ON UNDECIDABLE PROPOSITIONS OF FORMAL MATHEMATICAL SYSTEMS

Notes on lectures by

KURT GÖDEL

February - May, 1934

(Notes by S.C.Kleene and J.B.Rosser)

Institute for Advanced Study

Princeton, N.J.

ON UNDECIDABLE PROPOSITIONS OF FORMAL MATHEMATICAL SYSTEMS

Notes on lectures by Kurt Gödel

1. INTRODUCTION

A formal mathematical system is a system of symbols together with rules for employing them. The individual symbols are called undefined terms. Formulas are finite sequences of the undefined terms. There shall be defined a class of formulas called meaningful formulas, and a class of meaningful formulas called axioms. There may be a finite or infinite number of axioms. Further, there shall be specified a list of rules, called rules of inference; if such a rule be called R, it defines the relation of immediate consequence by R between a set of meaningful formulas $M_1, \ldots, M_k$ called the premises, and a meaningful formula N called the conclusion (ordinarily k = 1 or 2). We require that the rules of inference, and the definitions of meaningful formulas and axioms, be constructive; that is, for each rule of inference there shall be a finite procedure for determining whether a given formula B is an immediate consequence (by that rule) of given formulas $A_1, \ldots, A_n$, and there shall be a finite procedure for determining whether a given formula A is a meaningful formula or an axiom.

A formula N shall be called an immediate consequence of $M_1, \ldots, M_n$ if N is an immediate consequence of $M_1, \ldots, M_n$ by any one of the rules of inference. A finite sequence of formulas shall be a proof (specifically, a proof of the last formula of the sequence) if each formula of the sequence is either an axiom, or an immediate consequence of one or more of the preceding formulas. A formula is provable if a proof of it exists. Let the symbol $\sim$ be one of the undefined terms, and suppose it to express negation. Then the formal system shall be said to be complete, if for every meaningful formula A either A or $\sim$A is provable. We shall prove later that (under conditions to be stated) a system in which all propositions of arithmetic can be expressed as meaningful formulas is not complete.

2. RECURSIVE FUNCTIONS AND RELATIONS

Now we turn to some considerations which for the present have nothing to do with a formal system.

Small Roman letters x, y, z, ... will denote arbitrary natural numbers (i.e. non-negative integers); and German letters will be used in abbreviation for finite sequences of the former, e.g. $\mathfrak{x}$ for $x_1, \ldots, x_n$; $\mathfrak{y}$ for $y_1, \ldots, y_m$. Greek letters $\phi, \psi, \chi, \ldots$ will represent functions of one or more natural numbers whose values are natural numbers. Roman capitals R, S, T, ... will stand for classes of, or relations among, natural numbers. R(x) shall stand for the proposition that x is in the class R, and $S(x_1, \ldots, x_n)$ for the proposition that $x_1, \ldots, x_n$ stand in the relation S. Classes may be considered as relations with only one term, and relations as classes of ordered n-tuples. There shall correspond to each class or relation R a <u>representing function</u> ϕ such that $\phi(x_1, \ldots, x_n) = 0$ if $R(x_1, \ldots, x_n)$ and $\phi(x_1, \ldots, x_n) = 1$ if $\sim R(x_1, \ldots, x_n)$.

We use the following notations as abbreviations (p, q are to be replaced by any propositions): $(x)[A(x)]$ (for every natural number x, A(x)), $(Ex)[A(x)]$ (there exists a natural number x such that A(x)), $\epsilon x[A(x)]$ (the least natural number x such that A(x) if $(Ex)[A(x)]$; otherwise 0), $\sim p$ (not p), $p \vee q$ (p or q), $p \,\&\, q$ (p and q), $p \rightarrow q$ (p implies q, i.e. $(\sim p) \vee q$), $p \equiv q$ (p is equivalent to q, i.e. $(p \rightarrow q) \,\&\, (q \rightarrow p)$).

The function $\phi(x_1, \ldots, x_n)$ shall be <u>compound</u> with respect to $\psi(x_1, \ldots, x_n)$ and $\chi_i(x_1, \ldots, x_n)$ $(i = 1, \ldots, n)$ if, for all natural numbers $x_1, \ldots, x_n$,

(1) $$\phi(x_1, \ldots, x_n) = \psi(\chi_1(x_1, \ldots, x_n), \ldots, \chi_n(x_1, \ldots, x_n)).$$

$\phi(x_1, \ldots, x_n)$ shall be said to be <u>recursive</u> with respect to $\psi(x_1, \ldots, x_{n-1})$ and $\chi(x_1, \ldots, x_{n+1})$ if, for all natural numbers $k, x_2, \ldots, x_n$,

(2) $$\phi(0, x_2, \ldots, x_n) = \psi(x_2, \ldots, x_n)$$
$$\phi(k + 1, x_2, \ldots, x_n) = \chi(k, \phi(k, x_2, \ldots, x_n), x_2, \ldots, x_n).$$

In both (1) and (2), we allow the omission of each of the variables in any (or all) of its occurrences on the right side (e.g. $\phi(x, y) = \psi(\chi_1(x), \chi_2(x, y))$ is permitted under (1)). We define the class of recursive functions to be the totality of functions which can be generated by substitution, according to the scheme (1), and recursion, according to the scheme (2), from the successor function $x + 1$, constant functions $f(x_1, \ldots, x_n) = c$, and identity functions $U^n_j(x_1, \ldots, x_n) = x_j$ $(1 \leq j \leq n)$. In other words, a function ϕ shall be recursive if there exists a finite sequence of functions $\phi_1, \ldots, \phi_n$ which terminates with ϕ such that each function of the sequence is either the successor function $x + 1$ or a constant function $f(x_1, \ldots, x_n) = c$, or an identity function $U^n_j(x_1, \ldots, x_n) = x_j$, or is compound with respect to preceding functions, or is recursive with respect to preceding functions. A relation R shall be recursive if the representing function is recursive.

Recursive functions have the important property that, for each given set of values of the arguments, the value of the function can be computed by a finite procedure.[1] Similarly, recursive relations (classes) are decidable in the sense

[1] The converse seems to be true, if, besides recursions according to the scheme (2), recursions of other forms (e.g., with respect to two variables simultaneously) are admitted. This cannot be proved, since the notion of finite computation is not defined, but it serves as a heuristic principle.

that, for each given set of natural numbers, it can be determined by a finite procedure whether the relation holds or does not hold (the number belongs to the class or not), since the representing function is computable.

The functions $x + y$, xy, x^y and $x!$ are clearly recursive. Hence $\phi(\mathfrak{x}) + \psi(\mathfrak{y})$, $\phi(\mathfrak{x})\psi(\mathfrak{y})$, $\phi(\mathfrak{x})^{\psi(\mathfrak{y})}$, and $\phi(\mathfrak{x})!$ are recursive, if $\phi(\mathfrak{x})$ and $\psi(\mathfrak{y})$ are.

I. *If the relations* $R(\mathfrak{x})$ *and* $S(\mathfrak{y})$ *are recursive, then* $\sim R(\mathfrak{x})$, $R(\mathfrak{x}) \vee S(\mathfrak{y})$, $R(\mathfrak{x}) \,\&\, S(\mathfrak{y})$, $R(\mathfrak{x}) \rightarrow S(\mathfrak{y})$, $R(\mathfrak{x}) \equiv S(\mathfrak{y})$ *are recursive.*

By hypothesis, the representing functions $\rho(\mathfrak{x})$ and $\sigma(\mathfrak{y})$ of R and S,

respectively, are recursive. If

$$\alpha(0) = 1, \qquad \alpha(k + 1) = 0,$$

then $\alpha(x)$ and hence $\alpha(\rho(\mathfrak{x}))$ are recursive. But, since $\alpha(\rho(\mathfrak{x}))$ is 1 or 0 according as $\rho(\mathfrak{x})$ is 0 or 1, $\alpha(\rho(\mathfrak{x}))$ is the representing function of $\sim R(\mathfrak{x})$. Thus $\sim R(\mathfrak{x})$ is recursive. If $\beta(0, x) = 0$, $\beta(k + 1, x) = \alpha(\alpha(x))$, then

$$\beta(0, x) = \beta(x, 0) = 0 \quad \text{and} \quad \beta(x, y) = 1 \text{ when } x, y > 0.$$

Hence $\beta(\rho(\mathfrak{x}), \sigma(\mathfrak{y}))$, which is recursive, represents $R(\mathfrak{x}) \vee S(\mathfrak{y})$; that is, $R(\mathfrak{x}) \vee S(\mathfrak{y})$ is recursive. Since $R(\mathfrak{x}) \,\&\, S(\mathfrak{y}) \equiv \sim(\sim R(\mathfrak{x}) \vee \sim S(\mathfrak{y}))$, it follows that $R(\mathfrak{x}) \,\&\, S(\mathfrak{y})$ is recursive. Similarly $R(\mathfrak{x}) \to S(\mathfrak{y})$, $R(\mathfrak{x}) \equiv S(\mathfrak{y})$, and all other relations definable from $R(\mathfrak{x})$ and $S(\mathfrak{y})$ by use of $\sim$ and $\vee$, are recursive.

II. *If the functions $\phi(\mathfrak{x}), \psi(\mathfrak{y})$ are recursive, then the relations $\phi(\mathfrak{x}) = \psi(\mathfrak{y})$, $\phi(\mathfrak{x}) < \psi(\mathfrak{y})$, $\phi(\mathfrak{x}) \leq \psi(\mathfrak{y})$ are recursive.*

Let

$$\delta(0) = 0, \quad \delta(k + 1) = k,$$

and $x \dot{-} 0 = x$, $x \dot{-} (k + 1) = \delta(x \dot{-} k)$. Then

$$x \dot{-} y = x - y \text{ if } x \geq y \quad \text{and} \quad x \dot{-} y = 0 \text{ if } x \leq y.$$

Hence $\alpha(y \dot{-} x)$ is a representing function for $x < y$, and $\alpha(\psi(\mathfrak{y}) \dot{-} \phi(\mathfrak{x}))$ for $\phi(\mathfrak{x}) < \psi(\mathfrak{y})$. Thus $\phi(\mathfrak{x}) < \psi(\mathfrak{y})$ is recursive. $\phi(\mathfrak{x}) = \psi(\mathfrak{y})$ and $\phi(\mathfrak{x}) \leq \psi(\mathfrak{y})$ are likewise recursive, as may be seen directly, or inferred from the theorem for $\phi(\mathfrak{x}) < \psi(\mathfrak{y})$ by use of I.

III. *If the function $\phi(\mathfrak{x})$ and the relation $R(x, \mathfrak{y})$ are recursive, then the relations S, T, where*

$$S(\mathfrak{x}, \mathfrak{y}) \equiv (Ex)[x \leq \phi(\mathfrak{x}) \,\&\, R(x, \mathfrak{y})],$$
$$T(\mathfrak{x}, \mathfrak{y}) \equiv (x)[x \leq \phi(\mathfrak{x}) \to R(x, \mathfrak{y})],$$

and the function ψ, where

$$\psi(\mathfrak{x}, \mathfrak{y}) = \epsilon x[x \leq \phi(\mathfrak{x}) \,\&\, R(x, \mathfrak{y})],$$

are recursive.

Let the representing function of $R(x, \mathfrak{y})$ be $\rho(x, \mathfrak{y})$. Let

$\pi(0, 2g) = \rho(0, 2g)$ and $\pi(k + 1, 2g) = \pi(k, 2g) \cdot \rho(k + 1, 2g)$. Then $\pi(x, 2g) = \rho(0, 2g)\,\rho(1, 2g) \ldots \rho(x, 2g)$. Hence $\pi(x, 2g)$ is 0 or 1 according as some or none of $\rho(0, 2g), \ldots, \rho(x, 2g)$ are 0; that is, according as there do or do not exist natural numbers $n \leq x$ for which $R(n, 2g)$ holds. Hence $\pi(\phi(q), 2g)$, which is recursive, represents $(Ex)[x \leq \phi(q) \,\&\, R(x, 2g)]$. Thus $(Ex)[x \leq \phi(q) \,\&\, R(x, 2g)]$ is a recursive relation. It follows from this result and I that $(x)[x \leq \phi(q) \rightarrow R(x, 2g)]$ is recursive, since $(x)[x \leq \phi(q) \rightarrow R(x, 2g)] \equiv \sim[(Ex)[x \leq \phi(q) \,\&\, \sim R(x, 2g)]]$. Let $\mu(0, 2g) = 0$ and $\mu(k + 1, 2g) = (k + 1)[\pi(k, 2g) \dot{-} \pi(k + 1, 2g)] + \mu(k, 2g)[\alpha(\pi(k, 2g) \dot{-} \pi(k + 1, 2g))]$. Since $1 \geq \pi(k, 2g) \geq \pi(k + 1, 2g) \geq 0$, $\mu(k + 1, 2g) = k + 1$ if $\pi(k, 2g) = 1$ and $\pi(k + 1, 2g) = 0$, and otherwise $\mu(k + 1, 2g) = \mu(k, 2g)$. Both $\pi(k, 2g) = 1$ and $\pi(k + 1, 2g) = 0$ hold only when $\sim R(0, 2g), \ldots, \sim R(k, 2g)$ and $R(k + 1, 2g)$; that is, when $k + 1$ is the least value x' of x such that $R(x, 2g)$. Hence if such an x' exists and is > 0, $\mu(0, 2g) = \ldots = \mu(x' - 1, 2g) = 0$ and $\mu(x, 2g) = x'$ for all $x \geq x'$. If $x' = 0$, or x' does not exist, all $\mu(x, 2g)$ are 0. Hence $\mu(\phi(q), 2g)$ is the least $x \leq \phi(q)$ such that $R(x, 2g)$, if such exists, otherwise 0; i.e. $\mu(\phi(q), 2g) = \epsilon x[x \leq \phi(q) \,\&\, R(x, 2g)]$.

3. A FORMAL SYSTEM

We now describe in some detail a formal system which will serve as an example for what follows. While a formal system consists only of symbols and mechanical rules relating to them, the meaning which we attach to the symbols is a leading principle in the setting up of the system.

We shall depend on the theory of types as our means for avoiding paradox. Accordingly, we exclude the use of variables running over all objects, and use different kinds of variables for different domains. Specifically, p, q, r, ... shall be variables for propositions. Then there shall be variables of successive types as follows:

x, y, z, ... for natural numbers,

f, g, h, ... for functions (of one variable) whose domain and values are natural numbers,

F, G, H, ... for functions (of one variable) whose domain and values are functions f, g, h, ...,

and so on.[2] Different formal systems are determined according to how many of

[2] Functions for several variables need not be provided separately, since n-tuples of objects of each of these types can be mapped one-to-one on single objects of the same type.

Variables for classes and relations are unnecessary, since we can use, instead of the classes and relations, their representing functions.

these types of variables are used. We shall restrict ourselves to the first two types; that is, we shall use variables of the three sorts p, q, r, ...; x, y, z, ...; f, g, h, ... We assume that a denumerably infinite number of each are included among the undefined terms (as may be secured, for example, by the use of letters with numerical subscripts).

The undefined terms, in addition to variables, shall be the following: 0 (the number 0), N (N(x) denotes the next greater number than x, i.e. the successor of x), $\sim$, $\vee$, $\&$, $\rightarrow$, $\equiv$, Π ($\Pi x(F(x))$ means "F(x) is true for all natural numbers x", and may be regarded as the logical product of F(x) over all x), Σ ($\Sigma x(F(x))$ means "there is at least one natural number x such that F(x) is true", and may be regarded as the logical sum of F(x) over all x), $\in$, = (equals), (,) (f(x) is the value of f for the argument x, (and) being then interpreted as symbols for the operation of application of a function to an argument. Parentheses are also used as signs of inclusion, as in $\Pi x(A)$, $(A) \rightarrow (B)$, etc.).[3]

[3] $\sim$, $\vee$, $\&$, $\rightarrow$, $\equiv$, and $\in$ have the significances assigned to them in § 2. $\Pi x(A)$ and $\Sigma x(A)$, when A does not involve x, mean the same as A. The fact that the logical notions among our undefined terms are not independent does not matter for our purpose.

Next, the class of meaningful formulas must be defined. To do this we describe two classes of formulas which have significance - formulas which denote numbers, and formulas which denote propositions. The first comprises numerical symbols or expressions representing numbers (as 0, N(0), ...) together with functional expressions or expressions which become numerical expressions when numerical expressions are substituted in a suitable manner for variables which occur in them (as $\in x[y = N(x)]$). The second comprises propositions (e.g. $\Pi x [\sim(0 = N(x))]$), together with propositional functions or expressions which become propositions when numerical expressions are substituted in a suitable manner for variables which occur in them (e.g. $\Sigma x[y = N(x)]$). The exact definitions we give by complete induction, thus:

1. 0 and x, y, z, ... (variables for numbers) are expressions of the Ist kind, and p, q, r, ... (variables for propositions) are expressions of the IInd kind.

2. If A and B are expressions of the Ist kind, then A = B is an expression of the IInd kind.

3. If A exp. I, then N(A) exp. I.

4. If A exp. I, and f is a variable for a function, then f(A) exp. I.

5. If A and B exp. II, then $\sim$(A), (A) $\vee$ (B), (A) & (B), (A) $\rightarrow$ (B) and (A) $\equiv$ (B) exp. II.

6. If A exp. II, and x is a variable for a number, then Πx(A) and Σx(A) exp. II, and $\in x$(A) exp. I.

7. If A exp. II, and f is a variable for a function, then Πf(A) and Σf(A) exp. II.

8. If A exp. II, and p is a variable for a proposition, then Πp(A) and Σp(A) exp. II.

9. The class of expressions of the Ist (IInd) kind shall be the least class satisfying 1-8.

A formula shall be meaningful, if it is either an expression of the Ist kind or an expression of the IInd kind.

The occurrences of variables in a meaningful expression can be classified as free and bound in the following manner: There corresponds to each occurrence of Π in a meaningful expression A a unique part of A, beginning with the occurrence of

Π, of the form $\Pi t(B)$, where t is a variable and B is meaningful. This part of A will be called the scope of the given occurrence of Π in A. Similarly we define the scope of an occurrence of Σ or $\in$ in A. A given occurrence of the variable t in A shall be bound or free according as it is or is not in the scope of a Π, Σ or $\in$ followed by t.

In the above definitions of functional expressions and propositional functions, the substitutions which are meant are substitutions for the free occurrences of variables. (y is free and x is bound in $\in x[y = N(x)]$ and $\Sigma x[y = N(x)]$.)

We use Subst (A^t_G) to denote the expression obtained from A by substituting G for each occurrence of t in A as a free variable.[4]

---

[4] Subst by itself is not a formula of our system.

We may use F(t) to represent a meaningful formula in which t occurs as a free variable,[5] and F(A) to denote Subst $(F(t)^t_A)$.

---

[5] Then F by itself does not represent a formula.

If A is a meaningful formula, and f a variable for a function, then the occurrences of f in A as a free symbol are as the first symbol of parts of A of the form $f(U)$. We may list these parts as $f(U_1)$, ..., $f(U_n)$ in such an order that, if U_j contains $f(U_i)$, then $i < j$. Let G(x) be a meaningful formula in which x occurs as a free variable. Let A', $f(U'_2)$, ..., $f(U'_n)$ be obtained from A, $f(U_2)$, ..., $f(U_n)$ by substituting $G(U_1)$ for $f(U_1)$;[6] then let A", $f(U''_3)$, ...,

---

[6] More explicitly, let A' be the expression obtained from A by substituting $G(U_1)$ for the part $f(U_1)$, and let $f(U'_2)$, ..., $f(U'_n)$ be the parts of A' into which the parts $f(U_2)$, ..., $f(U_n)$ of A are transformed by the substitution.

$f(U''_n)$ be obtained from A', $f(U'_3)$, ..., $f(U'_n)$ by substituting $G(U'_2)$ for $f(U'_2)$, and so on. We shall denote by Subst' $(A^f_{G(x)})$ the expression $A^{(n)}$.

An expression A in which the distinct free variables $t_1, \ldots, t_n$ occur, shall mean the same as $\Pi t_1(\ldots(\Pi t_n(A)) \ldots)$.

The axioms shall be the following formulas (A1 - C2):[7]

---

[7] In writing down these axioms and other meaningful formulas we employ the usual conventions concerning the omission of parentheses.
All abbreviation of formulas is to be regarded as extraneous to the formal system; and each statement about a formula of the system refers to its unabbreviated form.

A. Axioms concerning the notions of the calculus of propositions

1. $(p \rightarrow q) \rightarrow ((q \rightarrow r) \rightarrow (p \rightarrow r))$,
2. $((\sim p) \rightarrow p) \rightarrow p$,
3. $p \rightarrow ((\sim p) \rightarrow q)$.
4. $p \,\&\, q \;.\equiv.\; \sim[(\sim p) \vee (\sim q)]$.
5. $p \vee q \;.\equiv.\; (\sim p) \rightarrow q$.
6. $p \equiv q \;.\equiv.\; (p \rightarrow q) \,\&\, (q \rightarrow p)$,
7. $p \equiv q \;.\rightarrow.\; p \rightarrow q$.
8. $p \equiv q \;.\rightarrow.\; q \rightarrow p$.

To give the theory of this group of axioms would require a study of the theory of the calculus of propositions.

B. Axioms concerning the notion of identity

1. $x = x$,
2. $x = y \;.\equiv.\; f(x) = f(y)$,
3. $(x = y) \,\&\, (y = z) \;.\rightarrow.\; z = x$.

C. Axioms which correspond to certain of Peano's axioms for the natural numbers

1. $\sim(0 = N(x))$
2. $N(x) = N(y) \;.\rightarrow.\; x = y$

To complete the definition of the formal system under consideration, it remains to list the rules of procedure. Each rule is to be interpreted as a statement of the conditions under which a formula N shall be an immediate consequence by

that rule of the meaningful formula(s) $M(M_1, M_2)$.[8]

---

[8] N also will be meaningful, whenever the conditions are realized.
Rule 1, for example, can be written more explicitly thus: N shall be an immediate consequence by Rule 1 of meaningful formulas M_1 and M_2, if and only if there exist formulas A and B such that M_1 is $(A) \rightarrow (B)$, M_2 is A, and N is B.

---

1. If $(A) \rightarrow (B)$ and (A), then B.

2. Suppose that A is meaningful,[9] that t is a variable, and that t does not occur in A. a. If $(A) \rightarrow (B)$, then $(A) \rightarrow (\Pi t(B))$. b. If $(A) \rightarrow (\Pi t(B))$, then $(A) \rightarrow (B)$.

---

[9] This condition ensures that, when $(A) \rightarrow (B)$ is a meaningful formula, the occurrence of $\rightarrow$ which separates (A) from (B) in $(A) \rightarrow (B)$ should be the last occurrence of $\rightarrow$ introduced in the construction of $(A) \rightarrow (B)$ according to the definition of meaningful formula. (We may say then that the main operation of $(A) \rightarrow (B)$ is an implication, whose 1st and 2nd terms are A and B, respectively.) It excludes such possibilities as that A be p when $(A) \rightarrow (B)$ is Axiom A1.

---

3. Suppose that t does not occur in B. a. If $(A) \rightarrow (B)$, then $(\Sigma t(A)) \rightarrow (B)$. b. If $(\Sigma t(A)) \rightarrow (B)$, then $(A) \rightarrow (B)$.

4a. Suppose that x is a variable for a number, that A contains x as a free variable, that G is an expression of the first kind, and that no free variable of G is bound in F. If A, then Subst (A^x_G).

4b. Suppose that f is a variable for a function, that A contains f as a free variable, that x is a variable for a number, that G(x) is an expression of the Ist kind in which x occurs as a free variable, and that no free variable of G(x) is bound in A. If A, then Subst $(A^f_{G(x)})$.

4c. Suppose that p is a variable for a proposition, that A contains p as a free variable, that P is an expression of the IInd kind, and that no free variable of P is bound in A. If A, then Subst (A^p_P).

4d. Suppose that x is a variable for a number, and that F(x) is meaningful and contains x as a free variable. If $(A) \rightarrow (F(x))$, then $(A) \rightarrow (F(\in x[F(x)]))$.

5. Suppose that x is a variable for a number, and that F(x) is a meaningful formula in which x occurs as a free variable. If F(0) and $(F(x)) \rightarrow (F(N(x)))$, then F(x).

6. Suppose that s and t are variables of the same kind, that s does not occur in A as a free variable, and that t does not occur in A. Let A' denote the result of substituting t for s throughout A. If A, then A'.

One process used in mathematical proof is not represented in this system, namely the definition and introduction of new symbols. However, this process is not essential, but merely a matter of abbreviation.

4. A REPRESENTATION OF THE SYSTEM BY A SYSTEM OF POSITIVE INTEGERS

For the considerations which follow, the meaning of the symbols is immaterial, and it is desirable that they be forgotten. Notions which relate to the system considered purely formally may be called metamathematical.

The undefined terms (hence the formulas and proofs) are countable, and hence a representation of the system by a system of positive integers can be constructed, as we shall now do.

We order the numbers 1-13 to symbols thus:

0	N	=	$\sim$	$\vee$	&	$\rightarrow$	$\equiv$	Π	Σ	$\in$	(	)
1	2	3	4	5	6	7	8	9	10	11	12	13,

the integers > 13 and $\equiv 0 \pmod 3$ to the variables for propositions, the integers > 13, $\equiv 1 \pmod 3$ to the variables for numbers, and the integers > 13, $\equiv 2 \pmod 3$ to the variables for functions. Thus a one-to-one correspondence is established between the undefined terms and the positive integers.

We order single integers to finite sequences of positive integers by means of the scheme

$$k_1, \ldots, k_n \text{ corresponds to } 2^{k_1} \cdot 3^{k_2} \cdot 5^{k_3} \ldots p_n^{k_n},$$

where p_i is the i-th prime number (in order of magnitude). A formula is a finite sequence of undefined terms, and a proof a finite sequence of formulas. To each formula we order the integer which corresponds to the sequence of the integers ordered to its symbols; and to each proof we order the integer which corresponds to the sequence of the integers which are then ordered to its member formulas. Then a one-to-one correspondence is determined between formulas (proofs) and a subset of the positive integers.

We may now define various metamathematical classes and relations of positive integers, including one corresponding to each class and relation of formulas. x shall be an f number (Bf number), if there is a formula (proof) to which x corres-

ponds.[10] The relation x, y U z between numbers shall mean that x, y and z are f

[10] $\mathfrak{B}$ for "Beweis". Below occur U for "unmittelbar Folge", Gl for "Glied", $\mathfrak{E}$ for "einklammern".

numbers, and the formula which z represents is an immediate consequence of the formulas which x and y represent. x $\mathfrak{B}$ y shall mean that x is a $\mathfrak{B}$ number and y an f number, and the proof which x represents is a proof of the formula which y represents. Also there are metamathematical functions of integers such as the following: Neg (x) = the number representing $\sim$(X), if x represents the formula X; and = 0 if x is not an f number. $Sb(x^y_z)$ = the number which represents the result Subst (F^t_G) of substituting G for the free occurrences of t in F, if x, z represent the formulas F, G, respectively, and y the variable t; and = 0 otherwise. These relations and functions, which we have defined indirectly, by using the correspondence between formulas and numbers, are constructive. Hence it is not surprising to find that they are recursive. We shall show this for some of the more important of them, by defining them directly, from relations and functions previously known to be recursive, by methods shown in § 2 to generate recursive relations and functions out of recursive relations and functions.[11]

[11] We use formal notations (including those explained in § 2) in the following definitions for the purpose of abbreviating the discussion. These formal notations must not be confused with the formulas of the formal mathematical system under consideration.

1. $x \mid y \equiv (Ez)[z \leq x \,\&\, x = yz]$.

"x | y" means "x is divisible by y". (yz is recursive. Hence, by II of § 2, x = yz is recursive. It follows by III that x | y is recursive. $z \leq x$ is inserted in the definition to make it clear that III applies and could be omitted without changing the meaning.)

2. Prime (x) $\equiv x > 1 \,\&\, \sim(Ez)[z \leq x \,\&\, \sim(z = 1) \,\&\, \sim(z = x) \,\&\, x \mid z]$.

"x is a prime number".

3. $\mathrm{Pr}\ (0) = 0$

$$\mathrm{Pr}\ (n+1) = \in y\big[y \leq \{\mathrm{Pr}\ (n)\}! + 1 \ \&\ \mathrm{Prime}\ (y) \ \&\ y > \mathrm{Pr}\ (n)\big].$$

Pr (n) is the n-th prime number (in order of magnitude).

4. $n\,G\!\!\not\!\chi x = \in y\big[y \leq x \ \&\ x \mid \{\mathrm{Pr}\ (n)\}^{y} \ \&\ \sim(x \mid \{\mathrm{Pr}\ (n)\}^{y+1})\big]$.

$n\,G\!\!\not\!\chi x$ is the n-th member of the sequence of positive integers which x represents (i.e. $n\,G\!\!\not\!\chi x$ is k_n if $x = 2^{k_1} \cdot 3^{k_2} \cdot \ldots \cdot p_n^{k_n} \cdot \ldots \cdot p_\chi^{k_\chi}$).

5. $L(x) = \in y\big[y \leq x \ \&\ (y+1)\,G\!\!\not\!\chi x = 0\big]$.

L(x) is the number of members in the sequence represented by x (if x represents a sequence of positive integers).

6. $x \oplus y = \in z\Big\{z \leq \big[\mathrm{Pr}\ (L(x) + L(y))\big]^{x+y} \ \&\ (n)\big[n \leq L(x) \longrightarrow n\,G\!\!\not\!\chi z = n\,G\!\!\not\!\chi x\big] \ \&\ (n)\big[0 < n \leq L(y) \longrightarrow (n + L(x))\,G\!\!\not\!\chi z = n\,G\!\!\not\!\chi y\big]\Big\}$.

$\oplus$ represents the operation of joining one finite sequence to another (i.e., if $x = 2^{k_1} \ldots p_r^{k_r}$ and $y = 2^{\chi_1} \ldots p_s^{\chi_s}$, then $x \oplus y = 2^{k_1} \ldots p_r^{k_r} p_{r+1}^{\chi_1} \ldots p_{r+s}^{\chi_s}$).

Note that the number of the sequence consisting of the single number x is 2^x.

7. $\mathcal{k}\,(x) = 2^{12} \oplus x \oplus 2^{13}$.

If x represents the formula A, $\mathcal{k}\,(x)$ represents (A) (for then the sequence of the numbers ordered to the symbols of (A) is $12, k_1, \ldots, k_n, 13$, where $k_1, \ldots, k_n$ is the sequence of the numbers ordered to the symbols of A).

8. $\mathrm{Neg}\ (x) = 2^4 \oplus \mathcal{k}\,(x)$.

If x represents the formula A, Neg (x) represents $\sim(A)$.

9. $\mathrm{Imp}\ (x, y) = \mathcal{k}\,(x) \oplus 2^7 \oplus \mathcal{k}\,(y)$.

If x, y represent the formulas A, B respectively, then Imp (x, y) represents $(A) \longrightarrow (B)$.

10. $u\mathrm{Gen}\ x = 2^9 \oplus 2^u \oplus \mathcal{k}\,(x)$

If u represents the variable t, and x the formula A, then u Gen x represents $\Pi t(A)$.

Similarly for $\sum t(A)$ and $\in x(A)$.

11. $x \equiv y \pmod{n} \;.\equiv.\; (Ez)[z \leq x + y \;\&\; (x = y + zn \vee y = x + zn)]$.

$x \equiv y \pmod{n}$ has the usual significance.

$t > 13$ expresses "t represents a variable". Also, using 11, recursive classes $Var_p(t)$, $Var_x(t)$, $Var_f(t)$ can be defined to express "t represents a variable for a proposition", "t represents a variable for a number", "t represents a variable for a function", respectively.

Recursive classes $M_I(x)$, $M_{II}(x)$, $M(x)$ expressing "x represents an exp, I", "x represents an exp. II", "x represents a meaningful formula", respectively, recursive relations corresponding to the relations "t occurs in A as a free (bound) variable", and recursive functions corresponding to the operations of substitution used in the rules of inference, can be defined.[12]

---

[12] For the details of the definition of classes, relations and functions of these sorts, relating to a formal system similar to the one under consideration, see K. Gödel, Über die formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I, Monatshefte für Math. u. Physik, Vol. XXXVIII (1931), pp. 173-198. Specifically, see the definitions 1-31, pp. 182-184.

---

12. $x, y U_1 z \equiv x = Imp\ (y,\ z) \;\&\; M(z)$.[13]

---

[13] M(z) is added to insure that z be an f number (i.e. represent a formula.)

"z represents a formula which is an immediate consequence by Rule 1 of the formulas represented by x, y" (if x, y represent meaningful formulas).

13. $xU_{2a}z \equiv (Et, v, w)[t, v, w \leq z \;\&\; M(v) \;\&\; M(w) \;\&\; x = Imp\ (v, w) \;\&\; z =$
$= Imp(v,\ t\ Gen\ w) \;\&\; t > 13 \;\&\; \sim(Ek)[k \leq L(v) \;\&\; k\,G\,v = t]]$.[14]

---

[14] $E(t, v, w)[t, v, w \leq z \;\&\; \ldots]$ stands for $(Et)[(Ev)[(Ew)[t \leq z \;\&\; v \leq z \;\&\; w \leq z \;\&\; \cdots]$
Similarly, $(x, y, z)[A]$ stands for $(x)[(y)[(z)[A]]]$.

"z represents a formula which is an immediate consequence by Rule 2a of the formula represented by x" (if x represents a meaningful formula).

Similarly for each of the other rules of inference.

14. $x, yUz \equiv x, yU_1 z \vee xU_{2a} z \vee \ldots \vee xU_6 z.$

"z represents a formula which is an immediate consequence of the formula(s) represented by x (x and y)" (if x and y represent formulas).

Each axiom is represented by a number. Let the numbers corresponding to the axioms be $\alpha_1, \ldots, \alpha_{13}$.

15. $Ax(x) \equiv x = \alpha_1 \vee x = \alpha_2 \vee \ldots \vee x = \alpha_{13}.$

"x represents an axiom."

16. $\mathfrak{B}(x) \equiv (n)\,[\,0 < n \leq L(x) \rightarrow \{Ax(n \cdot Gl\ x) \vee E(p, q)[0 < p, q < n \,\&\, p\ Glx,\ q\ Glx\ U\ nGlx]\}] \,\&\, L(x) > 0.$

"x represents a proof."

17. $x\mathfrak{B}y \equiv \mathfrak{B}(x) \,\&\, L(x)\ Gl\ x = y.$

"x represents a proof and y a formula, and the proof which x represents is a proof of the formula which y represents."

The assertion that the system is free from contradiction can be written as a proposition of arithmetic thus: $(x, y, z)[\sim(x\mathfrak{B}z \,\&\, y\mathfrak{B}Neg\ (z))]$ (i.e. for all natural numbers x, y and z, x does not represent a proof of the formula A, and y of $\sim(A)$, where z represents A).

5. REPRESENTATION OF RECURSIVE FUNCTIONS BY FORMULAS OF OUR FORMAL SYSTEM

We abbreviate certain formal expressions as follows: z_0 for 0, z_1 for N(0), z_2 for N(N(0)), etc. The z's then <u>represent</u> the natural numbers in the formal logic. Again, if $\phi(x_1, x_2, \ldots)$ is a function of positive integers, we shall say that the formal functional expression $G(u_1, u_2, \ldots)$ <u>represents</u> $\phi(x_1, x_2, \ldots)$ if $G(z_m, z_n, \ldots) = z_{\phi(m, n, \ldots)}$ is provable formally for each given set of natural numbers m, n, ...; in other words, if $G(z_m, z_n, \ldots) = z_k$ is provable formally whenever $\phi(m, n, \ldots) = k$ holds. If the value of $\phi(x_1, x_2, \ldots)$ is independent of some variable x_p, then $G(u_1, u_2, \ldots)$ need not contain the corresponding variable u_p. Similarly, if $R(x_1, x_2, \ldots)$ is a class or relation of natural numbers, we shall say that the formal functional expression $H(u_1, u_2, \ldots)$ <u>represents</u> $R(x_1, x_2, \ldots)$ if we can prove formally $H(z_m, z_n, \ldots)$ whenever $R(m, n, \ldots)$ holds, and $\sim H(z_m, z_n, \ldots)$ whenever $R(m, n, \ldots)$ does not hold.

We now sketch a proof that every recursive function, class, and relation is represented by some formula of our formal system.

The recursive function x + 1 is represented by N(w), because $N(z_n) = z_{n+1}$ can be proved formally for each natural number n. The proof is immediate, since under our abbreviations z_{n+1} <u>is</u> $N(z_n)$. The constant function $f(x_1, x_2, \ldots, x_n) = 0$ is represented by z_0, and the identity function $U^n_j(x_1, \ldots, x_n) = x_j$ is represented by u_j.

If $\phi(x_1, \ldots, x_n)$ is compound with respect to $\psi(x_1, \ldots, x_n)$ and $\chi_i(x_1, \ldots, x_n)$ (i = 1, ..., n), and if $G(w_1, \ldots, w_n)$ represents $\psi(x_1, \ldots, x_n)$ and $H_i(w_1, \ldots, w_n)$ represents $\chi_i(x_1, \ldots, x_n)$, then $G(H_1(w_1, \ldots, w_n), \ldots, H_n(w_1, \ldots, w_n))$ represents $\phi(x_1, \ldots, x_n)$.

If $\phi(x_1, \ldots, x_n)$ is recursive with respect to $\psi(x_1, \ldots, x_{n-1})$ and $\chi(x_1, \ldots, x_{n+1})$, and if $G(w_1, \ldots, w_{n-1})$ represents $\psi(x_1, \ldots, x_{n-1})$ and $H(w_1, \ldots, w_{n+1})$ represents $\chi(x_1, \ldots, x_{n+1})$, then

$$\in z\Big[\Sigma f\Big\{f(0) = G(w_2, \ldots, w_n) \,\&\, \Pi u\big[f(N(u)) = H(u, f(u), w_2, \ldots, w_n)\big]\Big\} \& f(w_1) = Z$$

represents $\phi(x_1, \ldots, x_n)$. This formula (call it $K(w_1, \ldots, w_n)$) intuitively has the desired significance. For each set of natural numbers $w_1, \ldots, w_n$, there is one and only one function f satisfying the conditions $f(0) = G(w_2, \ldots, w_n)$, $f(k + 1) = H(k, f(k), \ldots, w_2, \ldots, w_n)$, and therefore $K(w_1, \ldots, w_n)$ means "The value, which the function f satisfying the above conditions takes on for the argument w_1". This value obviously is $\phi(w_1, \ldots, w_n)$. The proof that $K(w_1, \ldots, w_n)$ actually represents $\phi(x_1, \ldots, x_n)$, if G represents ψ and H represents χ, is too long to give here.

If $R(x, y, \ldots)$ is a recursive class or relation, there is a recursive function $\phi(x, y, \ldots)$ such that $\phi(x, y, \ldots) = 0$ if $R(x, y, \ldots)$ and $\phi(x, y, \ldots) = 1$ if $\sim R(x, y, \ldots)$. Then there is a $G(u, v, \ldots)$ which represents $\phi(x, y, \ldots)$. $G(u, v, \ldots) = 0$ represents $R(x, y, \ldots)$. For if $R(m, n, \ldots)$, then $G(z_m, z_n, \ldots) = z_o = 0$ is provable formally; and if $\sim R(m, n, \ldots)$, then $G(z_m, z_n, \ldots) = z_1$, and therefore $\sim[G(z_m, z_n, \ldots) = 0]$, is provable formally.

Because certain metamathematical relations and propositions about our formal system can be expressed by recursive relations and statements, these relations and propositions can be expressed in the formal system. Thus parts of the theory whose object is our formal system can be expressed in the same formal system. This leads to interesting results.

We have noted that $x \mathfrak{B} y$ is a recursive relation; and we can also prove that $\gamma(x, y)$ is recursive, where $\gamma(x, y)$ is the number of the formula which results when we replace all free occurrences of w by z_y in the formula whose number is x. (In fact, if a is the number of w and $\chi(n)$ the number of z_n, $\gamma(x, y)$ is $Sb(x \overset{a}{\chi(y)})$). Hence there is a formula $B(u, v)$ which represents $x \mathfrak{B} y$, and a formula $S(u, v)$ which represents $\gamma(x, y)$.

Let $U(w)$ be the formula $\pi v[\sim B(v, S(w, w))]$, and let p be the number of $U(w)$. Now $U(z_p)$ is the formula which results when we replace all free occurrences of w by z_p in the formula whose number is p, and hence has the number $\gamma(p, p)$.

Hence if $U(z_p)$ is provable, there is a k such that $k \mathfrak{B} \gamma(p, p)$. But since $S(u, v)$ represents $\gamma(x, y)$ and $B(u, v)$ represents $x \mathfrak{B} y$, it follows that $B(z_k, S(z_p, z_p))$ is provable. Also, it is a property of our system that if $\Pi v F(v)$ is provable, then $F(z_\chi)$ is provable for all χ; consequently, if $U(z_p)$ is provable, $\sim B(z_k, S(z_p, z_p))$, as well as $B(z_k, S(z_p, z_p))$, is provable, and the system contains a contradiction. Thus we conclude that $U(z_p)$ cannot be proved unless the system contains a contradiction.

Next we raise the question of whether $\sim U(z_p)$ can be proved if the system is not contradictory. If the system is not contradictory, $U(z_p)$ cannot be proved, as just seen. But $U(z_p)$ is the formula with the number $\gamma(p, p)$, so that, for all k, $\sim k \mathfrak{B} \gamma(p, p)$. Therefore $\sim B(z_k, S(z_p, z_p))$ is provable for all k. If furthermore $\sim U(z_p)$, i.e. $\sim \Pi v [\sim B(v, S(z_p, z_p))]$, is provable, then we have that a formula is provable which asserts that $\sim B(z_k, S(z_p, z_p))$ is not true for all k, and this, together with the fact that $\sim B(z_k, S(z_p, z_p))$ is provable for all k, makes the system intuitively contradictory. In other words, if we consider the system to be contradictory not merely if there is an A such that both A and $\sim A$ are provable, but also if there is an F such that all of the formulas $\sim \Pi v F(v)$, $F(z_0)$, $F(z_1)$, ... are provable, then, if $\sim U(z_p)$ is provable, the system is contradictory in this weaker sense. Hence neither $U(z_p)$ nor $\sim U(z_p)$ is provable, unless the system is contradictory.

If our system is free from contradiction in the strong sense (i.e., if A and $\sim A$ are not both provable for any A), then $U(z_p)$ is not provable. But $(x, y, z)[\sim \{x \mathfrak{B} y \,\&\, z \mathfrak{B} \text{Neg } y\}]$ is a statement that our system is free from contradiction in the strong sense. Hence we have shown that $(x, y, z)[\sim \{x \mathfrak{B} y \,\&\, z \mathfrak{B} \text{Neg } y\}] \rightarrow (x) \sim x \mathfrak{B} \gamma(p, p)$. The fairly simple arguments of this proof can be paralleled in the formal logic to give a formal proof of

$$\text{Contrad} \rightarrow \Pi v [\sim B(v, S(z_p, z_p))].$$

where Contrad is a formula of the system which expresses the proposition $(x, y, z)[\sim\{x\ \mathfrak{B}\ y\ \&\ z\ \mathfrak{B}\ \text{Neg}\ y\}]$. Then if Contrad could also be proved formally, we could use Rule I to infer $\Pi v[\sim B(v, S(z_p, z_p))]$ or $U(z_p)$, in which case as we have seen, the system would contain a contradiction. Hence Contrad cannot be proved in the system itself, unless the system contains a contradiction.

6. CONDITIONS THAT A FORMAL SYSTEM SHOULD SATISFY IN ORDER THAT THE FOREGOING ARGUMENTS APPLY

Now consider any formal system (in the sense of § 1) satisfying the following five conditions:

(1) Supposing the symbols and formulas to be numbered in a manner similar to that used for the particular system considered above, then the class of axioms and the relation of immediate consequence shall be recursive.

This is a precise formulation of the requirement of § 1 that the class of axioms and relation of immediate consequence be constructive.

(2) There shall be a certain sequence of meaningful formulas z_n (standing for the natural numbers n), such that the relation between n and the number representing z_n is recursive.

(3) There shall be a symbol $\sim$ (negation) and two symbols v and w (variables) such that to every recursive relation of two variables there corresponds a formula R(v, w) of the system such that $R(z_p, z_q)$ is provable if the relation holds of p and q and $\sim R(z_p, z_q)$ is provable if the relation does not hold of p and q; or, instead of a single symbol $\sim$, there may be a formula F(x) not containing v or w such that the foregoing holds when $\sim(A)$ stands for the formula F(A).

The formulas R(v, w) which represent recursive relations, and their negations $\sim R(v, w)$, shall be called <u>recursive propositional functions of two variables</u>; and $R(v, z_n)$ and $\sim R(v, z_n)$ <u>recursive propositional functions of one variable.</u>

(4) There shall be a symbol such that if $\Pi vF(v)$ is provable for a recursive propositional function F(v) of one variable, then $F(z_k)$ shall be provable for all k;

or, instead of a single symbol Π, there may be a formula G(x) not containing w such that the foregoing holds when $\Pi v\ F(v)$ stands for G(F(v)).

(5) The system shall be free from contradiction in the two following senses:

(a) If R(v, w) is a recursive propositional function of two variables, then $R(z_p, z_q)$ and $\sim R(z_p, z_q)$ shall not both be provable.

(b) If F(v) is a recursive propositional function of one variable, then the formulas $\sim \Pi v F(v)$, $F(z_0)$, $F(z_1)$, $F(z_2)$, ... shall not all be provable.

Now, using condition (1), $x \mathfrak{B} y$, $\mathfrak{Y}(x, y)$ (defined as before), and $k \mathfrak{B} \mathfrak{Y}(\chi, \chi)$ are recursive. Then, by (3), there is an R(v, w) such that $R(z_k, z_\chi)$ is provable if $k \mathfrak{B} \mathfrak{Y}(\chi, \chi)$, and $\sim R(z_k, z_\chi)$ is provable if $\sim k \mathfrak{B} \mathfrak{Y}(\chi, \chi)$. Noting that R(v, w) plays the same role as B(v, S(w, w)) in our special system, we can prove by reasoning similar to that of § 5 that, if p is the number of $\Pi v \sim R(v, w)$, (5a) implies that $\Pi v \sim R(v, z_p)$ is not provable and (5b) implies that $\sim \Pi v \sim R(v, z_p)$ is not provable. Also, as before, $(x, y, z)\left[\sim\{x \mathfrak{B} y \,\&\, z \mathfrak{B} \text{Neg } y\}\right] \rightarrow (x) \sim x \mathfrak{B} \mathfrak{Y}(p, p)$ can be established. We shall not list the further conditions under which it is possible to convert the intuitive proof of this into a formal proof of Contrad $\rightarrow \Pi v \sim R(v, z_p)$ (Contrad defined as before). However, they are conditions satisfied by all systems of the type under consideration which contain a certain amount of ordinary arithmetic, and these systems therefore cannot contain a proof of their own freedom from contradiction.

7. RELATION OF FOREGOING ARGUMENTS TO THE PARADOXES

We have seen that in a formal system we can construct statements about the formal system, of which some can be proved and some cannot, according to what they say about the system. We shall compare this fact with the famous Epimenides paradox ("Der Lügner"). Suppose that on May 4, 1934, A makes the single statement "Every statement which A makes on May 4, 1934, is false." This statement clearly cannot be true. Also it can't be false, since the only way for it to be false is for A to

have made a true statement in the time specified and in that time he made only the single statement.

The solution suggested by Whitehead and Russell, that a proposition cannot say something about itself, is too drastic. We saw that we can construct propositions which make statements about themselves, and, in fact, these are arithmetic propositions which involve only recursively defined functions, and therefore are undoubtedly meaningful statements. It is even possible, for any metamathematical property f which can be expressed in the system, to construct a proposition which says of itself that it has this property. For suppose that $F(z_n)$ means that n is the number of a formula that has the property f. Then if $F(S(w, w))$ has the number p, $F(S(z_p, z_p))$ says that it itself has the property f[15]. This construction can

---

[15] Of course we can find properties f such that $F(S(z_p, z_p))$ is provable, just as we found ones for which it was not provable.

---

only be carried out if the property f can be expressed in the system, and the solution of the Epimenides lies in the fact that the latter is not possible for every metamathematical property. For consider the above statement made by A. A must specify a language B and say that every statement that he made in the given time was a false statement in B. But "false statement in B" cannot be expressed in B, and so his statement was in some other language, and the paradox disappears.

The paradox can be considered as a proof that "false statement in B" cannot be expressed in B. We now will establish this fact in a more formal manner, and in doing so obtain a heuristic argument for the existence of undecidable propositions. Suppose that $T(z_n)$ means that the formula whose number is n is true. That is, if n is the number of N, $T(z_n)$ shall be equivalent to N. Then we could apply our procedure to $\sim T(S(w, w))$, obtaining $\sim T(S(z_p, z_p))$, which says that it is itself false, and this leads to a contradiction similar to the "Epimenides". But, on the other hand, $\sum v\ B(v, z_k)$ is a statement in the system of the fact that the formula with number k is provable. So we see that the class κ of numbers of true

formulas cannot be expressed by a propositional function of our system, whereas the class β of provable formulas can. Hence $\alpha \neq \beta$, and if we assume $\beta \leq \alpha$ (i.e. every provable formula is true) we have $\beta < \alpha$, i.e. there is a proposition A which is true but not provable. $\sim$A then is not true and therefore not provable either, i.e. A is undecidable.

8. DIOPHANTINE EQUIVALENTS OF UNDECIDABLE PROPOSITIONS

Suppose $F(x_1, \ldots, x_n)$ a polynomial with non-negative integral coefficients. By use of logical quantifiers (x) and (Ex)[16], we can make certain statements about

---

[16] Where x is any variable for a natural number.

---

the solutions in natural numbers of the Diophantine equation $F(x_1, \ldots, x_n) = 0$. Thus $(Ex_1)(Ex_2) \ldots (Ex_n)(F(x_1, \ldots, x_n) = 0)$ says that there is a solution; $(x_3)(Ex_1)(Ex_2)(Ex_4) \ldots (Ex_n)(F(x_1, \ldots, x_n) = 0)$ says that for any assigned value of x_3, the resulting equation has a solution; etc. We wish to prove that there is a sequence of logical quantifiers, say (P), and a Diophantine equation, F = 0, such that our undecidable proposition is equivalent to (P)(F = 0).

To prove this we find it convenient to make use of the intermediary concept of an arithmetical expression, that is an expression built up out of $\sim$, $\vee$, $\&$, $\rightarrow$, $\equiv$, +, $\times$, =, natural numbers, variables running over natural numbers, and the quantifiers (x) and (Ex)[16], according to the following induction:

1. If f and g are built up out of variables, natural numbers, +, and $\times$[17], then f = g is an arithmetical expression.

---

[17] That is, if f and g are polynomials with natural number coefficients.

---

2. If A and B are arithmetical expressions, then $\sim$A, A $\vee$ B, A $\&$ B, A $\rightarrow$ B, and A $\equiv$ B are arithmetical expressions.
3. If A is an arithmetical expression which contains x as a free variable, then (x)A and (Ex)A are arithmetical expressions.

We shall prove first that, if $\phi(x_1, \ldots, x_n)$ is recursive, then there is

an arithmetical expression $A(x_1, \ldots, x_n, y)$ such that

$A(x_1, \ldots, x_n, y) \cdot \equiv \cdot \phi(x_1, \ldots, x_n) = y$; and second that if $B(x_1, \ldots, x_n)$ is an arithmetical expression, then there are polynomials $Q(x_1, \ldots, x_n, y_1, \ldots, y_m)$ and $R(x_1, \ldots, x_n, y_1, \ldots, y_m)$ with natural number coefficients and a sequence (P) of quantifiers such that

$B(x_1, \ldots, x_n) \cdot \equiv \cdot (P)\left[Q(x_1, \ldots, x_n, y_1, \ldots, y_m) = R(x_1, \ldots, x_n, y_1, \ldots, y_m)\right]$, where the x's and y's range over the natural numbers. Since our undecidable proposition has the form $(x)F(x)$ where F is recursive, there is a recursive function $\phi(x)$ such that our undecidable proposition is equivalent to $(x)\left[\phi(x) = 0\right]$. Then there is an arithmetical expression $A(x, y)$ such that $\phi(x) = y \cdot \equiv \cdot A(x, y)$, and there are polynomials $Q(x, y, z_1, \ldots, z_m)$ and $R(x, y, z_1, \ldots, z_m)$ with natural number coefficients, and a sequence of quantifiers (P) such that

$A(x, y) \cdot \equiv \cdot (P)\left[Q(x, y, z_1, \ldots, z_m) = R(x, y, z_1, \ldots, z_m)\right]$. Then our undecidable proposition is equivalent to $(x)(P)\left[Q(x, 0, z_1, \ldots, z_m) = R(x, 0, z_1, \ldots, z_m)\right]$.

We prove first that recursive functions are expressible arithmetically.

If $f(x) = x + 1$, $f(x) = y \cdot \equiv \cdot x + 1 = y$.

If $f(x_1, \ldots, x_n) = c$, $f(x_1, \ldots, x_n) = w \cdot \equiv \cdot w = c$

Similarly for the identity function $U^n_j(x_1, \ldots, x_n)$.

If $\psi(x_1, \ldots, x_n) = y \cdot \equiv \cdot A(x_1, \ldots, x_n, y)$, $\chi_i(x, \ldots, x_n) = y \cdot \equiv \cdot B_i(x_1, \ldots, x_n, y)$, and $\phi(x_1, \ldots, x_n) = \psi(\chi_1(x_1, \ldots, x_n), \ldots, \chi_n(x_1, \ldots, x_n))$, then $\phi(x_1, \ldots, x_n) = y \cdot \equiv \cdot (Et_1) \ldots (Et_n)$

$\left[B_1(x_1, \ldots, x_n, t_1) \,\&\, \ldots \,\&\, B_n(x_1, \ldots, x_n, t_n) \,\&\, A(t_1, \ldots, t_n, y)\right]$.

To handle the case where ϕ is recursive with respect to ψ and χ, we require an arithmetical expression for $\beta(c, d, i) = y$, where $\beta(c, d, i)$ is a certain function which has the property that if a function $f(i)$ of natural numbers and a natural number γ are given, then natural numbers c and d such that $\beta(c, d, i) = f(i)$ $(i = 0, \ldots, \gamma)$ can be found. We may define $x \equiv y \pmod z$ as

$(Et)[x = y + tz \vee y = x + tz]$, and $x \geqq y$ as $(Et)[x = y + t]$.[18] Then we define

[18] If we were allowing the variables to run over the positive and negative integers instead of just the natural numbers we could define $x \geqq y$ as $(Et_1) \ldots (Et_4)[x = y + t_1^2 + t_2^2 + t_3^2 + t_4^2]$, since every positive integer is the sum of four squares.

$\beta(c, d, i)$ to be the least non-negative residue of c modulo $1 + (i + 1)d$, i.e. $\beta(c, d, i) = z \cdot \equiv \cdot z \equiv c \pmod{[1 + (i + 1)d]}$ & $z \leqq (i + 1)d$. To prove that $\beta(c, d, i)$ has the aforesaid property, suppose $f(i)$ and χ given. Choose s greater than all of the numbers χ, $f(0)$, $f(1)$, ..., $f(\chi)$. Then the numbers $1 + s!$, $1 + 2s!$, ..., $1 + (\chi+1)s!$ are relatively prime. For if a prime number divides two of them, it divides their difference $(i-j)s!$; but it cannot divide $s!$, since it divides $1 + is!$; then also it cannot divide $i - j$, since $i - j \leqq \chi < s$ and hence $i - j$ is a factor of $s!$. Then if we let $d = s!$, we can find a c such that $c \equiv f(i) \pmod{[1 + (i+1)d]}$ $(i = 0, \ldots, \chi)$ since $1 + s!$, ..., $1 + (\chi+1)s!$ are relatively prime. Since $s > f(i)$ and therefore $1 + (i+1)s! > f(i)$, we have $f(i) = \beta(c, d, i)$ as was to be shown.

If $\psi(x_1, \ldots, x_{n-1}) = y \cdot \equiv \cdot A(x_1, \ldots, x_{n-1}, y)$, $\chi(x_1, \ldots, x_{n+1}) = y \cdot \equiv \cdot B(x_1, \ldots, x_{n+1}, y)$, $\phi(0, x_2, \ldots, x_n) = \psi(x_2, \ldots, x_n)$, and $\phi(k + 1, x_2, \ldots, x_n) = \chi(k, \phi(k, x_2, \ldots, x_n), x_2, \ldots, x_n)$, then $\phi(x_1, \ldots, x_n) = y \cdot \equiv \cdot (Ef)[A(x_2, \ldots, x_n, f(0))$ & $(t)\{t + 1 \leqq x_1 \rightarrow B(t, f(t), x_2, \ldots, x_n, f(t+1))\}$ & $f(x_1) = y]$. But if there is an f satisfying the condition in square brackets, then there is a c and a d such that $\beta(c, d, i) = f(i)$ $(i = 0, \ldots, x_1)$ and therefore $(Ec)(Ed)[A(x_2, \ldots, x_n, \beta(c, d, 0))$ & $(t)\{t + 1 \leqq x_1 \rightarrow B(t, \beta(c, d, t), x_2, \ldots, x_n, \beta(c, d, t+1))\}$ & $\beta(c, d, x_1) = y]$. Conversely, this obviously implies the original expression. The latter formula can be transformed into the arithmetical one

$(Ec)(Ed)[(Ev)\{A(x_2, \ldots, x_n, v)$ & $v = \beta(c, d, 0)\}$ & $(t)\{t + 1 \leqq x_1 \rightarrow (Ev)(Ew)[B(t, v, x_2, \ldots, x_n, w)$ & $v = \beta(c, d, t)$ & $w = \beta(c, d, t + 1)]\}$

& $y = \beta(c, d, x_1)]$ by substituting $(Ev)[A(x_2, \ldots, x_n, v) \;\&\; v = \beta(\not{x}, d, 0)]$ for $A(x_2, \ldots, x_n, \beta(\not{x}, d, 0))$, and

$(Ev)(Ew)[B(t, v, x_2, \ldots, x_n, w) \;\&\; v = \beta(\not{x}, d, t) \;\&\; w = \beta(\not{x}, d, t+1)]$ for

$B(t, \beta(\not{x}, d, t), x_2, \ldots, x_n, \beta(\not{x}, d, t+1))$. This completes the proof that all recursive functions are arithmetical.

We next show that all arithmetical expressions can be given the equivalent normal form $(P)[Q = R]$ where Q and R are polynomials with natural number coefficients.

If $\sim$, $\vee$, &, $\rightarrow$, $\equiv$, and quantifiers do not occur in an arithmetical expression, then it has the required normal form by definition (p. 22)

Suppose that $A \equiv (P)[Q = R]$ where x does not occur in the quantifiers denoted by (P). Then $(x)A \cdot \equiv \cdot (x)(P)[Q = R]$ and $(Ex)A \cdot \equiv \cdot (Ex)(P)[Q = R]$.

Suppose also that $B \equiv (P')[Q' = R']$ where the variables of (P') are distinct from those of (P). Then owing to the fact that $p \vee (Ex)F(x) \cdot \equiv \cdot (Ex)p \vee F(x)$ and $p \vee (x)F(x) \cdot \equiv \cdot (x)p \vee F(x)$,

$$A \vee B \cdot \equiv \cdot (P)(P')[Q = R \vee Q' = R']$$
$$\equiv \cdot (P)(P')[Q - R = 0 \vee Q' - R' = 0]$$
$$\equiv \cdot (P)(P')[(Q - R)(Q' - R') = 0]$$
$$\equiv \cdot (P)(P')[QQ' + RR' = Q'R + QR']$$

Moreover $\sim A \cdot \equiv \cdot \sim(P)[Q = R]$. Then, since $\sim(x)p \cdot \equiv \cdot (Ex) \sim p$ and $\sim(Ex)p \equiv \cdot (x) \sim p$, we can shift the negative sign through the prefix (P) and find a P" such that

$$\sim A \cdot \equiv \cdot (P")[\sim Q = R]$$
$$\equiv \cdot (P")[(Q - R)^2 > 0]$$
$$\equiv \cdot (P")[Q^2 + R^2 \geq 2QR + 1]$$
$$\equiv \cdot (P")(Et)[Q^2 + R^2 = 2QR + t + 1]$$

&, $\rightarrow$, and $\equiv$ are expressible by means of $\sim$ and $\vee$.

If the argument is modified slightly, the variables can be allowed to run over the integers instead of just the natural numbers.

Thus there exists a statement about the solutions of a Diophantine equation which is not decidable in our formal system. It can be shown that it is decidable in the next higher type, but there is another such statement which is not decidable even in that type, but which is decidable by going into the next higher type; and so on. In other words, there can be no complete theory of Diophantine analysis.

9. GENERAL RECURSIVE FUNCTIONS

If $\psi(y)$ and $\chi(x)$ are given recursive functions, then the function $\phi(x, y)$, defined inductively by the relations $\phi(0, y) = \psi(y)$, $\phi(x+1, 0) = \chi(x)$, $\phi(x+1, y+1) = \phi(x, \phi(x+1, y))$, is not in general recursive in the limited sense of § 2. This is an example of a definition by induction with respect to two variables simultaneously.

To get arithmetical definitions of such functions, we have to generalize our β function. The consideration of various sorts of functions defined by inductions leads to the question what one would mean by "every recursive function".

One may attempt to define this notion as follows: If ϕ denotes an unknown function, and $\psi_1, \ldots, \psi_k$ are known functions, and if the ψ's and the ϕ are substituted in one another in the most general fashions and certain pairs of the resulting expressions are equated, then if the resulting set of functional equations has one and only one solution for ϕ, ϕ is a recursive function[19].

---

[19] This was suggested by Herbrand.

Thus we might have

$$\phi(x, 0) = \psi_1(x),$$
$$\phi(0, y+1) = \psi_2(y),$$
$$\phi(1, y+1) = \psi_3(y),$$
$$\phi(x+2, y+1) = \psi_4(\phi(x, y+2), \phi(x, \phi(x, y+2))).$$

We shall make two restrictions on Herbrand's definition. The first is that the left-hand side of each of the given functional equations defining ϕ shall be of the form

$$\phi(\psi_{i1}(x_1, \ldots, x_n), \psi_{i2}(x_1, \ldots, x_n), \ldots, \psi_{i\chi}(x_1, \ldots, x_n)).$$

The second (as stated below) is equivalent to the condition that all possible sets of arguments $(n_1, \ldots, n_\chi)$ of ϕ can be so arranged that the computation of the value of ϕ for any given set of arguments $(n_1, \ldots, n_\chi)$ by means of the given equations requires a knowledge of the values of ϕ only for sets of arguments which precede $(n_1, \ldots, n_\chi)$.

From the given set of functional equations, we define by induction a set of derived equations, thus:

(1a) Any expression obtained by replacing all the variables of one of the given equations by natural numbers shall be a derived equation.

(1b) $\psi_{ij}(k_1, \ldots, k_n) = m$ shall be a derived equation if $k_1, \ldots, k_n$, are natural numbers, and $\psi_{ij}(k_1, \ldots, k_n) = m$ is a true equality.

2a) If $\psi_{ij}(k_1, \ldots, k_n) = m$ is a derived equation, the equality obtained by substituting m for an occurrence of $\psi_{ij}(k_1, \ldots, k_n)$ in a derived equation shall be a derived equation.

(2b) If $\phi(k_1, \ldots, k_\chi) = m$ is a derived equation where $k_1, \ldots, k_\chi$, m are natural numbers, the expression obtained by substituting m for an occurrence of $\phi(k_1, \ldots, k_\chi)$ on the right-hand side of a derived equation shall be a derived equation.

Now our second restriction on Herbrand's definition of recursive function is that for each set of natural numbers $k_1, \ldots, k_\chi$ there shall be one and only one m such that $\phi(k_1, \ldots, k_\chi) = m$ is a derived equation.

Using this definition of the notion of a recursive function, we can prove that, if $\phi(x_1, \ldots, x_\chi)$ is recursive, there is an arithmetical expression $A(x_1, \ldots, x_\chi, y)$ such that $\phi(x_1, \ldots, x_\chi) = y \cdot \equiv \cdot A(x_1, \ldots, x_\chi, y)$.

Presburger has given a set of axioms for the relations built up out of +, =, and logical symbols, together with a method of deciding such relations[20].

20 Presburger, Über die Vollständigkeit eines gewissen Systems der Arithmetik etc., Comptes Rend. du I. Congrès des Math. des Pays Slaves, Warszawa 1929.

Skolem has sketched a method of deciding relations constructed similarly using $\times$ instead of +[21].

21 Th. Skolem, Über einige Satzfunktionen in der Arithmetik, Vidensk. Akad. Skrifter, Mat.-Nat. Kl., 1930, No. 7.

However there is no general method of deciding relations in which both + and $\times$ occur, since (as shown above) there can be no general theory of Diophantine equations.

NOTES AND ERRATA

p. 3 The first sentence may be omitted, since the removal of any of the occurrences of variables on the right may be effected by means of the function U^n_j.

p. 5, line 9 Insert an additional parenthesis) after the last 2f.

p. 7, ℙ1. As indicated on the next page, the substitutions which are meant in the case of $\in x[y = N(x)]$ and $\sum x[y = N(x)]$ are substitutions for y.

p. 7, last ℙ. By an occurrence of a symbol (or formula) K in an expression A, we shall mean a particular part of A of the form K.

p. 8, last ℙ. Here we describe the proper method of substituting an expression G(x) for a functional variable f in an expression A, and denote the result of the substitution by Subst' $(A^f_{G(x)})$. If G(x) does not contain f (which can always be made the case in the course of formal proofs by a change in the notation), Subst' $(A^f_{G(x)})$ may also be defined as follows: Replace f(U) by G(U) for one of the free occurrences of f in A, do the same thing with the resulting expression, and so on, until an expression is obtained in which f no longer occurs as a free variable.

p. 10, Rule 1. Omit the third pair of parentheses.

p. 10, Rule 3. Instead of "Suppose" read "Suppose that A is meaningful, that t is a variable, and".

p. 10, Rule 4d. If F(x) is an expression of the IInd kind containing the variable x for a number as a free variable, then, with the aid of this rule, $\sum x\, F(x) \longrightarrow F(\in x[F(x)])$ is provable in our formal system. For Rule 4c allows us to infer $(p \longrightarrow [(\sim p) \longrightarrow p]) \longrightarrow ((([(\sim p) \longrightarrow p] \longrightarrow p) \wedge (p \longrightarrow p))$ from Axiom A1 (i.e., by substituting $(\sim p) \longrightarrow p$ for q and p for r), and $p \longrightarrow [(\sim p) \longrightarrow p]$ from Axiom A3 (by substituting p for q). Then Rule 1 allows us from these two results to infer $([(\sim p) \longrightarrow p] \longrightarrow p) \longrightarrow (p \longrightarrow p)$, and then from the latter and Axiom A2 to infer $p \longrightarrow p$. Thence we can successively

infer $F(x) \longrightarrow F(x)$ by Rule 4c (by substituting $F(x)$ for p), $F(x) \longrightarrow F(\in x[F(x)])$ by Rule 4d, $F(x) \longrightarrow F(\in y[F(y)])$ by Rule 6, $\sum x[F(x)] \longrightarrow F(\in y[F(y)])$ by Rule 3a, and $\sum x[F(x)] \longrightarrow F(\in x[F(x)])$ by Rule 6. Thus the last formula, which expresses the essential property of $\in$, is proved in our formal system. If the system admitted the use of $\in$ with variables for functions (i.e., if a rule of inference 4d', obtained from 4d by replacing "x" by "f" and "number" by "function", were added), then similarly, for any expression G(f) containing the variable f for a function as a free variable, $\sum f[F(f)] \longrightarrow G(\in f[G(f)])$ would be provable. The latter formula expresses the axiom of choice for classes of functions of integers.

as given below

Note that by our formal rule for $\in$ we cannot prove that $\in xF(x)$ is the smallest integer x for which F(x), nor that $\in xF(x) = 0$ if there is no such integer, but we can prove only that if there are integers x satisfying F(x), then $\in xF(x)$ is one of them $\left(\text{i.e. } \sum x[F(x)] \longrightarrow F(\in x[F(x)])\right)$. This however suffices for all applications.

p. 10, Rule 6. Instead of "If A, then A'", read "Suppose that A is meaningful, and let B' denote the expression obtained from B by the substitution of A' for a given occurrence of A in B. If B, then B'." Note that B may be A itself (then B' is A').

p. 14, def. 13, 2nd line. Read "= Imp (v, ...".