MATHEMATICAL LOGIC: CONSTRUCTIVE AND NON-CONSTRUCTIVE OPERATIONS

By S. C. KLEENE

1. Mathematical logic

Early in the century, especially in connection with Hilbert's treatment of geometry (1899), it was being said that the theorems of an axiomatic theory express truths about whatever systems of objects make the axioms true.

In the simplest case, a system S consists of a non-empty set D (the *domain*), in which there are distinguished certain *individuals*, and over which there are defined certain n-place *functions* (or operations) taking values in D, and certain n-place *predicates* (or properties and relations), i.e. functions taking propositions as values.

The *elementary* (or *first-order*) *predicate calculus* provides a language for discussing such systems. To a preassigned list of (non-logical) *constants* for the distinguished individuals, functions and predicates, we add the *propositional connectives* $\rightarrow$ ('implies' or 'if ... then ...'), & ('and'), $\vee$ ('or'), $\neg$ ('not'), the *universal quantifier* (a) ('for all a (in D)'), and the *existential quantifier* (Ea) ('(there) exists (an) a (in D such that)').

For example, when S is the arithmetic of the natural numbers 0, 1, 2, ..., with 0, 1, $+$, $\cdot$, $=$, $>$ in their usual senses,

$$(\alpha)\ a = b+1, \quad (\beta)\ (Eb)(a = b+1), \quad (\gamma)\ a > 0,$$

$$(\delta)\ a > 0 \rightarrow (Eb)(a = b+1), \quad (\epsilon)\ (a)[a > 0 \rightarrow (Eb)(a = b+1)],$$

are *formulas*. Formula (α) (containing a, b, *free*) expresses a 2-place predicate (relation), (β)–(δ) (containing a free) express 1-place predicates (properties), and (ϵ) (containing no variable free, i.e. a *sentence*) expresses a proposition.

When (a, b) are $(3, 2)$, (α) is true. Hence when a is 3, (β) is true, also (γ); and hence by the truth table for $\rightarrow$ (right), (δ) is true. Similarly, for any other a, (δ) is true. Hence (ϵ) is true. Truth tables, which in principle go back to Peirce (1885) and Frege (1891), were first fully exploited by Łukasiewicz (1921) and Post (1921), and truth definitions generally by Tarski (1933).

$A \rightarrow B$: A \ B	True	False
True	True	False
False	True	True

We need one elementary technical result of logic. In any formula, the quantifiers can be advanced (step by step) to the front, preserving the truth or falsity of the proposition, or of any value of the predicate, expressed. (For example,

$$[(a)\,A(a)] \rightarrow (a)\,B(a)$$

is equivalent to

$$(Ea)\,(b)\,[A(a) \rightarrow B(b)].)$$

The resulting formula we call a *prenex form* of the original.

I. {Löwenheim (1915). / Skolem (1920).} *If* {*a sentence* A *is* / *sentences* $A_0, A_1, A_2, \ldots$ *are*} *true of a given system* S*, then* {*it is* / *they are all*} *true of a system* S_1 *with countable domain* D_1.

Proof. Say a prenex form of A is

$$(Eb)\,(c)\,(Ed)\,(e)\,(f)\,(Eg)\,A(b, c, d, e, f, g) \qquad \text{(i)}$$

(all quantifiers shown). This being true of S with domain D, there are an individual β and (by the axiom of choice) functions $\delta(c)$ and $\gamma(c, e, f)$ such that

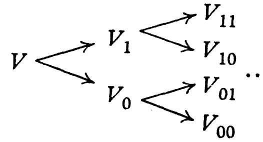
$$(c)\,(e)\,(f)\,A(\beta, c, \delta(c), e, f, \gamma(c, e, f)) \qquad \text{(ii)}$$

is true. Now (ii), and hence (i), will remain true if we cut down the domain (without otherwise altering the functions and predicates) from D to its least subset D_1 containing β (and the distinguished individuals of S) and closed under δ, γ (and the functions of S). The new domain D_1 is countable; indeed all its members have names in the list $t_0, t_1, t_2, \ldots$, of the distinct *terms* without variables formable using β, δ, γ and the symbols for the distinguished individuals and functions of S. (We can always arrange to have at least one individual, and one function, symbol.) For the version with $A_0, A_1, A_2, \ldots$, we use different symbols in the role of β, δ, γ with each prenex form.

Continuing the example, (i) will be true of a system S_1 with domain D_1 whose members are named by $t_0, t_1, t_2, \ldots$, if each of the expressions $A(\beta, t_c, \delta(t_c), t_e, t_f, \gamma(t_c, t_e, t_f))$ $(c, e, f = 0, 1, 2, \ldots)$ is true; enumerate these (or for $A_0, A_1, A_2, \ldots$, the expressions arising similarly from the various prenex forms) as $A^0, A^1, A^2, \ldots$.

For the next theorem we simply try in all possible ways to make $A^0, A^1, A^2, \ldots$ simultaneously true. We obtain the greatest freedom to do this by interpreting each term t_i as representing a different individual, say i. Thereby we can choose the value of each expression $P(t_{c_1}, \ldots, t_{c_n})$ (P an n-place predicate symbol) as true or false independently of the others. Enumerate these (without repetitions) as $Q_0, Q_1, Q_2, \ldots$. Choosing their values successively can be correlated to following a path (indicated by arrows) in the tree (right); e.g. if we choose Q_0 true, Q_1 false, Q_2 false, ..., we follow the path $VV_0V_{01}V_{011}\ldots$. As soon as the values already chosen make any one of

$$A^0, A^1, A^2, \ldots$$



false, we are defeated for that sequence of choices, and terminate the path.

Now by König's Unendlichkeitslemma (1926) (= a classical version of Brouwer's fan theorem, 1924), if (*Case* 1) arbitrarily long finite paths exist, there is an infinite path. (We follow such a path by choosing each time an arrow belonging to arbitrarily long finite paths.) Thereby we obtain the first alternative of:

II. *Either* (1) *all of* $A^0, A^1, A^2, \ldots$ $\left(\textit{and hence} \begin{Bmatrix} A \\ \textit{all of } A_0, A_1, A_2, \ldots \end{Bmatrix}\right)$ *are true of some system* S_1 *with the domain* $D_1 = \{0, 1, 2, \ldots\}$, *or else* (2) *some 'Herbrand conjunction'* $A^{j_1} \& \ldots \& A^{j_m}$

$$\left(\textit{and hence} \begin{Bmatrix} A \\ \textit{some } A_{k_1} \& \ldots \& A_{k_n} \end{Bmatrix}\right)$$

is false of every system S.

If (*Case* 2) there is a finite upper bound $b+2$ to the lengths of paths, then for each of the 2^{b+1} ways of choosing the values of $Q_0, \ldots, Q_b$ some particular A^j will be false. The conjunction $A^{j_1} \& \ldots \& A^{j_m}$ ($m \leqslant 2^{b+1}$) of these A^j's will be false for all 2^{b+1} ways, and thus of all systems S. Likewise A itself (or the conjunction $A_{k_1} \& \ldots \& A_{k_n}$ of those $A_0, A_1, A_2, \ldots$ from which $A^{j_1}, \ldots, A^{j_m}$ arise); for were A true of an S, we would be led as under I to values of $Q_0, \ldots, Q_b$ making $A^{j_1}, \ldots, A^{j_m}$ all true. (Here we need δ and γ for only finitely many arguments, symbolized by terms occurring in $A^{j_1} \& \ldots \& A^{j_m}$, so I is reproved without using the axiom of choice.)

II includes as much of Gödel's completeness theorem for the predicate calculus (1930), and of Herbrand's theorem (1930), as we can state in *model theory*. The theory of models concerns 'mutual relations between sentences of formalized theories and mathematical systems [*models*] in which these sentences hold' (Tarski, 1954–5).

Gödel's completeness theorem (II_G) has $(2_G) \begin{Bmatrix} \neg A \\ \textit{some } \neg(A_{k_1} \& \ldots \& A_{k_n}) \end{Bmatrix}$ *is provable in the predicate calculus* in place of (2), and Herbrand's theorem (II_H) gives the equivalence of (2_G) to (2).

However, *if* we agree here that a 'proof' of a sentence should be a finite linguistic construction, recognizable as being made in accordance with preassigned rules and whose existence assures the 'truth' of the sentence in the appropriate sense, we already have (II), since the verification of (2) for a given $A^{j_1} \& \dots \& A^{j_m}$ is such a construction.

What usual proofs of Gödel's completeness theorem add is that the *proof* of $\neg A$ (or $\neg(A_{k_1} \& \dots \& A_{k_n})$) for ($2_G$) can be effected in a usual *formal system* of *axioms* and *rules of inference* for the predicate calculus as given in *proof theory*.

Proof theory is a modern version of the axiomatic-deductive method, which goes back to Pythagoras (reputedly), Aristotle and Euclid. Since Frege (1879), it has been emphasized that, in order to exclude hidden assumptions, the axioms and rules of inference should be specified by referring only to the *form* of the linguistic expressions (i.e. not to the interpretations or models); hence the term 'formal system'.

With Hilbert since 1904 appeared the idea of proving in a *metatheory* or *metamathematics* theorems about formal systems (cf. Hilbert–Bernays, 1934, 1939; Kleene, 1952). Thus we can talk of proving (metamathematically) that in (2_G) there is a (formal) proof of $\neg A$.

In *Hilbert's* metamathematics it was intended that only safe ('constructive' or 'finitary') methods should be used. That certain methods outrun intuition and even consistency, the mathematical public was forced to recognize by the paradoxes in which Cantor's set theory culminated in 1895. Hilbert hoped to save 'classical mathematics' (including the usual arithmetic and analysis and a suitably restricted axiomatized set theory), which he acknowledged to outrun intuition, by codifying it as a formal system, and proving this system *consistent* (i.e. that no 'contradictory' pair of sentences C and $\neg C$ are provable in it) by finitary metamathematics. Kronecker earlier (in the 1880's), and others later, proposed rather a direct redevelopment of mathematics on a less or more wide constructive basis, such as the intuitionistic (Brouwer, 1908; Heyting, 1956) or the operative (Lorenzen, 1950, 1955).

In a model S_1 as constructed above for II, $=$ may not express equality (identity). (For I, it will if it does for S.) But *if* $A_0, A_1, A_2, \dots$ include the usual axioms for equality, then the relation $\{x = y$ is true of the *above* $S_1\}$ will be an equivalence relation under which the equivalence classes will constitute the domain (countably infinite *or finite*) of a *new* model S_1 *with* $=$ *as equality* (Gödel, 1930). For our applications, we may take II to be thus strengthened.

Applying (II_G) with $\begin{Bmatrix} \neg C \\ \neg C, B_0, B_1, B_2, \ldots \end{Bmatrix}$ as the $\begin{Bmatrix} A \\ A_0, A_1, A_2, \ldots \end{Bmatrix}$: ($II'_G$).

In $\begin{Bmatrix} \textit{the predicate calculus} \\ \textit{theories formalized by the predicate calculus with axioms } B_0, B_1, B_2, \ldots \end{Bmatrix}$, *each sentence C which is true of*

$$\begin{Bmatrix} \textit{every system } S \\ \textit{every system } S \textit{ which makes } B_0, B_1, B_2, \ldots \textit{ true} \end{Bmatrix}$$

is provable as a theorem. This confirms that the predicate calculus fully accomplishes (for 'elementary theories') what we started out by considering as the role of logic. But what is combined with this in Gödel's completeness theorem (including Löwenheim's theorem) is more than was sought, and makes the theorem as much an incompleteness theorem for axiom systems as it is a completeness theorem for logic.

Thus the Löwenheim–Skolem theorem I shows that the axioms of an axiomatic set theory have a countable model (if they have any model at all), despite Cantor's theorem holding in the theory (the Skolem 'paradox', 1922–3).

Furthermore, II entails: (II″) *If the sentences of each finite subset* $A_{k_1}, \ldots, A_{k_n}$ *of* $A_0, A_1, A_2, \ldots$ *are true of a respective system S, then there is a system* S_1*, with countable domain, of which* $A_0, A_1, A_2, \ldots$ *are all true.* This gives the following theorem, found by Skolem (1933, 1934) using another method (and partially anticipated by Tarski, 1927–8).

III. *Say the constants include* $0, +1, =$, *and suppose* $B_0, B_1, B_2, \ldots$ *are true of the system* S_0 *of the natural numbers. Then there is a system* S_1*, with countable domain, not isomorphic to* S_0 *of which* $B_0, B_1, B_2, \ldots$ *are also true.*

Proof. Let $A_0, A_1, A_2, \ldots$ be $B_0, B_1, B_2, \ldots, \neg 0 = \pi, \neg 1 = \pi, \neg 2 = \pi, \ldots$ where π is a new individual symbol. Each $A_{k_1}, \ldots, A_{k_n}$ is true of an S obtained from S_0 by interpreting π as a natural number different from each n for which $\neg n = \pi$ is among $A_{k_1}, \ldots, A_{k_n}$.

Applications of Gödel's completeness theorem to algebra were noted about 1946–7 by Tarski, Henkin and A. Robinson, and have been cultivated since. We have been supposing the number of symbols at most countably infinite, as must be the case of any language in actual use. However, Malcev (1936) extended the completeness theorem to languages with arbitrarily (possibly uncountably) many constants, and Henkin (1947) used such languages to represent the complete addition and multiplication tables, etc., of algebraic systems in the set of formulas for application of the extensions of I–II.

Returning to countable languages, we may consider ones with more than one type of variables, e.g. a second-order predicate calculus with variables ranging over a domain D of individuals and also variables ranging over a collection M of subsets of D. A *standard* model for a set of sentences $A_0, A_1, A_2, \dots$ is one with $M = \{$the set 2^D of all subsets of $D\}$. The above results do not extend when only standard models are used, in view of the categoricity of Peano's axioms for the natural numbers (using a variable over 2^D to express induction). However, Henkin (1947, 1950) introduced the notion of a *general* model in which M may be an appropriate subset of 2^D, and with which he obtained an extension of Gödel's completeness theorem. Thus we are still unable to characterize the natural numbers, except by reading into the axioms the notion of all possible subsets, which is hardly simpler.

We have given the foregoing model theory as part of the familiar classical mathematics, and for the classical 'two-valued' form of the predicate calculus. The negative results obtain all the more from the constructive standpoints. The axiomatic method cannot provide an autonomous foundation for mathematics. The rules of the language of the axioms must (at some level) be understood, and not merely described by more axioms; and this amounts to presupposing the natural numbers intuitively.

2. Constructive and non-constructive operations

The awareness that some mathematical operations are 'constructive', and others are not (at least directly) such, must go far back in mathematical history; witness the word 'algorithm'. A computer cannot tabulate the truth or falsity of $(Ex)\,R(a,x)$, where the variables range over the natural numbers, unless for the particular R he has some theory which gives him an equivalent 'constructive' definition of $(Ex)\,R(a,x)$. Say triples b_0, b_1, b_2 are mapped constructively into single numbers b, with constructive inverses $(b)_0$, $(b)_1$, $(b)_2$. Such a theory is known for $R(a,x) \equiv (a)_0\,(x)_0 + (a)_1\,(x)_1 = (a)_2$, using Euclid's algorithm; but not today for $R(a,x) \equiv ((x)_0+1)^{(x)_3} + ((x)_1+1)^{(x)_3} = ((x)_2+1)^{(x)_3}$ & $(x)_3 > a$, where the value just for $a = 2$ would 'decide' Fermat's 'last theorem'.

In 1936 the claim was made, by Church first and independently by Turing and by Post, that a certain class of functions definable mathematically (in one of several equivalent ways) includes all that are 'computable' or 'effectively calculable' or 'constructively defined' (*Church's thesis*), and conversely that all the functions of this class are 'computable' (*Converse of Church's thesis*).

The definition of this class of functions is not itself constructive. It consists in specifying constructively a type of computation procedure. But a given such procedure may or may not terminate for all arguments, so as to compute a (completely defined) function. (Otherwise, by Cantor's diagonal method one could get constructively outside the class, so Church's thesis could not hold.)

The converse of Church's thesis constructively interpreted means that, whenever one has a constructive proof that the computation procedure always terminates, the function is computable. It is hardly debatable then. A possibility for skepticism remains to one who wishes computability to include constructive provability that the computation procedure always terminates, while allowing the condition that it always terminate to be understood classically; he may imagine that there might be cases when the procedure does always terminate but without there being any constructive proof of that fact.

Much work has been done, especially by Péter since 1932, on special classes of computable functions, for which classes proofs are known that all the computation procedures always terminate.

To Church's thesis itself, the only suggested counterexamples involve 'computation procedures' in which the computer is to perform steps depending on some unpredictable future state of his mind, or in which the 'procedure' is somehow to vary with the argument of the function. But for the thesis, 'computation' is intended to mean of a predetermined function independent of the computer, by only preassigned rules independent of the argument.

We shall now present (essentially) Turing's definition of the class of the 'computable' functions. (Among the equivalents that appear in the literature are the Church–Kleene λ-definable functions, 1933–5, the Herbrand–Gödel general recursive functions, 1934, and definitions using Post's canonical systems, 1943, and Markov's algorithms, 1951.)

Instead of a human computer subjected to preassigned instructions, we can speak of a *machine*. Turing's theory is about *ideal* (digital) computing machines, unhampered by finiteness of storage space or fallibility of functioning. More recently the notion of an *automaton* has been used, by von Neumann (1951); the automaton should not be finite (Kleene, 1956), but potentially infinite (Church, 1957). We want a fixed finite amount of structure (or information) to establish the computation *procedure* for a function $\phi(a)$, while an unbounded amount of space and time must be available to accomodate the argument a and the computation.

The machine or automaton shall accordingly consists of $\aleph_0$ *cells*, each adjacent to at most a given finite number of other cells; but only a finite diversity of structure shall be built into it, the rest of the infinity consisting of identical repetition. Here we use the idea from information theory that information is conveyed only when the signal is not predictable. In order to simplify our brief discussion, we can specialize to the case when the cells are $c_0, c_1, c_2, \ldots$, in the order type of the natural numbers, each c_i (except c_0) being adjacent to exactly two others c_{i-1} and c_{i+1}. The general defense of the Church–Turing thesis then requires arguing that no other arrangement of the cells (with only a finite diversity of structure) would make a function computable that is not computable in this space.

Discrete *moments* of time $0, 1, 2, \ldots$ are distinguished. *States* $s_0, \ldots, s_l$ are given, in one of which each cell shall be at each moment. At moment 0, all but a finite number of the cells shall be in the *passive* state s_0. A *table* is given which determines the state of each cell c_i at moment $t+1$ from its state and the states of the adjacent cells (for $i = 0$, s_0 replacing the state of c_{i-1}) at moment t; the output of this table shall differ from s_0 only when an input does.

To set the problem, say of computing $\phi(a)$ for a as argument, we can take the states at $t = 0$ of the cells $c_0, c_1, c_2, \ldots$ to be

$$s_0 \underbrace{s_1 \ldots s_1}_{a \text{ times}} s_2 \, s_0 \, s_0 \, s_0 \ldots.$$

The answer shall be receivable by the states being

$$s_0 \underbrace{s_1 \ldots s_1}_{a \text{ times}} s_1 \, s_0 \underbrace{s_1 \ldots s_1}_{\phi(a) \text{ times}} s_3 \, s_0 \, s_0 \, s_0 \ldots$$

at a later moment $t = x$ when s_3 first occurs. (The fundamental representation of a natural number b is by b successive marks, so it can be argued that a computation problem is solved only when it is possible to present the solution in this representation.)

One may for example imagine the cells $c_0, c_1, c_2, \ldots$ as representing sheets of paper, each admitting one of finitely many symbols on each of finitely many squares, and one of them carrying as part of its state a human computer in one of finitely many states of mind (cf. Kleene, 1952).

Machines can be used similarly to compute n-place functions $\phi(a_1, \ldots, a_n)$; and they can be used to 'decide' predicates $P(a_1, \ldots, a_n)$ by computing 0 to represent truth and 1 falsity.

The behavior of a machine is completely described by its table, which can be written in code form as a natural number, its *index*.

Let $T(i, a, x) \equiv \{i$ is the index of a Turing machine M_i, which, when applied to compute for a as argument, first at moment x has computed a value $\phi_i(a)\}$.

Here $\phi_i(a)$ is an incompletely defined function of i and a, its condition of definition being $(Ex)\,T(i, a, x)$.

We can constructively decide whether a given i is the index of a machine M_i, and if so given also a and x imitate M_i's behavior for a as argument at moments $0, \ldots, x$ successively. Thus, given i, a, x, we can decide whether $T(i, a, x)$ is true or false. (So there is by Church's thesis, and in a detailed treatment of the subject we would actually construct, a machine that decides $T(i, a, x)$.)

IV. *The function*

$$\psi(a) = \begin{cases} \phi_a(a) + 1 & \textit{if} \quad (Ex)\,T(a, a, x), \\ 0 & \textit{otherwise} \end{cases} \tag{A}$$

is uncomputable.

Proof. Were $\psi(a)$ computable, it would be computed by a machine M_q; so for each a, (B) $\psi(a) = \phi_q(a)$ and (C) $(Ex)\,T(q, a, x)$. Substituting q for a in (C) and using (A), $\psi(q) = \phi_q(q) + 1$, which contradicts (B) with q substituted for a.

V. *The predicate* $(Ex)\,T(a, a, x)$ *is undecidable.*

Proof. Were $(Ex)\,T(a, a, x)$ decidable, we could compute $\psi(a)$ by first deciding $(Ex)\,T(a, a, x)$, and according to the answer, either imitating machine M_a applied to a as argument to compute $\phi_a(a)$ and adding 1, or writing 0. This is Church's theorem 1936, but with a different example of an absolutely undecidable predicate.

In a standard formal system N of arithmetic (or 'number theory'), each decidable predicate, such as $T(i, a, x)$, can be expressed; hence also $(Ex)\,T(a, a, x)$, by a sentence C_a (constructively obtainable from a). Now, for particular a, $(Ex)\,T(a, a, x)$ when true can be 'proved' by doing the computation that shows $T(a, a, x)$ to be true for the appropriate x. This intuitive proof is available formally in a standard N. Thus

$$(Ex)\,T(a, a, x) \rightarrow \{C_a \text{ is provable}\}. \tag{a}$$

Also we are assuming of N that only true formulas are provable in it, so

$$\{C_a \text{ is provable}\} \rightarrow (Ex)\,T(a, a, x). \tag{b}$$

Now V gives:

VI. *There is no procedure for deciding whether a given sentence is* *provable in a formal system N of arithmetic; briefly, N is 'undecidable'* (Church 1936).

Continuing, could we in N also prove $\neg C_a$ whenever $(Ex)\,T(a,a,x)$ is false, besides only then so

$$\{\neg C_a \text{ is provable}\} \rightarrow \neg (Ex)\,T(a,a,x), \tag{c}$$

we would be able, by searching for C_a or $\neg C_a$ among the provable sentences, to decide $(Ex)\,T(a,a,x)$. So, again from V:

VII. *In a formal system N of arithmetic, there is a sentence C_q such that C_q and $\neg C_q$ are both unprovable, though $\neg C_q$ is true (i.e. $\neg(Ex)\,T(q,q,x)$).*

This gives Gödel's famous incompleteness theorem (1931), generalized to apply to all formal systems N satisfying very general conditions, and with the 'formally undecidable' sentence C_q expressing the value, for an argument q depending on the system, of a preassigned predicate $(Ex)\,T(a,a,x)$. The above proof is indirect, the existence of q being inferred from the absurdity that $\neg C_a$ is provable for all a for which it is true. But we can make it direct, by taking as q the index of a machine M_q which, given a, searches through the proofs in N for one of $\neg C_a$, and if one is found writes 0 (but otherwise never computes a value), so

$$(Ex)\,T(q,a,x) \equiv \{\neg C_a \text{ is provable}\}. \tag{d}$$

Substituting q for a in (b)–(d), the three conclusions of VII follow.

Here we have used the feature of formal systems, essential for the purpose which they are intended to serve, that a proof of a sentence can be constructively recognized as being such (and also that C_a can be constructively found from a). Without this feature, we would have a trivial counterexample to VII by taking all the true sentences as the axioms of N. With it, by Church's thesis we conclude the existence of an M_q to any such system. Here the computability notion can be applied directly to the linguistic symbolism, or the latter can be converted to natural numbers as we have already done with machine tables (by a 'Gödel numbering').

The application of Church's thesis by which we obtain VII for all systems N can be avoided for a particular system by actually constructing the M_q for it. This in effect Gödel did in proving his theorem for a particular system before Church's thesis had appeared.

In retrospect, Skolem's theorem III on the existence of unintended models S_1 of systems of sentences $B_0, B_1, B_2, \ldots$ intended to describe the natural numbers suggests Gödel's theorem VII. (Compare the example of Euclid's fifth postulate.) Indeed, for an N based on the elementary predicate calculus, (II'_G) shows that C_q is false of such an S_1. However, III applies even when $B_0, B_1, B_2, \ldots$ are all the true sentences, unlike VII.

I do not consider that VII means we must give up the emphasis on formal systems. The reasons which make a formal system the only accurate way of saying explicitly what assumptions go into a proof are still cogent. Rather VII indicates that, contrary to Hilbert's program, the path of mathematical conquest (even within the already fixed territory of arithmetic) shall not consist solely in discovering new proofs from given axioms by given rules of inference, but also in adducing new axioms or rules. There remains the question whether mathematicians can agree on the validity of the new methods.

In VII, no sooner are we aware that $\neg C_q$ is unprovable than we also know that $\neg C_q$ is true, so we can extend N by adding $\neg C_q$ as a new axiom. This process can be repeated, finitely often, and indeed transfinitely often within the limits of structural constructiveness.

It is illuminating to consider wherein the intuitive proof of $\neg C_q$ transcends N. We only conclude the truth of $\neg C_q$ when we accept (c). By (a), (c) reduces to the consistency of N, which is expressible in N via Gödel numbering by a sentence 'Consis'. The rest of the reasoning that $\neg C_q$ is true is elementary, though tedious when executed in full detail; so we may expect (as has been confirmed by Hilbert and Bernays (1939) for the usual systems as N) that it can be formalized in N. So Consis cannot be provable in N, or $\neg C_q$ would be, contrary to VII. Thus:

VIII. *In a usual formal system N of arithmetic, the sentence* Consis *expressing the consistency of N is unprovable* (Gödel's second incompleteness theorem, 1931).

Thus a system N formalizing classical mathematics cannot be proved consistent, as Hilbert hoped, by a 'subset' of the methods formalized in N.

Gentzen (1936, 1938) gave a proof of the consistency of a system N of arithmetic, in which the method transcending N is a form of transfinite induction over the ordinal numbers $<$ Cantor's first epsilon-number ϵ_0; and other such proofs have appeared since. It is a rather subjective matter whether this should make us feel safer about N than we already feel on the basis of its axioms being true, and its rules of inference preserving truth, under an interpretation ('truth definition') that as classical mathematicians we presumably accept. By a reduction of classical to intuitionistic logic given by Kolmogorov (1925), Gödel (1932–3), Gentzen (1936) and Bernays, the consistency proof by a truth definition can even be managed intuitionistically.

Kreisel (1951–2, 1958) finds the significance of the consistency proofs using ϵ_0-induction in by-products. When a sentence $(a)(Eb)R(a, b)$ (R decidable) is proved, then $(a)R(a, \beta(a))$ will be true for certain functions β, including $\beta(a) = \{$the least b such that $R(a, b)$ is true$\}$, which is computable. It is clear that in a given system N only a subclass of the computable functions can thus be proved to exist; indeed Kleene (1936) gave a proof of Gödel's incompleteness theorem from this idea. Kreisel, however, extracts from Ackermann's consistency proof (1940) a different characterization (not directly from N) of this subclass of the computable functions. The possibility thus appears that some true formula $(a)(Eb)R(a, b)$ might be shown to be unprovable in N because no β for it is in this subclass.

From Church's theorem other undecidability results follow. The theory of $(Ex)T(a, a, x)$ can be formalized in a system N_1 consisting of finitely many axioms $B_1, \ldots, B_k$ adjoined to the (elementary) predicate calculus. So $(Ex)T(a, a, x) \equiv \{C_a$ is provable in $N_1\} \equiv \{B_1 \& \ldots \& B_k \rightarrow C_a$ is provable in the predicate calculus$\}$. Thence from V:

IX. *The elementary predicate calculus is undecidable* (Church, 1936a; Turing, 1936–7).

Various formal systems obtained by adjoining axioms for algebraic systems to the predicate calculus have been shown undecidable by Tarski and others using a method of Tarski (1949) (cf. Tarski *et al.* 1953).

Negative solutions to the problems of the existence of various algebraic algorithms have been obtained by Post (1947), Markov since 1947, and others; in particular, Novikov (1952, 1955) showed the word problem for groups unsolvable.

Turing (1939) introduced the notion of a function $\phi(a)$ *computable from* another function $\psi(a)$ (or predicate $Q(a)$). A simple plan under the above treatment is to print the values of ψ into the space, in this respect alone violating the demand that only a finite amount of information be incorporated, by accenting successions of $\psi(0)+1, \psi(1)+1, \psi(2)+1, \ldots$ cells, preceded and separated by single unaccented cells. In effect, we double the number of states from $s_0, \ldots, s_l$ to $s_0, \ldots, s_l, s_0', \ldots, s_l$.

When the theory is thus relativized to a given predicate $Q(a)$, the decidable predicate $T(i, a, x)$ becomes a predicate $T^Q(i, a, x)$ decidable from Q, and IV, V assume relativized versions IV*, V*.

X. *If $R^Q(a, x)$ is decidable from Q, there is a computable function $\theta(a)$ such that $(Ex)R^Q(a, x) \equiv (Ex)T^Q(\theta(a), \theta(a), x)$.*

Proof. Given a, let $M_{\theta(a)}$ be a machine which tries to compute from Q the constant function whose value is the least x such that $R^Q(a, x)$, by testing successively $x = 0, x = 1, x = 2, \ldots$.

Thus $(Ex)\,R^Q(a,x)$ is decidable from $(Ex)\,T^Q(a,a,x)$ by first computing $\theta(a)$. In particular (taking $R^Q(a,x) \equiv Q(a)\,\&\,x = x$), $Q(a)$ is decidable from $(Ex)\,T^Q(a,a,x)$; but by V*, not conversely. This Post (1948) expressed by saying $(Ex)\,T^Q(a,a,x)$ is of 'higher degree (of unsolvability)' than $Q(a)$. Predicates and functions are of the 'same degree' when each is decidable (or computable) from the other. A decidable predicate is of the lowest degree ('solvability'). Starting from say $H_{(0)}(a) \equiv a = a$, and for each n defining $H_{n+1}(a) \equiv (Ex)\,T^{H_{(n)}}(a,a,x)$, we obtain predicates $H_{(n)}(a)$ $(n = 0, 1, 2, \ldots)$ of ascending degrees. These predicates, together with those decidable from them, turn out to be exactly the predicates (called *arithmetical* by Gödel, 1931) expressible in the usual system of arithmetic. Thus the arithmetical predicates fall into a hierarchy, first described by Kleene (1943) and Mostowski (1946) in terms of the numbers of quantifiers necessary to define them in prenex form from decidable predicates.

The hierarchy can be extended into the transfinite (Davis, Kleene, Mostowski, Post, about 1950; cf. Mostowski, 1951; Kleene, 1955). One method is to consider $H_{(n)}(a)$ as a predicate $H(n,a)$ of both variables; this is of higher degree than each $H_{(n)}(a)$, and thus is non-arithmetical. 'Contracting' $H(n,a)$ to a one-place predicate $H((a)_1,(a)_0)$, which we write $H_{(\omega)}(a)$, we can proceed as before to $H_{(\omega+1)}(a), H_{(\omega+2)}(a), \ldots$. In general, at a limit ordinal ξ of Cantor's second number class approached through an increasing sequence $\{\xi_n\}$, we consider $H_{(\xi_n)}(a)$ as a predicate of n, a, and contract.

However, we have no uniform method, or justification, for picking a particular increasing sequence $\{\xi_n\}$ for ξ. So a diversity of predicates $H_{(\xi)}$ arise, for each transfinite ξ, depending on the selections of increasing sequences. Worse than this, even for $\xi = \omega$, the use of arbitrary increasing sequences $\{\xi_n\}$ with $\lim_n \xi_n = \xi$ (above we used $\xi_n = n$) will give predicates of arbitrarily high degree. This suggests restricting the sequences $\{\xi_n\}$ to be computable, after rendering ordinals accessible to the above notion of computability by representing them in a suitable system of notations, which can be natural numbers (Church–Kleene, 1936; Kleene, 1938). This being done, the diversity in predicates at a given transfinite level ξ, which remains due to the possibility of using different computable increasing sequences, was shown by Spector (1955) to be confined always within a degree. The predicates thus definable corresponding to constructive ordinals, together with all predicates decidable (and functions computable) from them, we call *hyperarithmetical* (Kleene, 1955a).

It was noticed, about 1957, by Addison, Büchi, Grzegorczyk, Kleene, Kuznecov and Myhill (cf. Grzegorczyk *et al.* 1958) that the hyperarithmetical predicates are exactly the predicates expressible unambiguously by a formula of the elementary predicate calculus, when the domain is the natural numbers.

Kleene (1957) formulated computability from higher-type objects, such as from the existential quantifier (Ex) considered as a functional E which operates on a predicate to produce a truth value (or on a function ψ to produce the number 0 if $(Ex)(\psi(x) = 0)$ and 1 otherwise). The hyperarithmetical functions $\phi(a_1, \ldots, a_n)$ are exactly those computable from E; thus, operating constructively, except for using a number quantifier, we obtain not merely the usual predicates of arithmetic but the hyperarithmetical predicates.

REFERENCES

Ackermann, W.

1940. Zur Widerspruchsfreiheit der Zahlentheorie. *Math. Ann.* 117, 162–194.

Brouwer, L. E. J.

1908. De onbetrouwbaarheid der logische principes. (The untrustworthiness of the principles of logic.) *Tijdschrift voor wijsbegeerte*, 2, 152–158.

1924. Beweis, dass jede volle Funktion gleichmässig stetig ist. *Proc. Akad. Wet. Amst.* 27, 189–193.

Church, A.

1933. A set of postulates for the foundation of logic (second paper). *Ann. Math.* (2), 34, 839–864.

1936. An unsolvable problem of elementary number theory. *Amer. J. Math.* 58, 345–363.

1936a. A note on the Entscheidungsproblem. *J. Symb. Logic*, 1, 40–41. Correction, *ibid.* 101–102.

1941. *The Calculi of Lambda-Conversion.* Ann. of Math. Studies, no. 6. Princeton University Press, Princeton, N.J.

1957. Application of recursive arithmetic to the problem of circuit synthesis. *Summaries of Talks Presented at the Summer Institute of Symbolic Logic in* 1957 *at Cornell University* (mimeographed), 1, 3–50; 3, 429.

Church, A. and Kleene, S. C.

1933–5. See Church (1933), Kleene (1935), Church (1941).

1936. Formal definitions in the theory of ordinal numbers. *Fundam. Math.* 28, 11–21.

Frege, G.

1879. *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens.* Nebert, Halle.

1891. *Funktion und Begriff.* Jena.

Gentzen, G.

1936. Die Widerspruchsfreiheit der reinen Zahlentheorie. *Math. Ann.* 112, 493–565.

1938. Neue Fassung des Widerspruchsfreiheitsbeweises für die reine Zahlentheorie. *Forschungen zur Logik und zur Grundlegung der exakten Wissenschaften*, N.S., no. 4, 19–44. Hirzel, Leipzig.

Gödel, K.

1930. Die Vollständingkeit der Axiome des logischen Funktionenkalküls. *Monatsh. Math. Phys.* 37, 349–360.

1931. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme. I. *Monatsh. Math. Phys.* 38, 173–198.

1932–3. Zur intuitionistischen Arithmetik und Zahlentheorie. *Ergebn. math. Kolloq.* Heft 4, 34–38 (for 1931–2, publ. 1933).

1934. *On Undecidable Propositions of Formal Mathematical Systems* (mimeographed). Princeton, N.J.

Grzegorczyk, A., Mostowski, A. and Ryll-Nardzewski, C.

1958. The classical and the ω-complete arithmetic. *J. Symb. Logic*, 23, 188–206.

Henkin, L.

1947. *The Completeness of Formal Systems.* Princeton University Ph.D. Thesis, Princeton, N.J.

1950. Completeness in the theory of types. *J. Symb. Logic*, 15, 81–91.

Herbrand, J.

1930. *Recherches sur la théorie de la démonstration.* Travaux de la Société des Sciences et des Lettres de Varsovie, Classe III, sciences mathématiques et physiques, no. 33.

Herbrand, J. and Gödel, K.

1934. See Gödel (1934), Kleene (1936, 1952).

Heyting, A.

1956. *Intuitionism, An Introduction.* North Holland Publ. Co., Amsterdam.

Hilbert, D.

1899. *Grundlagen der Geometrie*, 7th ed. (1930), Teubner, Leipzig and Berlin.

1904. Über die Grundlagen der Logik und der Arithmetik. *Verhand. Dritten Int. Math.-Kong. Heidelberg* 1904, 247–261 (publ. Leipzig 1905).

Hilbert, D. and Bernays, P.

1934. *Grundlagen der Mathematik*, vol. 1. Springer, Berlin.

1939. *Grundlagen der Mathematik*, vol. 2. Springer, Berlin.

Kleene, S. C.

1935. A theory of positive integers in formal logic. *Amer. J. Math.* 57, 153–173, 219–244.

1936. General recursive functions of natural numbers. *Math. Ann.* 112, 727–742.

1938. On notation for ordinal numbers. *J. Symb. Logic*, 3, 150–155.

1943. Recursive predicates and quantifiers. *Trans. Amer. Math. Soc.* 53, 41–73.

1952. *Introduction to Metamathematics.* North Holland Publ. Co. (Amsterdam), Noordhoff (Groningen), Van Nostrand (New York and Toronto).

1955. Arithmetical predicates and function quantifiers. *Trans. Amer. Math. Soc.* 79, 312–340.

1955a. Hierarchies of number-theoretic predicates. *Bull. Amer. Math. Soc.* 61, 193–213.

1956. Representation of events in nerve nets and finite automata. *Automata Studies.* Ann. of Math. Studies, no. 34, 3–41. Princeton University Press, Princeton, N.J.

Kleene, S. C.
1957. Recursive functionals of higher finite types. *Summaries of Talks Presented at the Summer Institute of Symbolic Logic in* 1957 *at Cornell University* (mimeographed), 1, 148–154. Errata, 3, 429,

Kolmogorov, A.
1925. Sur le principe de tertium non datur. *Rec. Math., Moscou,* 32, 646–667.

König, D.
1926. Sur les correspondences multivoques des ensembles. *Fundam. Math.* 8, 114–134.

Kreisel, G.
1951–2. On the interpretation of non-finitist proofs. *J. Symb. Logic,* 16, 241–267; 17, 43–58.
1958. Mathematical significance of consistency proofs. *J. Symb. Logic,* 23, 155–182.

Lorenzen, P.
1950. Konstruktive Begründung der Mathematik. *Math. Z.* 53, 162–202.
1955. *Einführung in die operative Logik und Mathematik.* Springer, Berlin, Göttingen and Heidelberg.

Löwenheim, L.
1915. Über Möglichkeiten im Relativkalkül. *Math. Ann.* 76, 447–470.

Łukasiewicz, Jan
1921. Logika dwuwartościowa. (Two-valued logic.) *Przegląd Filozoficzny,* 23, 189–205.

Malcev, A.
1936. Untersuchungen aus dem Gebiete der mathematischen Logik. *Mat. Sbornik,* 1 (43), 323–336.

Markov, A. A.
1947. Névozmožnosť nékotoryh algorifmov v téorii associativnyh sistém. (On the impossibility of certain algorithms in the theory of associative systems.) *Dokl. Akad. Nauk, SSSR,* N.S., 55, 587–590.
1951. Téoriá algorifmov. (The theory of algorithms.) *Trudy Matématičéskogo Instituta iméni V. A. Steklova,* 38, 176–189.

Mostowski, A.
1946. On definable sets of positive integers. *Fundam. Math.* 34, 81–112.
1951. A classification of logical systems. *Studia Philosophica,* 4, 237–274.

Novikov, P. S.
1952. Ob algoritmíčeśkoj nérazréšimosti problémy toždéstva. (On algorithmic unsolvability of the word problem.) *Dokl. Akad. Nauk, SSSR,* N.S., 85, 709–712.
1955. Ob algoritmiiceśkoj nérazréšimosti problémy toždéstva slov v téorii grupp. (On the algorithmic unsolvability of the word problem in group theory.) *Trudy Mat. Inst. im. Steklov,* no. 44. *Izdat. Akad. Nauk, SSSR.* Moscow.

Peirce, C. S.
1885. On the algebra of logic: A contribution to the philosophy of notation. *Amer. J. Math.* 7, 180–202.

Péter, R.
1951. *Rekursive Funktionen.* Akadémiai Kiadó (Akademischer Verlag), Budapest.

Post, E.
1921. Introduction to a general theory of elementary propositions. *Amer. J. Math.* 43, 163–185.
1936. Finite combinatory processes—formulation. I. *J. Symb. Logic,* 1, 103–105.
1943. Formal reductions of the general combinatorial decision problem. *Amer. J. Math.* 65, 197–215.
1947. Recursive unsolvability of a problem of Thue. *J. Symb. Logic,* 12, 1–11.
1948. Degrees of recursive unsolvability (abstract). *Bull. Amer. Math. Soc.* 54, 641–642.

Robinson, A.
1951. *On the Metamathematics of Algebra.* North Holland Publ. Co., Amsterdam.

Skolem, T.
1920. Logisch-kombinatorische Untersuchungen über die Erfüllbarkeit oder Beweisbarkeit mathematischer Sätze nebst einem Theoreme über dichte Mengen. *Skrifter utgit av Videnskapsselskapet i Kristiania,* I. *Mathematisk-naturvidenskabelig klasse,* no. 4.
1922–3. Einige Bemerkungen zur axiomatischen Begründung der Mengenlehre. *Wissenschaftliche Vorträge gehalten auf dem Fünften Kongress der Skandinavischen Mathematiker in Helsingfors vom* 4. *bis* 7. *Juli* 1922 (publ. Helsingfors, 1923), 217–232.
1933. Über die Unmöglichkeit einer vollständigen Charakterisierung der Zahlenreihe mittels eines endlichen Axiomensystems. *Norsk. mat. Foren. Skr.* ser. 2, no. 10, 73–82.
1934. Über die Nicht-charakterisierbarkeit der Zahlenreihe mittels endlich oder abzählbar unendlich vieler Aussagen mit ausschliesslich Zahlenvariablen. *Fundam. Math.* 23, 150–161.

Spector, C.
1955. Recursive well-orderings. *J. Symb. Logic,* 20, 151–163.

Tarski, A.
1927–8. See the Bemerkung der Redaktion in Skolem, 1934, p. 161.
1933. Der Wahrheitsbegriff in den formalisierten Sprachen. *Studia philosophica,* 1 (1936, tr. from Polish original, 1933). Engl. tr. in A. Tarski, *Logic, Semantics, Metamathematics,* Oxford University Press, Oxford, 1956.
1954–5. Contributions to the theory of models. *Proc. Akad. Wet. Amst.,* ser. A, 57, 572–578; 58, 56–64.

Tarski, A., Mostowski, A. and Robinson, R. M.
1953. *Undecidable Theories.* North Holland Publ. Co., Amsterdam.

Turing, A. M.
1936–7. On computable numbers, with an application to the Entscheidungsproblem. *Proc. Lond. Math. Soc.* (2), 42, 230–265. A correction, *ibid.* 43, 544–546.
1939. Systems of logic based on ordinals. *Proc. Lond. Math. Soc.* (2), 45, 161–228.

von Neumann, J.
1951. The general and logical theory of automata. *Cerebral Mechanisms in Behavior, The Hixon Symposium,* pp. 1–31 (editor, Jeffress, Lloyd A.). Wiley, New York.