let R be the relation which θ numerates in \mathcal{Q} . By 3.7 we can find a PR -formula α_1 such that

$$\vdash_{\mathcal{M}} \alpha_1(x) \leftrightarrow \bigvee_y \bigvee_z [z < x \wedge y < x \wedge \theta(z, y) \wedge x \approx z \wedge \bigwedge_{v_m} \forall y \approx \forall y].$$

Let A_1 be the primitive recursive set which α_1 numerates in \mathcal{Q} (the same as in \mathcal{M} , by the consistency of \mathcal{M}). Then a sentence φ of K belongs to A_1 if and only if $\varphi = \psi \wedge \bigwedge_{v_m} (v_m \approx v_m)$ for some ψ, m where $R[\psi, m]$.

It follows that if $\varphi \in A_1$ then $\varphi = \psi \wedge \bigwedge_{v_m} (v_m \approx v_m)$ for some $\psi \in A$; conversely (since \mathcal{Q} is weakly ω -consistent), with each $\psi \in A$ is associated an m such that $\psi \wedge \bigwedge_{v_m} (v_m \approx v_m) \in A_1$. But $\vdash_{\mathcal{L}_K} \bigwedge_{v_m} v_m \approx v_m$, so that $Pr_{\mathcal{A}} = Pr_{\mathcal{A}_1}$. Since

$$\vdash_{\mathcal{M}} Pr_{\mathcal{Q}}(\bigwedge_{v_y} \forall y \approx \forall y),$$

it is seen that

$$\vdash_{\mathcal{M}} a(z) \rightarrow \text{Pr}_{a_1}(z) \quad \text{and} \quad \vdash_{\mathcal{M}} a_1(x) \rightarrow \text{Pr}_{a_1}(x).$$

Hence by 4.7 (i) we obtain (iii). Also, if a'_1 is any numeration of A_1 in \mathcal{S} , when we set

$$a'(x) = \bigvee_y a'_1(x) \wedge \bigwedge_{\forall y} \text{Pr}_y \approx \text{Pr}_y^{(SM)},$$

it follows by ω -consistency of \mathcal{S} that a' numerates A in \mathcal{S} and, by the same argument as just given,

$$\vdash_{\mathcal{P}} \text{Pr}_a(x) \leftrightarrow \text{Pr}_{a'_1}(x)$$

a' is equivalent in \mathcal{P} to an *RE*-formula whenever a'_1 is an *RE*-formula.

A particular case in which *RE*-formulas cannot be used in the same way as *PR*-formulas is in connection with the formulas $\text{Prf}_{a \uparrow n}$. Even though $\text{Prf}_{a \uparrow n}$ is a numeration of $\text{Prf}_{\mathcal{A} \uparrow n}$ in a theory where a numerates A , we do not necessarily have $\text{Prf}_{a \uparrow n}(x, y)$ provably equivalent to $\text{Prf}_{[A \uparrow n]}(x, y)$ for any n . The reason for this is that we may not be able to show $a(x) \wedge x \leq \bar{n}$ provably equivalent to $[A \uparrow n](x)$; for if this were true for every n , a would be a bi-numeration of A . In fact, it is just for bi-numerations that we can obtain this result. The proof is clear.

4.14. LEMMA. Let $\mathcal{A} = \langle A, K \rangle$ and \mathcal{S} be axiom systems with $\mathcal{P} \subseteq \mathcal{S}$. Suppose a bi-numerates A in \mathcal{S} . Then for any n ,

$$\vdash_{\mathcal{S}} a(x) \wedge x \leq \bar{n} \leftrightarrow [A \uparrow n](x)$$

and hence

$$\vdash_{\mathcal{S}} \text{Prf}_{a \uparrow n}(x, y) \leftrightarrow \text{Prf}_{[A \uparrow n]}(x, y) \quad \text{and} \quad \vdash_{\mathcal{S}} \text{Con}_{a \uparrow n} \leftrightarrow \text{Con}_{[A \uparrow n]}.$$

5. The Gödel underivability theorems in a general setting

Up to now we have dealt with the arithmetization of the syntactical and logical notions relative to an arbitrary set of constants K . We now restrict attention to the case that $K_0 \subseteq K$. Moreover we shall deal primarily with theories \mathcal{A} which are extensions of \mathcal{Q} or of \mathcal{P} , i. e. with theories which are "suitably strong".

As we remarked in § 1, the first Gödel underivability theorem, about the incompleteness of certain theories, has a purely extensional formulation and has been proved in full generality. We shall, however, consider such a proof, since it will inform us, to a certain extent, about the proper formulation of the second underivability theorem.

Basic to these proofs is the construction, by a diagonal argument, of "self-referential" statements. This can be formulated as follows.

5.1. LEMMA. Let $\varphi \in \text{Fm}_{K_0}$, $\text{Fv}(\varphi) \subseteq \{x\}$. Then we can effectively find $\varphi \in \text{Fm}_{K_0}$ such that

$$\vdash_{\mathcal{Q}} \varphi \leftrightarrow \psi(\bar{\varphi}).$$

This is proved by taking χ to be the formula, with free variable x ,

$$\chi = \bigvee_y [(y \approx \text{Sb}_{\text{mm}_x}^{\text{FV}_0} x) \wedge \psi(y)].$$

Then we set $\varphi = \chi(\bar{\chi}) = \text{Sb}_{\bar{\chi}}^{\text{FV}_0} \chi$ ⁽¹⁴⁾.

5.2. DEFINITION. For each $a \in \text{Fm}_{K_0}$, with $\text{Fv}(a) \subseteq \{x\}$ we take v_a to be the sentence associated with $\psi = \sim \text{Pr}_a$ in 5.1, such that

$$\vdash_{\mathcal{Q}} v_a \leftrightarrow \sim \text{Pr}_a(\bar{v}_a).$$

One half of the first underivability theorem can now be stated as follows (not in its most general form, but sufficient for our purposes).

5.3. THEOREM. Suppose that $\mathcal{A} = \langle A, K \rangle$, $A \subseteq \text{St}_K$, $\mathcal{Q} \subseteq \mathcal{A}$ and that \mathcal{A} is consistent. Suppose that a numerates A in \mathcal{S} , where $\mathcal{Q} \subseteq \mathcal{S} \subseteq \mathcal{A}$. Then

$$\text{not } \vdash_{\mathcal{A}} v_a.$$

This follows immediately from the fact that Prf_a numerates $\text{Prf}_{\mathcal{A}}$ in \mathcal{S} (4.4 (i)). Hence, if for some Ψ , $\text{Prf}_{\mathcal{A}}[v_a, \Psi]$, we would have $\vdash_{\mathcal{A}} \text{Prf}_a(\bar{v}_a, \bar{\Psi})$, and thence $\vdash_{\mathcal{A}} \text{Pr}_a(\bar{v}_a)$, i. e. $\vdash_{\mathcal{A}} \sim v_a$.

To obtain a result about non-derivability in \mathcal{A} of $\sim v_a$, one would have to restrict \mathcal{A} to being ω -consistent (cf. [10], Satz VI). However, if one is interested in such incompleteness results, the use of Rosser's method instead serves to remove this restriction (cf. [29] Theorem II or [15], p. 208-209).

Since there exists at least one a which numerates A in \mathcal{Q} if A is recursively enumerable, it follows from 5.3 that there exists such a sentence v_a which is not provable in \mathcal{A} , for consistent recursively enumerable extensions \mathcal{A} of \mathcal{Q} . For our considerations, it is a bit more informative to look at the proof of 5.3 in the case that a is (logically) equivalent to a bounded prenex formula. Then we know by 4.5 (ii) that so also is Prf_a . Hence, by 3.10, if $\text{Prf}_a(\bar{v}_a, \bar{\Psi})$ is true then $\vdash_{\mathcal{Q}} \text{Prf}_a(\bar{v}_a, \bar{\Psi})$. The formalization of 3.10 is essential to showing that, under this restriction on a , the Theorem 5.3 can itself be formalized.

5.4. THEOREM. Suppose that $\varphi \in \text{BPF}$, $\text{Fv}(\varphi) \subseteq \{v_0, \dots, v_{n-1}\}$. Let t, u be distinct variables other than v_0, \dots, v_{n-1} . Then

$$\vdash_{\mathcal{M}} \text{Sq}(t) \wedge \text{Sq}(u) \wedge \text{L}(t) \approx \text{L}(u) \wedge \text{L}(t) \approx \bar{n} \wedge (t)_{\bar{v}_0} \approx \text{nm}_{v_0} \wedge \dots \wedge (t)_{\bar{v}_{n-1}} \approx \text{nm}_{v_{n-1}} \\ \wedge (u)_{\bar{v}_0} \approx \text{Pr}_{\bar{v}_0} \wedge \dots \wedge (u)_{\bar{v}_{n-1}} \approx \text{Pr}_{\bar{v}_{n-1}} \wedge \psi(v_0, \dots, v_{n-1}) \rightarrow \text{Pr}_{\mathcal{Q}}(\text{SSb}(t, u, \bar{\varphi})).$$

⁽¹⁴⁾ The construction of φ in this lemma is an obvious generalization of Gödel's diagonal construction in his proof of Satz VI of [10].

As suggested in § 4, we can also indicate this by writing

$$\vdash_{\mathcal{M}} \psi(v_0, \dots, v_{n-1}) \rightarrow \text{Pr}_{(Q)} \left(\text{Sb} \left(\frac{\text{vR}_0}{\text{nm}_{v_0}} \dots \frac{\text{vR}_{n-1}}{\text{nm}_{v_{n-1}}} \bar{\psi} \right) \right).$$

A proof of 5.4 would follow very directly the mathematical proof of 3.10.

5.5. COROLLARY. *Suppose that $\varphi \in \text{St}_{K_0}$ is such that for some $\psi \in \text{BPF}$,*

$$\vdash_Q \varphi \leftrightarrow \psi.$$

Then

$$\vdash_{\mathcal{P}} \varphi \rightarrow \text{Pr}_{(Q)}(\bar{\varphi}).$$

This follows from 5.4 and the fact that $\vdash_{\mathcal{M}} \text{Pr}_{(Q)}(\bar{\varphi} \leftrightarrow \bar{\psi})$, hence $\vdash_{\mathcal{P}} \text{Pr}_{(Q)}(\bar{\varphi}) \leftrightarrow \text{Pr}_{(Q)}(\bar{\psi})$.

In most applications we shall want to show that

$$\vdash_{\mathcal{M}} \varphi \rightarrow \text{Pr}_a(\bar{\varphi}),$$

for a certain formula a ; this will be possible whenever we can show $\vdash_{\mathcal{M}} [Q](x) \rightarrow \text{Pr}_a(x)$, by 4.7 (i) ⁽¹⁵⁾.

5.6. THEOREM. *Suppose that $\mathcal{A} = \langle A, K \rangle$ is a consistent axiom system with $\mathcal{P} \subseteq \mathcal{A}$. Suppose that a is an RE-formula which numerates A in \mathcal{S} , where $\mathcal{Q} \subseteq \mathcal{S} \subseteq \mathcal{A}$. Then*

$$\vdash_{\mathcal{A}} \text{Con}_a \leftrightarrow v_a$$

and hence

$$\text{not } \vdash_{\mathcal{A}} \text{Con}_a.$$

Proof. First observe that $\sim v_a$ is equivalent in Q to $\text{Pr}_a(\bar{v}_a)$ and hence is equivalent by 3.9 and 4.5 (ii) to a bounded prenex formula. Then, by 5.5,

$$\vdash_{\mathcal{P}} \sim v_a \rightarrow \text{Pr}_{(Q)}(\sim \bar{v}_a).$$

Since each $\varphi \in Q$ is provable in \mathcal{A} and Q is finite, it follows that

$$\vdash_{\mathcal{S}} [Q](x) \rightarrow \text{Pr}_a(x).$$

Hence by 4.7 (i), (ii)

$$\vdash_{\mathcal{A}} \sim v_a \rightarrow \text{Pr}_a(\sim \bar{v}_a).$$

This gives

$$\vdash_{\mathcal{A}} \text{Con}_a \wedge \sim v_a \rightarrow \sim \text{Pr}_a(\bar{v}_a),$$

i. e.

$$\vdash_{\mathcal{A}} \text{Con}_a \wedge \sim v_a \rightarrow v_a,$$

⁽¹⁵⁾ 5.4 and 5.5 take the place in our work of the third derivability condition of Hilbert-Bernays [14], Vol. II, p. 286; for a proof of that condition in $Z\mu$, cf. loc. cit. p. 309-313. The remark of the text shows that the condition applies to every proof definition Pr_a for which a "provably contains" Q .

which shows that

$$\vdash_{\mathcal{A}} \text{Con}_a \rightarrow v_a.$$

(Actually, it can be seen that $\vdash_{\mathcal{S} \cup \mathcal{P}} \text{Con}_a \rightarrow v_a$.) On the other hand, clearly $\vdash_{\mathcal{P}} \sim \text{Pr}_a(\bar{v}_a) \rightarrow \text{Con}_a$, so

$$\vdash_{\mathcal{P}} v_a \rightarrow \text{Con}_a.$$

The desired conclusion is then obtained by 5.3.

Thus we see how to state and prove Gödel's second undervability theorem under quite general conditions. Moreover, it is easily seen that this theorem comprehends all known particular instances as special cases. For example, the formulation of the axioms of \mathcal{P} as a finite set of axioms and axiom schemata quickly leads to a PR -formula π which bi-numerates P in \mathcal{Q} ⁽¹⁶⁾.

The theorem could further be strengthened by showing that 5.4, 5.5 and all the statements in the proof of 5.6 involving provability in \mathcal{P} can be obtained for a certain finitely axiomatizable subtheory of \mathcal{P} (independent of \mathcal{A}), and hence that there is a certain finite subtheory \mathcal{F} of $\mathcal{S} \cup \mathcal{P}$ which can be used in place of \mathcal{A} in 5.6.

The main feature of 5.6 which we wish to bring attention to is that, in contrast to 5.3, it is not stated for arbitrary numerations a of A in \mathcal{Q} , let alone of A in any subsystem of \mathcal{A} . Indeed, our next main step will be to show that under certain circumstances, it is not possible to obtain such improvements. In order to see why this is so, we must first introduce a new concept.

5.7. DEFINITION. *Let $\mathcal{A} = \langle A, K \rangle$, $K_0 \subseteq K$. \mathcal{A} is said to be reflexive if for each finite $F \subseteq A$,*

$$\vdash_{\mathcal{A}} \text{Con}_{(F)}.$$

It is readily seen that \mathcal{A} is reflexive just in case

$$\text{for each } n, \quad \vdash_{\mathcal{A}} \text{Con}_{[A \uparrow n]}.$$

The following theorem has been proved in [23] by Mostowski.

5.8 (i) \mathcal{P} is reflexive;

more generally,

5.8 (ii) *Every consistent extension \mathcal{A} of \mathcal{P} , with the same constants as \mathcal{P} , is reflexive.*

⁽¹⁶⁾ By begging the question as to exactly what an axiom schema is and what a "natural" formal definition of such is, we can say that for any system A of finitely many axioms and axiom schemata, the natural definition a of A is equivalent to a PR -formula a' which bi-numerates A in \mathcal{Q} .

This result has also been proved by Kreisel and Wang in [19]. It follows immediately from 5.6 that *no consistent reflexive theory is finitely axiomatizable*; indeed this was the main motive of Mostowski's work. We shall obtain a stronger result in 6.8 (ii). Formulation of the notion of reflexivity for a wider class of theories than in 5.7 will be given in § 8, together with a description of another class of reflexive theories discovered by Montague. At any rate, it is already seen that the notion applies to a wide class of theories.

5.9. THEOREM. *Suppose that $\mathcal{A} = \langle A, K \rangle$ is a consistent, reflexive axiom system with $\mathcal{P} \subseteq \mathcal{A}$. Suppose further that A is recursive. Then there is an α^* which bi-numerates A in \mathcal{A} for which*

$$\vdash_{\mathcal{P}} \text{Con}_{\alpha^*} \text{ }^{(17)}$$

Proof. By 3.12 there is an α which bi-numerates A in \mathcal{Q} , and hence in \mathcal{A} . Set α^* to be the formula, with one free variable x ,

$$(1) \quad \alpha^*(x) = \alpha(x) \wedge \bigwedge_z (z \leq x \rightarrow \text{Con}_{\alpha \uparrow z} \wedge \text{St}_K^{(\mathcal{Q})}(x)).$$

Suppose $n \in A$. Then

$$\vdash_{\mathcal{A}} \alpha(\bar{n}) \wedge \text{St}_K^{(\mathcal{Q})}(\bar{n})$$

and, by 4.14 and the reflexive property of \mathcal{A} ,

$$\vdash_{\mathcal{A}} \text{Con}_{\alpha \uparrow \bar{0}} \wedge \dots \wedge \text{Con}_{\alpha \uparrow \bar{n}};$$

hence

$$\vdash_{\mathcal{A}} \alpha^*(\bar{n}).$$

On the other hand, it is clear that

$$\text{if } \vdash_{\mathcal{A}} \sim \alpha(\bar{n}) \text{ then } \vdash_{\mathcal{A}} \sim \alpha^*(\bar{n}).$$

Thus α^* is also a bi-numeration of A in \mathcal{A} . To obtain the desired result, $\vdash_{\mathcal{P}} \text{Con}_{\alpha^*}$, we can work in \mathcal{M} as well as \mathcal{P} . First, we use the fact, established in 4.10 (ii), that

$$(2) \quad \vdash_{\mathcal{M}} \text{Con}_{\alpha} \leftrightarrow \bigwedge_z \text{Con}_{\alpha \uparrow z}.$$

Hence

$$(3) \quad \vdash_{\mathcal{M}} \sim \text{Con}_{\alpha} \rightarrow \bigvee_z [\sim \text{Con}_{\alpha \uparrow z} \wedge \bigwedge_y (y < z \rightarrow \text{Con}_{\alpha \uparrow y})].$$

Now, by 5.8 (i), $\vdash_{\mathcal{M}} \text{Con}_{\{0\}}$, so we see that $\vdash_{\mathcal{M}} \text{Con}_{\alpha \uparrow \bar{0}}$.

⁽¹⁷⁾ The corresponding Theorem 3 of [5] was incorrectly stated, as was Theorem 3 of [6] (we can only say that α^* bi-numerates A in \mathcal{A} , not in \mathcal{P} —however, the theorems agree with 5.9 when $\mathcal{A} = \mathcal{P}$).

Thus we can modify (3) to

$$(4) \quad \vdash_{\mathcal{M}} \sim \text{Con}_{\alpha} \rightarrow \bigvee_z [\sim \text{Con}_{\alpha \uparrow z} \wedge \bigwedge_y (y \leq z \rightarrow \text{Con}_{\alpha \uparrow y})].$$

But then it is seen from (1) that

$$(5) \quad \vdash_{\mathcal{M}} \sim \text{Con}_{\alpha} \rightarrow \bigvee_z [\text{Con}_{\alpha \uparrow z} \wedge \bigwedge_x (\alpha^*(x) \leftrightarrow \alpha(x) \wedge x \leq z)],$$

so that

$$(6) \quad \vdash_{\mathcal{M}} \sim \text{Con}_{\alpha} \rightarrow \bigvee_z [\text{Con}_{\alpha \uparrow z} \wedge \bigwedge_x (\text{Pr}_{\alpha^*}(x) \leftrightarrow \text{Pr}_{\alpha \uparrow z}(x))].$$

But then

$$(7) \quad \vdash_{\mathcal{M}} \sim \text{Con}_{\alpha} \rightarrow \text{Con}_{\alpha^*}.$$

On the other hand, it follows from (1) that

$$(8) \quad \vdash_{\mathcal{M}} \bigwedge_x (\alpha^*(x) \wedge \text{Fm}_K(x) \rightarrow \alpha(x)),$$

hence by 4.10 (iii),

$$(9) \quad \vdash_{\mathcal{M}} \text{Con}_{\alpha} \rightarrow \text{Con}_{\alpha^*}.$$

Thus

$$(10) \quad \vdash_{\mathcal{M}} \text{Con}_{\alpha^*},$$

which is the desired result by 3.5 ⁽¹⁸⁾.

In particular, 5.8 (i) and 5.9 show that

5.10. COROLLARY. *There is a π^* which bi-numerates P in \mathcal{P} for which*

$$\vdash_{\mathcal{P}} \text{Con}_{\pi^*}.$$

Rather than contradicting Gödel's second undecidability theorem, 5.9 and 5.10 show the importance of a precise method of dealing with consistency statements, at any rate for theories with infinitely many axioms. We have maintained that insofar as a formula a expresses membership in A , the formula Pr_{α} expresses provability of \mathcal{A} in $\mathcal{M}(\mathcal{P})$ and the sentence Con_{α} expresses the consistency of \mathcal{A} in \mathcal{M} and \mathcal{P} . Thus, one particular conclusion we can draw is that the formula α^* , although it extensionally corresponds to A , does not properly express membership in \mathcal{A} . Indeed, inspection of the proof of 5.9 reveals that it expresses membership in a certain subsystem of A which, independent of the consistency of \mathcal{A} , is always consistent.

⁽¹⁸⁾ Suppose that A is recursively enumerable and that α numerates A in \mathcal{A} . Construct α^* as in (1) above. Then (10) still holds but, as we saw in the remarks leading to 4.14, we cannot expect that α^* numerates A in \mathcal{A} . S. Orey has communicated to us an interesting technique for modifying the proof so as to obtain such α^* . By this means, several theorems below, in particular 5.11 and 6.6, can be obtained in more general form.

A first reaction following such realizations might be to restrict attention to a certain class of "natural" formulas α in problems of arithmetization; 5.6 suggests that the class of RE -formulas or some subclass of it might be appropriate for this purpose. However, we shall obtain in § 6 results through the use of arbitrary formulas α which should be of interest even to those who would otherwise thus restrict attention. There is nothing "wrong" with the use of arbitrary formulas α ; rather, the guiding consideration should be to investigate how different restrictions on the choice of α affect the results by arithmetization.

Let us return to the proof of 5.9. Although the arguments on the grounds of expressibility give us one way of seeing why the theorem is possible, we should also ask why it is not possible to carry through the proof of

$$\text{Con}_{\alpha^*} \rightarrow \nu_{\alpha^*},$$

as in 5.6—otherwise contradicting the non-provability of ν_{α^*} in \mathcal{A} obtained in 5.3. One answer is that there occurs an unbounded universal quantifier appearing in the part $\text{Con}_{\alpha \wedge x}$ of the definition of α^* ; hence, we cannot expect to find a bounded prenex formula equivalent to Pr_{α^*} . Then we cannot hope to prove that

$$\sim \nu_{\alpha^*} \rightarrow \text{Pr}_{\alpha^*}(\sim \bar{\nu}_{\alpha^*}),$$

which corresponds to the essential step in the proof of 5.6. Nevertheless, by the remark following 5.5, we shall be able to prove

$$\sim \nu_{\alpha} \rightarrow \text{Pr}_{\alpha^*}(\sim \bar{\nu}_{\alpha}),$$

when α is an RE -formula. This, when combined with 5.3 and 5.6, leads to the following statement.

5.11. THEOREM. *Suppose that $\mathcal{A} = \langle A, K \rangle$ is a consistent, reflexive axiom system with $\mathcal{P} \subseteq \mathcal{A}$ and that A is primitive recursive. Suppose that α is a PR -formula which bi-numerates A in \mathcal{A} , and let $B = A \cup \{\sim \text{Con}_{\alpha}\}$. Then there is a β^* which bi-numerates B in \mathcal{A} for which*

$$\vdash_{\mathcal{A}} \text{Con}_{\beta^*}.$$

Proof. Let α^* be the formula constructed in the proof of 5.9, statement (1), from the given α , such that $\vdash_{\mathcal{P}} \text{Con}_{\alpha^*}$. As in 5.6, we see that

$$(1) \quad \vdash_{\mathcal{A}} \sim \nu_{\alpha} \rightarrow \text{Pr}_{\mathcal{Q}}(\sim \bar{\nu}_{\alpha}).$$

Let A' be $Pr_{\mathcal{A}} \cap Fm_K$, and $\mathcal{A}' = \langle A', K_0 \rangle$; thus also $\mathcal{P} \subseteq \mathcal{A}'$. Since Pr_{α^*} numerates $\text{Pr}_{\mathcal{A}}$ in \mathcal{A} , and thence in \mathcal{A}' , and since \mathcal{Q} is finite, we see that

$$(2) \quad \vdash_{\mathcal{A}'} [\mathcal{Q}](x) \rightarrow \text{Pr}_{\alpha^*}(x),$$

hence, by 5.5,

$$(3) \quad \vdash_{\mathcal{A}' \cup \mathcal{M}} \sim \nu_{\alpha} \rightarrow \text{Pr}_{\alpha^*}(\sim \bar{\nu}_{\alpha}).$$

But $\vdash_{\mathcal{A}'} \text{Con}_{\alpha^*}$. Thus

$$(4) \quad \vdash_{\mathcal{A}' \cup \mathcal{M}} \sim \nu_{\alpha} \rightarrow \sim \text{Pr}_{\alpha^*}(\bar{\nu}_{\alpha}).$$

By definition (5.2) of ν_{α} , this shows

$$(5) \quad \vdash_{\mathcal{A}' \cup \mathcal{M}} \text{Pr}_{\alpha^*}(\bar{\nu}_{\alpha}) \rightarrow \sim \text{Pr}_{\alpha}(\bar{\nu}_{\alpha}).$$

On the other hand (8) of 5.9 and 4.7 (i) shows that

$$(6) \quad \vdash_{\mathcal{M}} \text{Pr}_{\alpha^*}(\bar{\nu}_{\alpha}) \rightarrow \text{Pr}_{\alpha}(\bar{\nu}_{\alpha}).$$

From (5) and (6) it follows that

$$(7) \quad \vdash_{\mathcal{A}' \cup \mathcal{M}} \sim \text{Pr}_{\alpha^*}(\bar{\nu}_{\alpha}).$$

We see from 5.6 that for some finite subsystem $\mathcal{F} = \langle F, K_0 \rangle$ of \mathcal{P} ,

$$(8) \quad \vdash_{\mathcal{F}} \text{Con}_{\alpha} \leftrightarrow \nu_{\alpha}$$

and hence

$$(9) \quad \vdash_{\mathcal{M}} \text{Pr}_{\mathcal{F}}(\overline{\text{Con}_{\alpha}} \leftrightarrow \bar{\nu}_{\alpha}).$$

But then we see, by the same argument as in (2) and (3), that

$$(10) \quad \vdash_{\mathcal{A}' \cup \mathcal{M}} \text{Pr}_{\alpha^*}(\overline{\text{Con}_{\alpha}} \leftrightarrow \bar{\nu}_{\alpha}),$$

and hence, from (7),

$$(11) \quad \vdash_{\mathcal{A}' \cup \mathcal{M}} \sim \text{Pr}_{\alpha^*}(\overline{\text{Con}_{\alpha}}).$$

Thus if we set

$$(12) \quad \beta^*(x) = \alpha^*(x) \vee x \approx \sim \overline{\text{Con}_{\alpha}},$$

we obtain

$$(13) \quad \vdash_{\mathcal{A}' \cup \mathcal{M}} \text{Con}_{\beta^*}.$$

From this it follows by 3.5 that, since the set of constants of \mathcal{A}' is K_0 , $\vdash_{\mathcal{A}'} \text{Con}_{\beta^*}$, i. e. $\vdash_{\mathcal{A}} \text{Con}_{\beta^*}$. That β^* is a bi-numeration of B in \mathcal{A} is clear from (12) and the fact that α^* bi-numerates A in \mathcal{A} .

The usefulness of this Theorem 5.11 will become apparent in the next section, especially in connection with Theorem 6.6.

6. Applications to problems of interpretability

The method of arithmetization is connected with problems of interpretability in at least two ways. In the first of these we consider the relation of interpretability between theories which number among their axioms some arithmetized statements. For example, what is the relation

between \mathcal{A} and $\mathcal{A} + \{\text{Con}_a\}$ or $\mathcal{A} + \{\sim \text{Con}_a\}$, for various choices of a ? The second connection with interpretability is less evident on the face of it; it is involved with the arithmetization of the semantical part of the proof of the completeness of predicate logic. However, these considerations lead to important results (based on earlier work of Hilbert-Bernays [14], Vol. II and of Wang [34]), which can in turn be used with problems of the first type. For this reason we treat the latter material first.

The main results to which we have just referred are stated as follows.

6.2. THEOREM. Let $\mathcal{A} = \langle A, K \rangle$ and \mathcal{S} be axiom systems, where $\mathcal{P} \subseteq \mathcal{S}$. Suppose that A numerates \mathcal{A} in \mathcal{S} . Then

$$\mathcal{A} \preceq \mathcal{S} + \{\text{Con}_a\}.$$

6.3. THEOREM. Let $\mathcal{A} = \langle A, K \rangle$ be an axiom system, and suppose that A numerates \mathcal{A} in some finite subtheory of \mathcal{P} . Then for some other finite subtheory $\mathcal{F} \subseteq \mathcal{P}$,

$$\mathcal{A} \preceq \mathcal{F} + \{\text{Con}_a\}.$$

It is seen that each of these is an immediate corollary of the following theorem, which we thus number "6.1". Its proof is closely related to the proof of the semantical Theorem 2.9 (which associated with each consistent axiom system \mathcal{A} a model in a subset of ω).

6.1. THEOREM. Let $\mathcal{A} = \langle A, K \rangle$ and $\mathcal{S} = \langle S, K' \rangle$ be axiom systems, $K_0 \subseteq K'$. Suppose that A numerates \mathcal{A} in \mathcal{S} . Then for some finite axiom system $\mathcal{F} \subseteq \mathcal{P}$,

$$\mathcal{A} \preceq (\mathcal{S} \cup \mathcal{F}) + \{\text{Con}_a\}.$$

Proof. We have seen in 4.12 that we can construct a formula γ of \mathcal{M} such that

$$(1) \quad \vdash_{\mathcal{M}} \text{Con}_a \rightarrow \bigwedge_x (a(x) \wedge \text{St}_K(x) \rightarrow \gamma(x)) \wedge \text{QCm}_\gamma.$$

If γ is not already a formula of \mathcal{P} , we can obtain such by forming $\gamma^{(2)}$; hence we assume this to be already done. It follows that there is a certain finite subsystem \mathcal{F}_0 of \mathcal{P} such that

$$(2) \quad \vdash_{\mathcal{F}_0} \text{Con}_a \rightarrow \bigwedge_x (a(x) \wedge \text{St}_K^{(2)}(x) \rightarrow \gamma(x)) \wedge \text{QCm}_\gamma.$$

We shall now proceed to show that there is a certain finite subsystem \mathcal{F}_1 of \mathcal{P} such that

$$(3) \quad \mathcal{A} \preceq (\mathcal{S} \cup \mathcal{F}_1) + \left\{ \bigwedge_x (a(x) \wedge \text{St}_K^{(2)}(x) \rightarrow \gamma(x)) \wedge \text{QCm}_\gamma \right\}.$$

Then our theorem will be proved when we take

$$(4) \quad \mathcal{F} = \mathcal{F}_0 \cup \mathcal{F}_1.$$

To prove (3), let us restrict ourselves, for simplicity, to the case $K = \{r_{1,0}, r_{1,1}\}$ (just as in the proof of 2.9). Let u, v, w be distinct variables, all different from x, y, z . We claim that there is a finite subsystem \mathcal{D}_1 of \mathcal{M} for which

(5) each of the following is provable in $\mathcal{D}_1 + \{\text{QCm}_\gamma\}$:

$$\begin{aligned} & \text{Vr}(u) \rightarrow \gamma(u \approx u), \\ & \text{Vr}(u) \wedge \text{Vr}(w) \wedge \gamma(u \approx w) \rightarrow \gamma(w \approx u), \\ & \text{Vr}(u) \wedge \text{Vr}(v) \wedge \text{Vr}(w) \wedge \gamma(u \approx v) \wedge \gamma(v \approx w) \rightarrow \gamma(u \approx w), \\ & \text{Vr}(u) \rightarrow \bigvee_z [\gamma(\text{Vr}_z \approx u) \wedge \bigwedge_w (\gamma(\text{Vr}_w \approx u) \rightarrow z \leq w)], \\ & \text{Fm}_K(x) \wedge \text{Vr}(u) \wedge \text{Vr}(w) \wedge \gamma(u \approx w) \wedge \gamma(x) \rightarrow \gamma(\text{Sb}_w^u x), \\ & \text{Pr}_\gamma(x) \rightarrow \gamma(x), \\ & \text{Fm}_K(x) \rightarrow [\gamma(\sim x) \leftrightarrow \sim \gamma(x)], \\ & \text{Fm}_K(x) \wedge \text{Fm}_K(y) \rightarrow [\gamma(x \rightarrow y) \leftrightarrow (\gamma(x) \rightarrow \gamma(y))], \\ & \text{Fm}_K(x) \wedge \text{Vr}(u) \rightarrow \{\gamma(\bigwedge_u x) \leftrightarrow \bigwedge_w [\text{Vr}(w) \rightarrow \gamma(\text{Sb}_w^u x)]\}. \end{aligned}$$

In addition we assume that certain elementary properties of the logical connectives and substitution are formalized in $\mathcal{D}_1 + \{\text{QCm}_\gamma\}$. To define the desired interpretation I , we set, as usual, $I(x \approx y) = (x \approx y)$ and then put

$$(6) \quad I(r_{1,1}[x, y]) = \gamma(\text{Rl}_{1,1}(\text{Vr}_x, \text{Vr}_y))^{(2)}$$

and for the "domain" of interpretation we take the formula

$$(7) \quad \delta(x) = \bigwedge_y [\gamma(\text{Vr}_x \approx \text{Vr}_y) \rightarrow x \leq y]^{(2)}.$$

The interpretation I is then determined for all formulas of K by the rules described in § 2 (d). The conditions (5) then guarantee that

$$(8a) \quad \vdash_{\mathcal{D}_1 + \{\text{QCm}_\gamma\}} \bigvee_x \delta(x),$$

and in fact

$$(8b) \quad \vdash_{\mathcal{D}_1 + \{\text{QCm}_\gamma\}} \bigwedge_y \bigvee_x [\delta(x) \wedge \gamma(\text{Vr}_x \approx \text{Vr}_y)].$$

The basic property which can be established in this theory is the following, corresponding to (4) of the proof of 2.9.

(9) if $\varphi \in \text{Fm}_K$ and $\text{Fv}(\varphi) \subseteq \{v_0, \dots, v_{n-1}\}$ then

$$\vdash_{\mathcal{D}_1 + \{\text{QCm}_\gamma\}} \delta(v_0) \wedge \dots \wedge \delta(v_{n-1}) \rightarrow \left[I(\varphi) \leftrightarrow \gamma \left(\text{Sb} \left(\frac{\text{Vr}_0}{\text{Vr}_{v_0}} \dots \frac{\text{Vr}_{n-1}}{\text{Vr}_{v_{n-1}}} \varphi \right) \right) \right].$$

This is proved by induction on φ , making use of the statements (5)-(8). At the induction step, to $\bigwedge_{v_n} \varphi$, where $Fv(\varphi) \subseteq \{v_0, \dots, v_n\}$, we must reduce

$$\bigwedge_{v_n} \left[\delta(v_n) \rightarrow \gamma \left(\text{Sb} \left(\frac{v_1 r_0}{v_1 r_{v_0}} \dots \frac{v_n r_n}{v_n r_{v_n}} \bar{\varphi} \right) \right) \right] \quad \text{to} \quad \bigwedge_{v_n} \gamma \left(\text{Sb} \left(\frac{v_1 r_0}{v_1 r_{v_0}} \dots \frac{v_n r_n}{v_n r_{v_n}} \bar{\varphi} \right) \right)$$

as an equivalent; this is possible by 8(b). Then using the last of the properties mentioned in (5), we can in turn reduce this, equivalently, to

$$\gamma \left(\text{Sb} \left(\frac{v_1 r_0}{v_1 r_{v_0}} \dots \frac{v_{n-1} r_{n-1}}{v_{n-1} r_{v_{n-1}}} \overline{\bigwedge_{v_n} \varphi} \right) \right).$$

From (9) we obtain

$$(10) \quad \text{if } \varphi \in \text{St}_K \text{ then } \vdash_{\mathcal{D}_1 + \{QCm_n\}} I(\varphi) \leftrightarrow \gamma(\bar{\varphi}).$$

Let \mathcal{F}_1 be a finite subsystem of \mathcal{P} for which $\vdash_{\mathcal{F}_1} \theta^{(St)}$ for each θ provable in \mathcal{D}_1 . Then we have

$$(11) \quad \text{if } \varphi \in \text{St}_K \text{ then } \vdash_{\mathcal{F}_1 + \{QCm_n\}} I(\varphi) \leftrightarrow \gamma(\bar{\varphi}).$$

We see that the choice of \mathcal{F}_1 depends just on γ and not on \mathcal{S} . Only now does \mathcal{S} enter the picture. We have presumed that α numerates A in \mathcal{S} . Hence

$$(12) \quad \text{for each } \varphi \in A, \vdash_{\mathcal{S}} \bigwedge_x (a(x) \wedge \text{St}_K^{(St)}(x) \rightarrow \gamma(x)) \rightarrow \gamma(\bar{\varphi}).$$

Combining (11) and (12) immediately gives the desired result (3).

Either of the main Theorems 6.2, 6.3 is a precise statement and generalization of the theorems of Hilbert-Bernays ([14], Vol. II, p. 252-253, for finite axiom systems) and of Wang [34] ⁽¹⁹⁾ ⁽²⁰⁾. Applications of these theorems will be found throughout the following, usually to obtain certain positive results about the existence of interpretations.

⁽¹⁹⁾ A discussion of the relationship of Wang's proof in [34] to the proof of the weaker theorem in [14] can be found in our review of [34]. *Journal of Symbolic Logic* 20 (1955), p. 76-77.

⁽²⁰⁾ Many of the same ideas used by Wang in [34] had also been explored by Novak in [26]. She deals there with theories (S') and (S) related, roughly speaking, as Gödel-Bernays set theory is related to Zermelo-Fraenkel set theory. Her main result is that "(S') has a denumerable consistent real model which is constructible in arithmetic under the arithmetical assumption equivalent to '(S) is consistent'" ([26], p. 106). However, her notion of "real model" does not provide that quantifiers transform into quantifiers, and it is not clear from her proof that this actually obtains. Wang [34], on the other hand, clearly obtains an interpretation in our sense.

A result analogous to Novak's could be obtained here, but would require somewhat more development, if it were to be done directly. Her result was used to establish the relative consistency of (S') to (S). Alternatively, one could formalize Shoenfields finitary proof of relative consistency in [30] to obtain $\vdash_{\mathcal{P}} \text{Con}_{(S)} \rightarrow \text{Con}_{(S')}$. Then one can apply 6.3 to see that (S') is relatively interpretable in $\mathcal{F} + \{\text{Con}_{(S)}\}$ (\mathcal{F} a finite subsystem of \mathcal{P}).

As a companion to these theorems, we want next to prove a lemma which will be very useful in drawing negative conclusions about existence of interpretations. The general form of these arguments will be to assume of the given systems \mathcal{A}, \mathcal{B} that $\mathcal{B} \preceq \mathcal{A}$, and then to obtain formal conclusions which can be shown contradictory. One such formal conclusion is that, for given a, β which numerate A, B respectively (say in \mathcal{P}), $\text{Con}_a \rightarrow \text{Con}_\beta$ is true. However, there is no reason to expect this implication to be provable in \mathcal{P} , for the verification that $\mathcal{B} \preceq \mathcal{A}$ when B is infinite may require a non-elementary argument. Nevertheless, it is possible to show that for suitable choice of a, β this can be carried through. A full discussion of the other logical possibilities connected with such formal relative consistency proofs will be found in § 7.

6.4. THEOREM. Suppose that $\mathcal{A} = \langle A, K \rangle$, $\mathcal{B} = \langle B, K' \rangle$ are axiom systems and that $\mathcal{B} \preceq \mathcal{A}$. Suppose further that B is recursively enumerable. Then for any RE-formula a numerating A in \mathcal{P} we can find an RE-formula β numerating B in \mathcal{P} for which

$$\vdash_{\mathcal{P}} \text{Con}_a \rightarrow \text{Con}_\beta.$$

If, further, B is finite then

$$\vdash_{\mathcal{P}} \text{Con}_a \rightarrow \text{Con}_{1B}.$$

Proof. Let α be given. Suppose $\mathcal{B} \stackrel{I}{\preceq} \mathcal{A}$.

I is a primitive recursive function determined, essentially, by its values on atomic formulas of \mathcal{B} . For other formulas it is defined by the conditions that it preserve \sim, \rightarrow and \wedge (the latter, at least, relativized); in other words, by construction, I preserves logical form. On this basis alone we see that

$$(1) \quad \text{if } \varphi \in \text{Fm}_{K'} \text{ and } \vdash_{K'} \varphi \text{ then } \vdash_{\mathcal{A}} I(\varphi).$$

Thus we can obtain a p. r. extension \mathcal{N} of \mathcal{P} such that $\mathcal{M} \subseteq \mathcal{N}$ and such that for a certain function symbol f of \mathcal{N} ,

$$(2) \quad f[x] \approx y \text{ numerates the function } I \text{ in } \mathcal{N},$$

$$(3) \quad \vdash_{\mathcal{N}} \text{Fm}_{K'}(x) \rightarrow f[\sim x] \approx \sim f[x],$$

$$(4) \quad \vdash_{\mathcal{N}} \text{Fm}_{K'}(x) \wedge \text{Fm}_{K'}(y) \rightarrow f[x \rightarrow y] \approx f[x] \rightarrow f[y],$$

and, corresponding to (1),

$$(5) \quad \vdash_{\mathcal{N}} \text{Pr}_{[0]}(x) \rightarrow \text{Pr}_a(f[x]).$$

Let β_0 be a fixed RE-formula which numerates B in \mathcal{P} . Then we can find a formula β in \mathcal{P} which is an RE-formula if a is and such that

$$(6) \quad \vdash_{\mathcal{N}} \beta(x) \leftrightarrow \beta_0(x) \wedge \text{Pr}_a(f[x]).$$

Since, by hypothesis, $\vdash_{\mathcal{A}} I(\varphi)$ for each $\varphi \in B$, we see that β numerates B in \mathcal{N} , and hence also in \mathcal{P} . It follows from (4)-(6), and the arithmetized version 4.6 (v) of proof by induction on provable formulas of K' , that

$$(7) \quad \vdash_{\mathcal{N}} \text{Pr}_{\beta}(\bar{x}) \rightarrow \text{Pr}_a(f[\bar{x}]).$$

Thus we see by (3) that

$$(8) \quad \vdash_{\mathcal{N}} \text{Con}_a \rightarrow \text{Con}_{\beta},$$

which gives us the desired result by 3.5. In case B is finite we can, in addition, apply 4.7 (i) to see that

$$(9) \quad \vdash_{\mathcal{P}} \text{Con}_{\beta} \rightarrow \text{Con}_{\lfloor B \rfloor}.$$

In case we had a primitive recursive translation I (in the sense of [35]) which is not an interpretation, we would not be able, in general, to verify that it preserves full logical form, and would not be able to derive the essential steps (4), (5) needed to carry through the induction for (7). This will be discussed further in § 7.

We can now obtain the following result.

6.5. THEOREM. *Suppose that $\mathcal{A} = \langle A, K \rangle$ is a consistent axiom system with $\mathcal{P} \subseteq \mathcal{A}$. Suppose further that a is an RE-formula numerating A in \mathcal{P} . Then*

$$\mathcal{A} + \{\text{Con}_a\} \not\leq \mathcal{A}.$$

Proof. Let $\mathcal{B} = \mathcal{A} + \{\text{Con}_a\}$. If $\mathcal{B} = \langle B, K \rangle$ is not consistent, we certainly do not have \leq . Let us suppose, then, that \mathcal{B} is consistent. Suppose further that $\mathcal{B} \leq \mathcal{A}$. Then by 6.4, there is an RE-formula β numerating B in \mathcal{P} , such that

$$\vdash_{\mathcal{P}} \text{Con}_a \rightarrow \text{Con}_{\beta}.$$

But this implies that

$$\vdash_{\mathcal{B}} \text{Con}_{\beta},$$

which contradicts 5.6.

In contrast to 6.5 we now obtain the following theorem.

6.6. THEOREM. *Suppose that $\mathcal{A} = \langle A, K \rangle$ is a reflexive axiom system with A primitive recursive and $\mathcal{P} \subseteq \mathcal{A}$. Let a be any PR-formula which bi-numerates A in \mathcal{A} . Then*

$$\mathcal{A} + \{\sim \text{Con}_a\} \leq \mathcal{A}.$$

Moreover, the same holds if A is recursively enumerable, \mathcal{A} is ω -consistent, and a is an RE-formula which numerates A in \mathcal{A} .

Proof. First consider the case that A is primitive recursive and a is a PR-formula. Let $\mathcal{B} = \mathcal{A} + \{\sim \text{Con}_a\}$. If \mathcal{A} is not consistent, it is of course true that $\mathcal{B} \leq \mathcal{A}$. If \mathcal{A} is consistent, we know by 5.11 that there is a β^* which bi-numerates \mathcal{B} in \mathcal{A} for which

$$\vdash_{\mathcal{A}} \text{Con}_{\beta^*}.$$

By 6.2, $\mathcal{B} \leq \mathcal{A} + \{\text{Con}_{\beta^*}\}$, hence $\mathcal{B} \leq \mathcal{A}$. Now consider the case that A is recursively enumerable and \mathcal{A} is ω -consistent; then a numerates A in \mathcal{Q} . By 4.13 we can construct a PR-formula a_1 and a system $\mathcal{A}_1 = \langle A_1, K \rangle$ such that a_1 bi-numerates A_1 in \mathcal{Q} , $\text{Pr}_{\mathcal{A}_1} = \text{Pr}_{\mathcal{A}}$, and

$$\vdash_{\mathcal{P}} \text{Pr}_a(\bar{x}) \leftrightarrow \text{Pr}_{a_1}(\bar{x}).$$

Then certainly

$$\vdash_{\mathcal{A}} \text{Con}_a \leftrightarrow \text{Con}_{a_1}.$$

By the preceding part of the theorem

$$\mathcal{A}_1 + \{\sim \text{Con}_{a_1}\} \leq \mathcal{A}_1,$$

hence we can obtain the desired conclusion for \mathcal{A} .

We see in particular that if π is the "natural" definition of P we have

$$\mathcal{P} + \{\sim \text{Con}_{\pi}\} \leq \mathcal{P}.$$

In other words, we can construct a "non-standard model" of \mathcal{P} within \mathcal{P} which, moreover, we can verify, axiom by axiom, to be a model of $\mathcal{P} + \{\sim \text{Con}_{\pi}\}$ ⁽²¹⁾.

So far, of the basic theorems 6.1-6.3, we have only applied 6.2. We now turn to some applications of 6.1, 6.3. If \mathcal{P} is ω -consistent then we can obtain from 6.2, 6.4 and the reflexiveness of \mathcal{P} (5.8) that the consistency of a system \mathcal{A} is provable in arithmetic,

$$\vdash_{\mathcal{P}} \text{Con}_a,$$

where a is an RE-numeration of A in \mathcal{Q} , just in case there is a finite subsystem $\mathcal{F} \subseteq \mathcal{P}$ for which $\mathcal{A} \leq \mathcal{F}$. We obtain now the following more general result, without hypothesis of ω -consistency, when A is primitive recursive.

6.7. THEOREM. *Let $\mathcal{A} = \langle A, K \rangle$ and \mathcal{S} be axiom systems with $\mathcal{P} \subseteq \mathcal{S}$ and A primitive recursive. Suppose further that \mathcal{S} is reflexive. Then the following three conditions are equivalent:*

(i) *there is an RE-formula a which numerates A in some finite subsystem of \mathcal{P} for which*

$$\vdash_{\mathcal{S}} \text{Con}_a;$$

(ii) *there is an a which numerates A in some finite subsystem of \mathcal{S} for which*

$$\vdash_{\mathcal{S}} \text{Con}_a;$$

⁽²¹⁾ This result should not be completely unexpected. Loosely speaking, 6.4 and 6.6 guarantee $\vdash_{\mathcal{P}} \text{Con}_{\mathcal{A}} \rightarrow \text{Con}_{\mathcal{A} + \{\sim \text{Con}_a\}}$. In fact, and speaking precisely, it already follows from 5.6 that $\vdash_{\mathcal{P}} \text{Con}_a \rightarrow \sim \text{Pr}_a(\overline{\text{Con}_a})$; however, this implication as it stands would not be sufficient to prove 6.6 from 6.2.

(iii) for some finite subsystem $\mathcal{F} \subseteq \mathcal{S}$,

$$\mathcal{A} \preceq \mathcal{F}.$$

Under the assumption that \mathcal{P} is ω -consistent, the above are equivalent even when A is recursively enumerable⁽²²⁾ (23).

Proof. Clearly (i) implies (ii). Suppose (ii). Let \mathcal{F}_0 be a finite subsystem of \mathcal{S} in which a numerates A . By 6.1 there is a finite subsystem \mathcal{F}_1 of \mathcal{P} such that

$$\mathcal{A} \preceq (\mathcal{F}_0 \cup \mathcal{F}_1) + \{\text{Con}_a\}.$$

The system on the right is a finite subsystem of \mathcal{S} . Suppose (iii). To show that (i) holds, we return to the proof of 6.4 and modify it slightly. We are now considering “ \mathcal{F} ” for “ \mathcal{A} ” and “ \mathcal{A} ” for “ \mathcal{B} ” there. Let $a_0(x)$ be a PR -formula which numerates A in \mathcal{Q} . Let $\alpha(x)$ be an RE -formula such that

$$\vdash_{\mathcal{N}} \alpha(x) \leftrightarrow a_0(x) \wedge \text{Pr}_{\Gamma[\mathcal{F}]}(f[x]).$$

Then for some finite subsystem \mathcal{Q}' of \mathcal{P} , where $\mathcal{Q} \subseteq \mathcal{Q}'$,

$$\vdash_{\mathcal{Q}'} \alpha(x) \leftrightarrow a_0(x) \wedge (\text{Pr}_{\Gamma[\mathcal{F}]}(f[x]))^{(24)},$$

i. e.

$$\vdash_{\mathcal{Q}'} \alpha(x) \leftrightarrow a_0(x) \wedge \bigvee_y \bigvee_w [(w \approx f[x])^{(25)} \wedge \text{Pr}_{\Gamma[\mathcal{F}]}(w, y)].$$

Then if $\varphi \in A$, and $\theta = I(\varphi)$, where $\mathcal{A} \preceq \mathcal{F}$, and $\text{Pr}_{\mathcal{F}}[\theta, \Psi]$, we have

$$\vdash_{\mathcal{Q}} a_0(\bar{\varphi}) \wedge (\bar{\theta} \approx f[\bar{\varphi}])^{(25)} \wedge \text{Pr}_{\Gamma[\mathcal{F}]}(\bar{\theta}, \bar{\Psi}),$$

hence

$$\vdash_{\mathcal{Q}'} \alpha(\bar{\varphi}).$$

On the other hand if $\varphi \notin A$, $\vdash_{\mathcal{Q}} \sim a_0(\bar{\varphi})$ and then $\vdash_{\mathcal{Q}'} \sim \alpha(\bar{\varphi})$, so that (by the consistency of \mathcal{P}), a numerates A in \mathcal{Q}' . The remainder of the proof of 6.4 still holds, so that

$$\vdash_{\mathcal{P}} \text{Con}_{\Gamma[\mathcal{P}]} \rightarrow \text{Con}_a.$$

⁽²²⁾ The implication, (iii) implies (i), was obtained for A finite and $\mathcal{S} = \mathcal{P}$ by Kreisel and Wang in [19], p. 108.

⁽²³⁾ Wang argues ([36], p. 449) that the consistency $\text{Con}(\mathcal{S})$ of every consistent decidable theory is provable in arithmetic, but this is erroneous except on a completely extensional understanding of the structure of $\text{Con}(\mathcal{S})$ —from which no interesting consequences could be obtained. This does suggest, though, the following interesting open problem: is it true that for every decidable system $\mathcal{A} = \langle A, K \rangle$ (A recursively enumerable) there exists an RE -numeration α of A for which $\vdash_{\mathcal{P}} \text{Con}_a$? As our Theorem 6.7 shows (if \mathcal{P} is ω -consistent), this is equivalent to the question: is it true that every decidable theory \mathcal{A} is (relatively) interpretable in some finite subtheory \mathcal{F} of \mathcal{P} ? Presumably, both conclusions hold for all known particular decidable theories; we conjecture, however, that the answer to the general question is negative.

Since \mathcal{S} is reflexive, this gives us the desired result (i). To obtain the last remark of our theorem, we need only modify this proof of: (iii) implies (i). Take a_0 to be an RE -formula which numerates A in \mathcal{Q} . Then under the hypothesis of ω -consistency, it will also numerate A in \mathcal{Q}' . Hence, the same will be true of a .

This theorem has two interesting consequences. It allows us, under certain circumstances, to extend the second undervivability theorem to arbitrary numerations in finite subsystems. It also allows us to sharpen the results of Mostowski, Kreisel-Wang and Montague (cf. 5.8 f.) on non-finitizability to results of non-interpretability.

6.8. THEOREM. Suppose that $\mathcal{A} = \langle A, K \rangle$ is a consistent, reflexive axiom system with $\mathcal{P} \subseteq \mathcal{A}$. Then we have:

(i) if A is primitive recursive and a is any numeration of A in any finite subsystem of \mathcal{A} then

$$\text{not } \vdash_{\mathcal{A}} \text{Con}_a;$$

(ii) for any finite subsystem $\mathcal{F} \subseteq \mathcal{A}$ we have

$$\mathcal{A} \not\preceq \mathcal{F}.$$

Under the assumption that \mathcal{P} is ω -consistent (i) holds even if A is recursively enumerable.

Proof. The proof of (i) is immediate by 5.6 and 6.7 (i), (ii). In (ii) we are not assuming that A is recursively enumerable. Suppose, however,

$$\mathcal{A} \preceq \mathcal{F} \text{ where } \mathcal{F} \subseteq \mathcal{A}.$$

Let A' consist of all sentences φ of K such that

$$\vdash_{\mathcal{F}} I(\varphi).$$

Then A' is recursively enumerable and $\mathcal{A}' \subseteq \mathcal{A}' = \langle A', K \rangle$, and clearly \mathcal{A}' is consistent. Moreover, \mathcal{A}' is reflexive. For suppose $\mathcal{D} = \langle D, K \rangle$ is a finite subsystem of \mathcal{A}' . Then by 6.4, from $\mathcal{D} \preceq \mathcal{F}$ we can obtain

$$\vdash_{\mathcal{D}} \text{Con}_{\Gamma[\mathcal{D}]} \rightarrow \text{Con}_{\mathcal{D}}.$$

But, since \mathcal{A} is assumed to be reflexive, $\vdash_{\mathcal{A}} \text{Con}_{\Gamma[\mathcal{P}]}$, hence $\vdash_{\mathcal{A}} \text{Con}_{\mathcal{D}}$ and then $\vdash_{\mathcal{A}'} \text{Con}_{\mathcal{D}}$. Now by 4.13 there is a primitive recursive set of sentences A'_1 such that, for $\mathcal{A}'_1 = \langle A'_1, K \rangle$, $\text{Pr}_{\mathcal{A}'} = \text{Pr}_{\mathcal{A}'_1}$. Then also $\mathcal{P} \subseteq \mathcal{A}'_1$ and \mathcal{A}'_1 is consistent and reflexive. Now we can apply 6.7 (i), (iii) to see that for some RE -formula a'_1 which numerates A'_1 in some finite subsystem of \mathcal{P} , $\vdash_{\mathcal{A}'_1} \text{Con}_{a'_1}$. However, this contradicts 5.6.

As a concluding application of the methods developed here we wish to present a theorem discovered by S. Orey. The proof has not been published elsewhere and is given here with his permission. Orey describes it as being an “arithmetical compactness theorem”.

6.9. THEOREM. Suppose that $\mathcal{A} = \langle A, K \rangle$ and $\mathcal{S} = \langle S, K' \rangle$ are axiom systems with A, S recursively enumerable. Suppose that \mathcal{S} is reflexive and $\mathcal{P} \subseteq \mathcal{S}$. Suppose, finally, that for each finite subsystem $\mathcal{D} \subseteq \mathcal{A}$ we have $\mathcal{D} \preceq \mathcal{S}$. Then

$$\mathcal{A} \preceq \mathcal{S}.$$

Proof. It is seen by 4.13 that we can restrict ourselves to the case A recursive. We may assume \mathcal{S} is consistent, for otherwise $\mathcal{A} \preceq \mathcal{S}$ is obvious. For each n there is a finite subsystem $\mathcal{F}_n = \langle F_n, K' \rangle$ of \mathcal{S} such that

$$\mathcal{A} \upharpoonright n \preceq \mathcal{F}_n.$$

Then by 6.4,

$$\vdash_{\mathcal{P}} \text{Con}_{[F_n]} \rightarrow \text{Con}_{[A \upharpoonright n]},$$

hence

$$\vdash_{\mathcal{S}} \text{Con}_{[A \upharpoonright n]}.$$

Let a bi-numerate A in \mathcal{S} . Then construct a^* from a just as in the proof of 5.9. That proof, applied to the present conditions shows that

$$(1) \quad a^* \text{ is a bi-numeration of } A \text{ in } \mathcal{S}$$

and

$$(2) \quad \vdash_{\mathcal{S}} \text{Con}_{a^*}.$$

Hence $\mathcal{A} \preceq \mathcal{S}$ by 6.3.

7. Relative consistency proofs in arithmetic

This section is devoted to a detailed investigation of two topics that were touched upon earlier. The first of these is connected with the problem of selecting a single sentence of arithmetic to express the consistency of a theory. The second of these is the problem of the relationship between interpretability and relative consistency proofs. As we shall see, these are closely tied together.

Let $\mathcal{A} = \langle A, K \rangle$, where A is recursively enumerable. As we have already suggested, 5.6 and 5.9 would seem to indicate that in a search for the "natural" formulas a to be used in expressing membership in A , one should restrict attention to, say, the *RE*-formulas. Moreover, one might hope to provide a method of associating a single *RE*-formula a with A , which would be the "most natural" for expressing consistency, or, at least, of associating a single such *PR*-formula a with each primitive recursive A . We believe the result 7.4 to be described below shows that such hopes must be given up.

One way of comparing different numerations a, a' of the same set A is to ask whether the implication

$$\text{Con}_{a'} \rightarrow \text{Con}_a,$$

or its converse, is provable in a given system \mathcal{B} .

7.1. DEFINITION. Let $\mathcal{B} = \langle B, K' \rangle$, $K_0 \subseteq K$, and suppose that $a, a' \in \text{Fm}_{K_0}$, $\text{Fv}(a) = \text{Fv}(a') = \{x\}$. We put:

- (i) $a \leq_{\mathcal{B}} a'$ if $\vdash_{\mathcal{B}} \text{Con}_{a'} \rightarrow \text{Con}_a$;
- (ii) $a <_{\mathcal{B}} a'$ if $a \leq_{\mathcal{B}} a'$ but $a' \not\leq_{\mathcal{B}} a$.

It is immediate that if $\mathcal{B} \subseteq \mathcal{C}$ and $a \leq_{\mathcal{B}} a'$ then $a \leq_{\mathcal{C}} a'$; however, it may be that $a <_{\mathcal{B}} a'$ but only $a' \leq_{\mathcal{C}} a$. On the other hand, if we show $a <_{\mathcal{C}} a'$, we have obtained more information than if we show $a <_{\mathcal{B}} a'$.

Our first result is to show that for finite sets A , the natural definition $[A]$ is minimal in the ordering $\leq_{\mathcal{P}}$.

7.2. LEMMA. Suppose that $\mathcal{A} = \langle A, K' \rangle$ and that A is finite. If a' is any numeration of A in \mathcal{P} (or any subsystem of \mathcal{P}) then

$$[A] \leq_{\mathcal{P}} a'.$$

The proof is immediate from 4.10 (iii), 3.5. It is not difficult to give examples of numerations a' of A where $a' \not\leq_{\mathcal{P}} [A]$; we shall obtain a more general result in 7.5. We can derive the following from 7.2.

7.3. THEOREM. Suppose that $\mathcal{A} = \langle A, K' \rangle$ is finitely axiomatizable. Then there is an *RE*-formula a which numerates A in \mathcal{P} such that for any numeration a' of A in \mathcal{P} (or any subsystem of \mathcal{P}) we have

$$a \leq_{\mathcal{P}} a'.$$

To see this, let a_0 be any fixed *RE*-formula which numerates A in \mathcal{P} . Let $\mathcal{F} = \langle F, K' \rangle$ be a finite subsystem of \mathcal{A} with $\text{Pr}_{\mathcal{A}} = \text{Pr}_{\mathcal{F}}$. Then take a to be an *RE*-formula for which

$$\vdash_{\mathcal{P}} a(x) \leftrightarrow a_0(x) \wedge \text{Pr}_{[F]}(x).$$

For any other numeration a' of A in \mathcal{P} we shall be able to show by 4.7 (i) that $\vdash_{\mathcal{P}} \text{Pr}_{[F]}(x) \rightarrow \text{Pr}_{a'}(x)$, and hence $\vdash_{\mathcal{P}} \text{Pr}_a(x) \rightarrow \text{Pr}_{a'}(x)$.

7.2 and 7.3 show that for finite or finitely axiomatizable systems, a "most natural" choice of numeration of A is a minimum in the relation $\leq_{\mathcal{P}}$. However, this cannot be extended to arbitrary A , as we now see from the following theorem.

7.4. THEOREM. Suppose that $\mathcal{A} = \langle A, K' \rangle$ is a consistent, reflexive axiom system with $\mathcal{P} \subseteq \mathcal{A}$. Then with each a which is a *PR*-formula numerating A in \mathcal{P} we can effectively associate a *PR*-formula a' numerating A in \mathcal{P} for which

$$a' <_{\mathcal{A}} a.$$

Under the assumption that \mathcal{P} is ω -consistent, the above also holds true with "RE" instead of "PR".

Proof. Consider first the case that a is a PR -formula. Let $B = A \cup \{\sim \text{Con}_a\}$. By 5.6, $\mathcal{B} = \langle B, K \rangle$ is consistent. Construct a numeration β of B which is a PR -formula and for which

$$(1) \quad \vdash_{\mathcal{P}} \beta(x) \leftrightarrow a(x) \vee x \approx \sim \overline{\text{Con}_a}.$$

Using the diagonal construction 5.1, determine φ such that

$$(2) \quad \vdash_{\mathcal{Q}} \varphi \leftrightarrow \bigwedge_z [\text{Prf}_\beta(\bar{\varphi}, z) \rightarrow \sim \text{Con}_{a \uparrow z}].$$

Suppose $\vdash_{\mathcal{B}} \varphi$. Then for some n , we would have $\vdash_{\mathcal{B}} \sim \text{Con}_{a \uparrow n}$, which would make \mathcal{B} inconsistent according to 4.14. Hence

$$(3) \quad \text{not } \vdash_{\mathcal{B}} \varphi.$$

Now construct a numeration a' of A which is a PR -formula and for which

$$(4) \quad \vdash_{\mathcal{P}} a'(x) \leftrightarrow a(x) \wedge \bigwedge_y [y < x \rightarrow \sim \text{Prf}_\beta(\bar{\varphi}, y)].$$

Clearly $a' \leq_{\mathcal{P}} a$ and hence

$$(5) \quad a' \leq_{\mathcal{A}} a.$$

To show that $a \not\leq_{\mathcal{A}} a'$, we shall show that

$$(6) \quad \vdash_{\mathcal{P}} \sim \varphi \rightarrow \text{Con}_{a'}$$

but

$$(7) \quad \text{not } \vdash_{\mathcal{A}} \sim \varphi \rightarrow \text{Con}_a.$$

In fact, (7) is just another way of stating (3). To prove (6), we see first from (2) that

$$(8) \quad \vdash_{\mathcal{P}} \sim \varphi \rightarrow \bigvee_z [\text{Prf}_\beta(\bar{\varphi}, z) \wedge \text{Con}_{a \uparrow z} \wedge \bigwedge_y (y < z \rightarrow \sim \text{Prf}_\beta(\bar{\varphi}, y))],$$

since

$$\vdash_{\mathcal{P}} \text{Con}_{a \uparrow z} \wedge y < z \rightarrow \text{Con}_{a \uparrow y}.$$

Hence, by (4), we have

$$(9) \quad \vdash_{\mathcal{P}} \sim \varphi \rightarrow \bigvee_z [\text{Con}_{a \uparrow z} \wedge \bigwedge_x (a'(x) \leftrightarrow a(x) \wedge x \leq z)],$$

which immediately gives (6). To prove the theorem in case a is an RE -formula, we apply 4.13, taking $\mathcal{S} = \mathcal{P}$. Construct primitive recursive A_1 and a PR -formula α_1 , numerating A_1 in \mathcal{Q} such that $Pr_{\mathcal{A}} = Pr_{\mathcal{A}_1}$ and

$$(10) \quad \vdash_{\mathcal{P}} \text{Con}_a \leftrightarrow \text{Con}_{\alpha_1}.$$

Now by the first part of our proof we can find a PR -formula α'_1 such that $\alpha'_1 <_{\mathcal{A}} \alpha_1$. By 4.13 we can then find an RE -formula α' such that α' numerates A in \mathcal{P} and

$$(11) \quad \vdash_{\mathcal{P}} \text{Con}_{\alpha'} \leftrightarrow \text{Con}_{\alpha'_1}.$$

But then (10) and (11) show that $\alpha' <_{\mathcal{A}} a$.

Of course, we could equivalently state the theorem in both cases for numerations in \mathcal{Q} , but we have done it for \mathcal{P} as a basis for comparison with 7.3.

Theorem 7.4 does not apply to arbitrary bi-numerations a , for we know by 5.9 that (in case A is recursive) there is such an a for which $\vdash_{\mathcal{P}} \text{Con}_a$ — but then a is a minimum in the relation $\leq_{\mathcal{P}}$.

The example mentioned following 7.2 can be obtained from the following theorem, according to which there is, in general, no maximum in the $\leq_{\mathcal{A}}$ relation.

7.5. THEOREM. *Suppose that $\mathcal{A} = \langle A, K \rangle$ is an ω -consistent axiom system with $\mathcal{P} \subseteq \mathcal{A}$. Then with each PR -formula (RE -formula, bi-numeration) a which numerates A in \mathcal{P} we can effectively associate a numeration a' of A of the same type for which*

$$a <_{\mathcal{A}} a'.$$

Proof. Consider first the case that a is a PR -formula or bi-numeration. Set $B = A \cup \{\text{Con}_a\}$. Then $\mathcal{B} = \langle B, K \rangle$ is consistent. Determine a numeration β of B of the same type as a for which

$$\vdash_{\mathcal{P}} \beta(x) \leftrightarrow a(x) \vee x \approx \overline{\text{Con}_a}.$$

By 5.3,

$$\text{not } \vdash_{\mathcal{B}} \nu_\beta.$$

Next determine a numeration a' of A of the same type as a for which

$$\vdash_{\mathcal{P}} a'(x) \leftrightarrow a(x) \vee \text{Fm}_K^{(M)}(x) \wedge \bigvee_y (y < x \wedge \text{Prf}_\beta(\bar{\nu}_\beta, y)).$$

then it is seen that

$$\vdash_{\mathcal{P}} \text{Con}_{a'} \rightarrow \nu_\beta$$

since

$$\vdash_{\mathcal{P}} \sim \nu_\beta \rightarrow \bigvee_y \bigwedge_x (y < x \wedge \text{Fm}_K^{(M)}(x) \rightarrow a'(x)).$$

On the other hand, by construction of \mathcal{B} ,

$$\text{not } \vdash_{\mathcal{A}} \text{Con}_a \rightarrow \nu_\beta.$$

Thus $a' <_{\mathcal{A}} a$. Clearly $a \leq_{\mathcal{A}} a'$, so that $a <_{\mathcal{A}} a'$. Since \mathcal{A} is presumed ω -consistent, we can also obtain the conclusion for RE -formulas, by applying the same argument as at the end of 7.4.

The moral of these theorems is not to reject the use of particular numerations for known particular axiom systems \mathcal{A} ; for example, we still consider “natural” the definition π of P as a finite set of axioms and axiom schemata. Rather, it is to reject the use, as a well defined idea, of sentences $\text{Con}_{\mathcal{A}}$ associated with arbitrary systems \mathcal{A} .

S. Orey has pointed out to us that the type of construction used in 7.5 can also be used to obtain the following unexpected theorem⁽²⁴⁾.

7.6. THEOREM. Suppose that $\mathcal{A} = \langle A, K \rangle$ and $\mathcal{B} = \langle B, K' \rangle$ are axiom systems with A, B recursively enumerable, and that \mathcal{B} is consistent. Then with each PR-formula (bi-numeration) β which numerates B in \mathcal{P} we can effectively associate a numeration a of A of the same type for which

$$\vdash_{\mathcal{P}} \text{Con}_a \rightarrow \text{Con}_\beta.$$

Under the assumption that \mathcal{P} is ω -consistent the same holds true with "RE" instead of "PR".

Proof. Consider first the case that β is a PR-formula or bi-numeration of B . Let a_0 be a fixed PR-formula numerating A in \mathcal{P} . Let φ_0 be any fixed K' sentence. Determine a numeration a of A of the same type as β for which

$$\vdash_{\mathcal{P}} a(x) \leftrightarrow a_0(x) \vee \text{Fm}_K^{(M)}(x) \wedge \bigvee_y [y < x \wedge \text{Prf}_\beta(\varphi_0 \wedge \sim \varphi_0, y)].$$

(That a is indeed a numeration of A follows from the consistency of B .) Then it seen again that the formal assumption $\sim \text{Con}_\beta$ would lead to the conclusion $\sim \text{Con}_a$. To obtain the result for RE-formulas we use the same argument as at the end of 7.4.

Thus it is seen that in one sense the relative consistency of two theories can be proved even where there is no connection between the theories. This returns us to the question concerning relative consistency proofs raised prior to 6.4. Wang ([35], Theorem 1) suggests that if a system $\mathcal{B} \preceq \mathcal{A}$ then we have

$$\vdash_{\mathcal{P}} \text{Con}_{\mathcal{A}} \rightarrow \text{Con}_{\mathcal{B}}$$

($\text{Con}_{\mathcal{A}}, \text{Con}_{\mathcal{B}}$ being ambiguously determined); actually he states this is even the case if \mathcal{B} is translatable into \mathcal{A} . He then goes on to argue (loc. cit., Theorem 2) that if $\mathcal{P} \subseteq \mathcal{A}$ and \mathcal{A} is ω -consistent then

$$\mathcal{A} + \{\text{Con}_{\mathcal{A}}\} \not\leq \mathcal{A};$$

for otherwise,

$$\vdash_{\mathcal{P}} \text{Con}_{\mathcal{A}} \rightarrow \text{Con}_{\mathcal{A} + \{\text{Con}_{\mathcal{A}}\}},$$

i. e.

$$\vdash_{\mathcal{A} + \{\text{Con}_{\mathcal{A}}\}} \text{Con}_{\mathcal{A} + \{\text{Con}_{\mathcal{A}}\}},$$

violating Gödel's second theorem. If his first argument holds for translatability as well as \preceq , this should also show $\mathcal{A} + \{\text{Con}_{\mathcal{A}}\}$ not translatable into \mathcal{A} . Contrary to this, Kreisel has shown in [18] that every consistent

⁽²⁴⁾ Several other interesting results connected with the material discussed in this section and the preceding have been announced by Orey in [27].

recursively axiomatizable theory is translatable into \mathcal{P} ; the proof will be reproduced below. Even for \preceq , the first statement of Wang appears doubtful, unless \mathcal{B} is finite, and we can formally verify the existence of the interpretation. (The other possibility is that it is intended to deal with "provable" translations and interpretations to begin with; in this case the desired conclusions are true but trivial.)

The question arises as to what sense can be ascribed to such statements as the above in our framework. One such sense has already been provided in 6.4, and, as we have seen in 6.5, this was just sufficient to allow us to conclude that

$$\mathcal{A} + \{\text{Con}_a\} \not\leq \mathcal{A},$$

when a is an RE-formula (\mathcal{A} as above). Another logical possibility would be to say that if $\mathcal{B} \preceq \mathcal{A}$ then there exist numerations α, β of A, B in \mathcal{P} such that

$$\vdash_{\mathcal{P}} \text{Con}_\alpha \rightarrow \text{Con}_\beta.$$

Indeed, this is true, for the Theorem 7.6 just proved shows it to hold without any assumption connecting \mathcal{B} and \mathcal{A} , except that \mathcal{B} is consistent. It could thus not be used to obtain 6.5 (which would also be seen by making an attempt to do so).

In order to proceed in this way to obtain a complete picture of the logical possibilities, we shall now show that 6.4 fails when we take instead of interpretability, one of the relations of translatability. We shall say that an axiom system $\mathcal{B} = \langle B, K' \rangle$ is translatable (in the weak sense) into $\mathcal{A} = \langle A, K \rangle$ by the function F if:

- (i) F is a recursive function with domain $St_{K'}$ and range a subset of St_K ;
- (ii) if $\varphi \in St_{K'} \cap Pr_{\mathcal{B}}$ then $F(\varphi) \in Pr_{\mathcal{A}}$;
- (iii) for each $\varphi \in St_{K'}$, $F(\sim\varphi) = \sim F(\varphi)$.

(That is one of the notions discussed by Wang in [35].) We would say that we have "provable" translatability of \mathcal{B} into \mathcal{A} within a system \mathcal{S} if, in addition, arithmetized versions of (i)-(iii) are provable in \mathcal{S} . Kreisel describes a proof in [18] that every system \mathcal{B} is provably translatable into \mathcal{P} within \mathcal{P} if a certain relative consistency statement involving \mathcal{B} and \mathcal{P} is provable in \mathcal{P} . However, he has pointed out to us that his argument serves to give ordinary translatability in all cases, without these further assumptions. Namely, we have: If \mathcal{B} is any consistent recursively axiomatizable axiom system then \mathcal{B} is translatable (in the weak sense) into \mathcal{P} .

The proof of this, which is closely related to the Rosser construction and argument in [29], can easily be described here. By Craig's theorem [1], or 4.13, it is clearly sufficient to carry it through for the case that B is primitive recursive. Let β be a PR-formula which numerates B in \mathcal{Q} .

Let UN be the set of all un-negated sentences of K' , i. e. all those $\varphi \in St_{K'}$ such that $\varphi \neq \sim\psi$ for all ψ . Every sentence of K' is logically equivalent either to a sentence φ or a sentence $\sim\varphi$, where $\varphi \in UN$. For $\varphi \in UN$ set

$$F(\varphi) = \bigvee_y [\text{Prf}_\beta(\bar{\varphi}, y) \wedge \bigwedge_z (z < y \rightarrow \sim \text{Prf}_\beta(\sim\bar{\varphi}, z))].$$

F is then defined on $St_{K'}$ by repeated application of the rule $F(\sim\varphi) = \sim F(\varphi)$. To show that (ii) holds, we need only show that for each $\varphi \in UN$,

$$(ii) (a) \quad \text{if } \vdash_{\mathcal{B}} \varphi \text{ then } \vdash_{\mathcal{P}} F(\varphi)$$

and

$$(ii) (b) \quad \text{if } \vdash_{\mathcal{B}} \sim\varphi \text{ then } \vdash_{\mathcal{P}} \sim F(\varphi).$$

That (ii) (a) is true is clear from the fact that \mathcal{B} is consistent and that Prf_β bi-numerates $\text{Prf}_{\mathcal{B}}$ in \mathcal{Q} . On the other hand, suppose $\vdash_{\mathcal{B}} \sim\varphi$; let $\text{Prf}_{\mathcal{B}}[\sim\varphi, n]$. Then

$$\vdash_{\mathcal{P}} \bigwedge_y [\bar{n} < y \rightarrow \bigvee_z (z < y \wedge \text{Prf}_\beta(\sim\bar{\varphi}, z))]$$

and

$$\vdash_{\mathcal{P}} \bigwedge_y [y \leq \bar{n} \rightarrow \sim \text{Prf}_\beta(\bar{\varphi}, y)],$$

by the consistency of \mathcal{B} . Hence

$$\vdash_{\mathcal{P}} \bigwedge_y [\text{Prf}_\beta(\bar{\varphi}, y) \rightarrow \bigvee_z (z < y \wedge \text{Prf}_\beta(\sim\bar{\varphi}, z))],$$

i. e.

$$\vdash_{\mathcal{P}} \sim F(\varphi).$$

Now if 6.4 were to hold for translatability (in the weak sense) as well as $\underline{\leq}$, it would follow from Kreisel's result that: if $\mathcal{B} = \langle B, K' \rangle$ is any consistent axiom system with B recursively enumerable, and π is any RE -formula numerating P in \mathcal{P} then there exists an RE -formula β numerating B in \mathcal{P} for which

$$\vdash_{\mathcal{P}} \text{Con}_\pi \rightarrow \text{Con}_\beta.$$

We shall show now that this is not possible, under the assumption that \mathcal{P} is ω -consistent. Indeed, let π be any fixed RE -formula numerating P in \mathcal{P} . We wish to construct a theory $\mathcal{B} = \langle B, K_0 \rangle$ such that \mathcal{B} is consistent, but such that for any RE -formula β numerating B in \mathcal{P} ,

$$\text{not } \vdash_{\mathcal{P}} \text{Con}_\pi \rightarrow \text{Con}_\beta.$$

In fact, simply take $\mathcal{B} = \mathcal{P} + \{\text{Con}_\pi\}$. \mathcal{B} is consistent under the hypothesis that \mathcal{P} is ω -consistent. Moreover,

$$\text{not } \vdash_{\mathcal{B}} \text{Con}_\beta$$

for any RE -formula β numerating B in \mathcal{P} , according to 5.6. Thus 6.4 fails to hold for translatability in the weaker sense. (Whether it holds for other notions of translatability intermediate between this and $\underline{\leq}$ is an open question.)

Another possibility which might be considered to hold in general is that if $\mathcal{B} \underline{\leq} \mathcal{A}$ then for any numerations α, β of A, B in \mathcal{P} we have

$$\vdash_{\mathcal{P}} \text{Con}_\alpha \rightarrow \text{Con}_\beta;$$

we might further restrict this to RE - or PR -formulas. However, the simplest possible counterexample to this, when $\mathcal{A} = \mathcal{B}$, is obtainable from 7.4. For if $\mathcal{A} = \langle A, K \rangle$ is any consistent, reflexive axiom system with A primitive recursive and $\mathcal{P} \subseteq \mathcal{A}$ (in particular, if $\mathcal{A} = \mathcal{P}$), then we can find PR -formulas α, α' numerating A in \mathcal{P} such that

$$\text{not } \vdash_{\mathcal{A}} \text{Con}_{\alpha'} \rightarrow \text{Con}_\alpha.$$

Without needing to go into any further details about other cases, we can now summarize the various logical possibilities for relative consistency statements as follows. For $\mathcal{A} = \langle A, K \rangle$ and $\mathcal{B} = \langle B, K' \rangle$ axiom systems, A, B recursively enumerable, let

$$(\forall\alpha)(\forall\beta) \vdash_{\mathcal{P}} \text{Con}_\alpha \rightarrow \text{Con}_\beta$$

abbreviate

for all numerations α of A in \mathcal{P} and numerations β of B in \mathcal{P}

$$\vdash_{\mathcal{P}} \text{Con}_\alpha \rightarrow \text{Con}_\beta.$$

Similarly, we shall use $(\exists\alpha)$, $(\exists\beta)$ when we wish to say that there exist numerations α of A in \mathcal{P} or β of B in \mathcal{P} , respectively. For simplicity, assume \mathcal{P} to be ω -consistent.

(1) The statement "If $\mathcal{B} \underline{\leq} \mathcal{A}$ then $(\forall\alpha)(\forall\beta) \vdash_{\mathcal{P}} \text{Con}_\alpha \rightarrow \text{Con}_\beta$ " is false, even when α, β are restricted to the class of PR -formulas. (By 7.4.)

Similarly, we have:

(2) The statements "If $\mathcal{B} \underline{\leq} \mathcal{A}$ then $(\exists\alpha)(\forall\beta) \vdash_{\mathcal{P}} \text{Con}_\alpha \rightarrow \text{Con}_\beta$ " and "If $\mathcal{B} \underline{\leq} \mathcal{A}$ then $(\exists\beta)(\forall\alpha) \vdash_{\mathcal{P}} \text{Con}_\alpha \rightarrow \text{Con}_\beta$ are false, even when α, β are restricted to the class of PR -formulas. (The first from 7.5, by taking $\mathcal{A} = \mathcal{B}$, " β " for " α " and " α " for " α ", and the second from 7.4, by taking $\mathcal{A} = \mathcal{B}$, " α " for " α " and " β " for " α ".)

(3) (a) The statement "If $\mathcal{B} \underline{\leq} \mathcal{A}$ then $(\forall\alpha)(\exists\beta) \vdash_{\mathcal{P}} \text{Con}_\alpha \rightarrow \text{Con}_\beta$ " is true, if α, β are restricted to the class of RE -formulas (By 6.4.)

(3) (b) The corresponding statement, when " $\mathcal{B} \underline{\leq} \mathcal{A}$ " is replaced by " \mathcal{B} is translatable (in the weak sense) into \mathcal{A} " is false. (By Kreisel's result.)

(4) The statement " $(\forall\beta)(\exists\alpha) \vdash_{\mathcal{P}} \text{Con}_\alpha \rightarrow \text{Con}_\beta$ " is true whenever \mathcal{B} is consistent and α, β are restricted to any one of the three classes, PR -formulas, RE -formulas, or bi-numerations. (By Orey's result 7.6.)

Hence

(5) The statement “ $(\exists\beta)(\exists a) \vdash_{\mathcal{A}} \text{Con}_a \rightarrow \text{Con}_\beta$ ” is true whenever \mathcal{B} is consistent and a, β are restricted to RE-formulas, and, for A, B primitive recursive, if restricted to PR-formulas.

These possibilities show rather well how difficult it is to make any interesting general statements about relative consistency proofs. In our work the statement (3) (a) has turned out to be most valuable. As with the problem of selecting particular sentences to express consistency, the moral of these results is to treat with care any general statements about relative consistency. However, they should by no means lessen our interest in the question of relative consistency proofs $\vdash_{\mathcal{A}} \text{Con}_a \rightarrow \text{Con}_b$ for specific axiom systems \mathcal{A}, \mathcal{B} and specific numerations a, β .

8. Formulation of the results for systems which do not contain arithmetic directly

In § 4 we described the arithmetization of the general metamathematics of arbitrary axiom systems $\mathcal{A} = \langle A, K \rangle$. However, the most interesting results of our work, in § 5 and § 6, depended on dealing with systems which contain a significant amount of arithmetic. We wish to show in this section how to formulate the corresponding results for those systems, such as various set theories, which “contain arithmetic” only via an interpretation. In most cases, the proper formulation is clear and the proofs, which we shall omit entirely, are like the previous ones. In fact, for the most part we can still work directly within arithmetic, only passing to the desired system, by interpretation, at the end-points of the arguments.

Let I be a function which is a possible relative interpretation of a system with basic symbols K_0 into a system with basic symbols K . If \mathcal{S} is any set of K_0 -sentences, we shall denote by $I(\mathcal{S})$ the set of K -sentences $I(\varphi)$ for $\varphi \in \mathcal{S}$, together with the sentences $\bigvee_x \delta(x)$ and the existence and unicity statements for function symbols (e. g., if $I(x' \approx y) = \psi(x, y)$, we take in $I(\mathcal{S})$ the statement

$$\bigwedge_x \{ \delta(x) \rightarrow \bigvee_y [\delta(y) \wedge \psi(x, y) \wedge \bigwedge_z (\delta(z) \wedge \psi(x, z) \rightarrow y \approx z)] \} .$$

If $\mathcal{S} = \langle \mathcal{S}, K_0 \rangle$, we put $I(\mathcal{S}) = \langle I(\mathcal{S}), K \rangle$. If $\mathcal{A} = \langle A, K \rangle$ and $\mathcal{S} = \langle \mathcal{S}, K_0 \rangle$ are axiom systems, $\mathcal{S} \stackrel{I}{\approx} \mathcal{A}$ is equivalent to $I(\mathcal{S}) \subseteq \mathcal{A}$. On the other hand, given a set A of K -sentences, we shall denote by $I^{-1}(A)$, the set of all φ such that $\vdash_{\mathcal{A}} I(\varphi)$, and shall write $I^{-1}(\mathcal{A}) = \langle I^{-1}(A), K_0 \rangle$. Then $I^{-1}(\mathcal{A}) \stackrel{I}{\approx} \mathcal{A}$ and hence $I(I^{-1}(\mathcal{A})) \subseteq \mathcal{A}$.

Further, if $I(\mathcal{S}) \subseteq \mathcal{A}$ then $\mathcal{S} \subseteq I^{-1}(\mathcal{A})$. We consider K and I to be fixed in the following.

We begin with a diagonal statement, formulated in arithmetic, which is useful in connection with such interpretations. It follows from the fact that I is primitive recursive.

8.1 (5.1) ⁽²⁵⁾ LEMMA. Let $\varphi \in Fm_{K_0}$, $Fv(\varphi) = \{x\}$. Then we can effectively find $q \in St_{K_0}$ such that

$$\vdash_{\mathcal{Q}} q \leftrightarrow \varphi(\overline{I(q)}) .$$

8.2 (5.2) DEFINITION. For each $a \in Fm_{K_0}$, $Fv(a) = \{x\}$, let μ_a be the sentence associated with $\varphi = \sim \text{Pr}_a$ in 7.1, such that

$$\vdash_{\mathcal{Q}} \mu_a \leftrightarrow \sim \text{Pr}_a(\overline{I(\mu_a)}) .$$

8.3 (5.3) THEOREM. Suppose that $\mathcal{A} = \langle A, K \rangle$ is an axiom system, $I(\mathcal{Q}) \subseteq \mathcal{A}$ and that \mathcal{A} is consistent. Suppose that a numerates A in $I^{-1}(\mathcal{A})$, or in some subtheory of $I^{-1}(\mathcal{A})$. Then

$$\text{not } \vdash_{\mathcal{A}} I(\mu_a) .$$

8.4 (5.6) THEOREM. Suppose that $\mathcal{A} = \langle A, K \rangle$ is an axiom system, $I(\mathcal{P}) \subseteq \mathcal{A}$ and that \mathcal{A} is consistent. Suppose that a is an RE-formula which numerates A in $I^{-1}(\mathcal{A})$, or in some subsystem of $I^{-1}(\mathcal{A})$. Then

$$\vdash_{\mathcal{A}} \text{Con}_a \leftrightarrow \mu_a$$

and hence

$$\text{not } \vdash_{\mathcal{A}} I(\text{Con}_a) .$$

8.5 (5.7) DEFINITION. Let $\mathcal{A} = \langle A, K \rangle$. \mathcal{A} is said to be reflexive (with respect to I) if $I(\mathcal{P}) \subseteq \mathcal{A}$ and if for each finite subsystem $\mathcal{F} = \langle F, K \rangle$ of \mathcal{A}

$$\vdash_{\mathcal{A}} I(\text{Con}_{\mathcal{F}}) .$$

Montague proved in his dissertation [21] that every extension, with the same constants, of Zermelo-Fraenkel set theory is reflexive. He has announced in [22] the improvement of this to all extensions of general set theory.

8.6 (5.9) Suppose that $\mathcal{A} = \langle A, K \rangle$ is a consistent and reflexive axiom system and that A is recursive. Then there is an a^* which bi-numerates A in $I^{-1}(\mathcal{A})$ for which

$$\vdash_{\mathcal{A}} \text{Con}_{a^*} .$$

and hence

$$\vdash_{\mathcal{A}} I(\text{Con}_{a^*}) .$$

8.7 (5.11) Suppose that $\mathcal{A} = \langle A, K \rangle$ is a consistent and reflexive axiom system and that A is primitive recursive. Let a be a PR-formula

⁽²⁵⁾ The numbers in parentheses will refer to the numbers of the corresponding statements in § 5, § 6.

which bi-numerates A in $I^{-1}(\mathcal{A})$. Let $B = A \cup \{\sim I(\text{Con}_a)\}$. Then there is a β^* which bi-numerates B in $I^{-1}(\mathcal{A})$ for which

$$\vdash_{\mathcal{A}} I(\text{Con}_{\beta^*}).$$

Let us turn now to the results of § 6. It would seem in 6.1 that we should consider $\mathcal{S} = \langle S, K' \rangle$ and J where J is a possible interpretation of a K_0 -system into a K' -system, in order to discuss interpretations of a K -system \mathcal{A} in certain extensions of \mathcal{S} . However, we shall indicate in the following how various of the desired consequences of such a theorem can already be derived from the results of § 6 as they stand.

8.8 (6.5) Suppose that $\mathcal{A} = \langle A, K \rangle$ is a consistent axiom system with $I(\mathcal{P}) \subseteq \mathcal{A}$. Suppose that α is an RE-formula which numerates A in \mathcal{P} . Then

$$\mathcal{A} + \{I(\text{Con}_\alpha)\} \not\leq \mathcal{A}.$$

For let $\mathcal{B} = \mathcal{A} + \{I(\text{Con}_\alpha)\}$; as before we can assume \mathcal{B} consistent. Suppose $\mathcal{B} \preceq \mathcal{A}$. Then by 6.4 there is an RE-formula numerating B in \mathcal{P} such that

$$\vdash_{\mathcal{P}} \text{Con}_\alpha \rightarrow \text{Con}_\beta.$$

Hence

$$\vdash_{\mathcal{B}} I(\text{Con}_\beta),$$

which contradicts 8.6.

8.9 (6.6) Suppose that $\mathcal{A} = \langle A, K \rangle$ is a reflexive axiom system with A primitive recursive and that α is a PR-formula which bi-numerates A in $I^{-1}(\mathcal{A})$. Then

$$\mathcal{A} + \{\sim I(\text{Con}_\alpha)\} \preceq \mathcal{A}.$$

The same holds if A is recursively enumerable, $I^{-1}(\mathcal{A})$ is ω -consistent, and α is an RE-formula which numerates A in $I^{-1}(\mathcal{A})$.

For let $\mathcal{B} = \mathcal{A} + \{\sim I(\text{Con}_\alpha)\}$. Again we can assume \mathcal{A} is consistent. Consider first the case that A is primitive recursive. By 8.7 there is a β^* which numerates B in $I^{-1}(\mathcal{A})$ for which

$$\vdash_{I^{-1}(\mathcal{A})} \text{Con}_{\beta^*}.$$

But $\mathcal{P} \subseteq I^{-1}(\mathcal{A})$, so by 6.3, $\mathcal{B} \preceq I^{-1}(\mathcal{A}) + \{\text{Con}_{\beta^*}\}$, i. e. $\mathcal{B} \preceq I^{-1}(\mathcal{A})$. Since $I^{-1}(\mathcal{A}) \preceq \mathcal{A}$, we obtain $\mathcal{B} \preceq \mathcal{A}$. To prove it in the second case we apply 4.13 as at the end of 6.6.

Finally, we mention the generalization of the arithmetical compactness theorem.

8.10 (6.9) THEOREM. Suppose that $\mathcal{A} = \langle A, K \rangle$ and $\mathcal{S} = \langle S, K' \rangle$ are axiom systems with A, S recursively enumerable. Suppose that \mathcal{S} is reflexive. Suppose, finally, that for each finite subtheory $\mathcal{D} \subseteq \mathcal{A}$ we have $\mathcal{D} \preceq \mathcal{S}$. Then

$$\mathcal{A} \preceq \mathcal{S}.$$

For we see as before that for each finite subsystem $\mathcal{D} = \langle D, K \rangle$ of \mathcal{A} there is a finite subsystem $\mathcal{F} = \langle F, K' \rangle$ of \mathcal{S} such that

$$\vdash_{\mathcal{F}} \text{Con}_{\langle \mathcal{D} \rangle} \rightarrow \text{Con}_{\langle \mathcal{D} \rangle}.$$

Hence, if J is the interpretation function giving $J(\mathcal{P}) \subseteq \mathcal{S}$,

$$\vdash_{J^{-1}(\mathcal{S})} \text{Con}_{\langle \mathcal{D} \rangle}.$$

Then, by 6.7 (it is easily seen that $J^{-1}(\mathcal{S})$ is reflexive), we have

$$\mathcal{D} \preceq J^{-1}(\mathcal{S})$$

for each \mathcal{D} . Hence, by 6.9,

$$\mathcal{A} \preceq J^{-1}(\mathcal{S}).$$

But $J^{-1}(\mathcal{S}) \preceq \mathcal{S}$, giving $\mathcal{A} \preceq \mathcal{S}$.

By such styles of reasoning all the other theorems of § 6 obtain analogues for systems in which arithmetic is interpretable. Similarly the results of § 7 could also be brought to a more general formulation, if desired.

References

- [1] W. Craig, *On axiomatizability within a system*, Journ. Symbolic Logic 18 (1953), p. 30-32.
- [2] M. Davis, *Arithmetical problems and recursively enumerable predicates*, Journ. Symbolic Logic 18 (1953), p. 33-41.
- [3] A. Ehrenfeucht, and S. Feferman, *Representability of recursively enumerable sets in formal theories*, Archiv. f. Math. Logik u. Grundlagenforschung 5 (1960), p. 37-41.
- [4] S. Feferman, *Formal consistency proofs and interpretability of theories*, Dissertation, Berkeley 1957.
- [5] — *Sentences of arithmetic which formally express consistency, and Formal consistency proofs and interpretability of theories*, (Abstracts) Journ. Symbolic Logic 22 (1957), p. 106-107.
- [6] — *Formal consistency proofs and interpretability of theories*, Summaries of talks presented at the Summer Institute of Symbolic Logic in 1957 at Cornell University, v. 1, p. 71-77 (mimeographed).
- [7] — *Degrees of unsolvability associated with classes of formalized theories*, Journ. Symbolic Logic 22 (1957), p. 161-175.
- [8] — *Ordinal logics re-examined, and On the strength of ordinal logics*, (Abstracts) Journ. Symbolic Logic 23 (1958), p. 105-106.
- [9] K. Gödel, *Die Vollständigkeit der Axiome des logischen Funktionenkalküls*, Monatshefte f. Math. u. Physik 37 (1930), p. 349-360.
- [10] — *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme*, Monatshefte f. Math. u. Physik 38 (1931), p. 173-198.
- [11] A. Grzegorzcyk, A. Mostowski and C. Ryll-Nardzewski, *The classical and the ω -complete arithmetic*, Journ. Symbolic Logic 23 (1958), p. 188-206.
- [12] G. Hasenjaeger, *Eine Bemerkung zu Henkin's Beweis für die Vollständigkeit des Prädikatenkalküls der ersten Stufe*, Journ. Symbolic Logic 18 (1953), p. 42-48.
- [13] L. Henkin, *The completeness of the first-order functional calculus*, Journ. Symbolic Logic 14 (1949), p. 159-166.

- [14] D. Hilbert, and P. Bernays, *Grundlagen der Mathematik*, Vol. I. Berlin 1934, xii + 471 pp.; Vol. II, Berlin 1939, xii + 498 pp.
- [15] S. C. Kleene, *Introduction to metamathematics*, Amsterdam 1952, x + 550 pp.
- [16] G. Kreisel, *Models, translations and interpretations, in Mathematical interpretation of formal systems*, Studies in Logic Series, Amsterdam 1955, p. 26-50.
- [17] — *On a problem of Henkin's*, Indag. Math. 15 (1953), p. 405-406.
- [18] — *Relative consistency and translatability*, (Abstract) Journ. Symbolic Logic 23 (1958), p. 108-109.
- [19] — and H. Wang, *Some applications of formalized consistency proofs*, Fund. Math. 42 (1955), p. 101-110; II (Addition to the preceding) Fund. Math. 45 (1958), p. 334-335.
- [20] M. H. Löb, *Solution of a problem of Leon Henkin*, Journ. Symbolic Logic 20 (1955), p. 115-118.
- [21] R. Montague, *Contributions to the axiomatic foundations of set theory*, Dissertation, Berkeley 1957.
- [22] — *Non-finite axiomatizability*, Summaries of talks presented at the Summer Institute of Symbolic Logic in 1957 at Cornell University, v. 2, p. 256-259 (mimeographed).
- [23] A. Mostowski, *On models of axiomatic systems*, Fund. Math. 39 (1952), p. 133-158.
- [24] — *Sentences undecidable in formalized arithmetic*, Amsterdam 1952, vii + 117 pp.
- [25] J. Myhill, *Creative sets*, Zeitschrift f. math. Logik u. Grund d. Math. 1 (1955), p. 97-108.
- [26] I. L. Novak, *A construction of models for consistent systems*, Fund. Math. 37 (1950), p. 87-110.
- [27] S. Orey, *Relative interpretations*, (Abstract) Journ. Symbolic Logic 24 (1959), p. 281-282.
- [28] R. M. Robinson, *Arithmetical representation of recursively enumerable sets*, Journ. Symbolic Logic 21 (1956), p. 162-186.
- [29] J. B. Rosser, *Extensions of some theorems of Gödel and Church*, Journ. Symbolic Logic 1 (1936), p. 87-91.
- [30] J. R. Shoenfield, *A relative consistency proof*, Journ. Symbolic Logic 19 (1954), p. 21-28.
- [31] A. Tarski, *Der Wahrheitsbegriff in den formalisierten Sprachen*, Studia Philosophica 1 (1936), p. 261-405.
- [32] A. Tarski with A. Mostowski and R. M. Robinson, *Undecidable theories*, Amsterdam 1953, ix + 98 pp.
- [33] A. M. Turing, *Systems of logic based on ordinals*, Proc. London Math. Soc., ser. 2, 45 (1939), p. 161-228.
- [34] H. Wang, *Arithmetical models for formal systems*, Methodos (1951), p. 217-232.
- [35] — *Arithmetic translations of axiom systems*, Trans. Amer. Math. Soc. 71 (1951), p. 283-293.
- [36] — *Remarks on the comparison of axiom systems*, Proc. Nat. Acad. Sci. 38 (1950), p. 448-453.
- [37] — *Truth definitions and consistency proofs*, Trans. Amer. Math. Soc. 73 (1952), p. 243-275.

STANFORD UNIVERSITY

Reçu par la Rédaction le 10. 12. 1959

Algebras which are independently generated by every n elements

by

S. Świerczkowski (Wrocław)

1. Preliminaries and results

By an algebra \mathcal{A} we mean a pair (A, F) where A is a set and F is a family of functions of finitely many variables defined on A and A -valued. F is called the *class of fundamental operations*. The *class of algebraic operations* is, by definition, the class of operations A generated by F , i. e. the smallest class A such that A contains F , all identity operations belong to A and A is closed with respect to composition. The subclass of all algebraic operations of n variables will be denoted by $A^{(n)}$. The above definitions are given in a more detailed form in [3]; we use here the same notation.

Following E. Marczewski [3] we say that $N \subset A$ is a *set of independent elements* if, for each sequence of n different elements $a_1, \dots, a_n \in N$ and for each pair of operations $f, g \in A^{(n)}$, the equality

$$f(a_1, \dots, a_n) = g(a_1, \dots, a_n)$$

implies that f and g are identical in \mathcal{A} .

We shall call the identity operations also *trivial operations*. More exactly: An operation $f(x_1, \dots, x_k)$ is called *trivial* if, for a certain $l \leq k$, we have $f(x_1, \dots, x_k) = x_l$ for all values of x_1, \dots, x_k . If all algebraic operations are trivial then the algebra will be called trivial. For $A = \{a_1, \dots, a_n\}$ and $F = \{f\}$ we shall write $(a_1, \dots, a_n; f)$ instead of (A, F) . Two algebras, (A, F_1) and (A, F_2) , having the same class of all algebraic operations will be treated here as identical.

We say that a set $B \subset A$ *generates* \mathcal{A} if each $x \in A$ is the result of an algebraic operation applied to some elements in B . Let \bar{S} denote the cardinal of the set S . We then say that *the algebra is independently generated by every n elements* if each set $B \subset A$ satisfying $\bar{B} = n$ is a set of independent elements and B generates \mathcal{A} . In this paper we show some properties of those algebras. The results were announced in paper [4].