$\vdash \exists! w P(x_1, \ldots, x_n, w)$ for the formulas numeralwise representing $[a/b]$, $\mathrm{rm}(a, b)$ and $\mathrm{rm}(c, (i' \cdot d)')$. However using *146a and *146b (with *123, *140, *141, *142a and *143a), this can be established also, and simpler representing formulas can be given equivalent to the former.

*178b. $\vdash Q(a, b, q) \sim \exists r(a=bq+r \;\&\; r<b) \vee b=q=0.$

*179b. $\vdash R(a, b, r) \sim \exists q(a=bq+r \;\&\; r<b) \vee (b=0 \;\&\; r=a).$

*180b. $\vdash B(c, d, i, w) \sim \exists v(c=(i' \cdot d)' \cdot v+w \;\&\; w<(i' \cdot d)').$

*178c. $\vdash \exists! q Q(a, b, q).$ *179c. $\vdash \exists! r R(a, b, r).$

*180c. $\vdash \exists! w B(c, d, i, w).$

LEMMA 18a. *The results* *(164)—*180c *and* (A)—(E) *of this section, excepting* *169 *and* *174a *when* t *is not a numeral,* *174b, *178b, c, *179b, c *and* *180b, c, *hold good for the formal system lacking Axiom Schema* 13 *but having as additional particular number-theoretic axioms the formulas of* *104—*107 *and of* *137 *or* *136 ("Robinson's system").

§ 42. Gödel's theorem. From a result of Presburger 1930, metamathematical proofs of consistency and completeness, and a decision procedure, can be given for the formal system with the formation rule and axioms for · omitted. (Cf. Example 2 § 79. Presburger deals with a classical system of the arithmetic of the integers, but Hilbert and Bernays 1934 pp. 359 ff. adapt his method to essentially the present classical system, and Joan Ross has verified that the adaptation works for the intuitionistic system as well.)

For the full system (or systems essentially equivalent to it), these questions proved to be very refractory. Consistency proofs by Ackermann 1924-5 and von Neumann 1927 lead to the result that the system is consistent under the restriction on the use of the induction postulate (Axiom Schema 13) to the case that the induction variable x does not occur free within the scope of a quantifier of the induction formula A(x). (Cf. Theorem 55 § 79. The restriction excludes e.g. our proofs of *105, *136 and *148.)

This situation was illuminated in 1931 by the appearance of two remarkable theorems of Gödel "on formally undecidable propositions of Principia Mathematica and related systems". We designate the first of these theorems, which entails the other as corollary, as "Gödel's theorem", although it is only one of a series of important contributions by its author. These two theorems, which became the most widely noted in the subject, bear on the whole program and philosophy of metamathematics.

The metamathematical results presented thus far in this book were reached along paths more or less suggested by the interpretation of the

system. These results of Gödel are obtained by a kind of metamathematical reasoning which goes more deeply into the structure of the formal system as a system of objects.

As is set forth in § 16, the objects of the formal system which we study are various formal symbols, formal expressions (i.e. finite sequences of formal symbols), and finite sequences of formal expressions. There are an enumerable infinity of formal symbols given at the outset. Hence, by the methods of § 1, the formal objects form an enumerable class. By specifying a particular enumeration of them, and letting our metamathematical statements refer to the indices in the enumeration instead of to the objects enumerated, metamathematics becomes a branch of number theory. Therewith, the possibility appears that the formal system should contain formulas which, when considered in the light of the enumeration, express propositions of its own metamathematics.

It will appear, on further study, that this possibility can be exploited, and with the use of Cantor's diagonal method (§ 2), a closed formula A can be found which, interpreted by a person who knows this enumeration, asserts its own unprovability.

This formula A bears an analogy to the proposition of the Epimenides paradox (§ 11). But now there is a way of escape from the paradox. By the construction of A,

(1) A means that A is unprovable.

Let us assume, as we hope is the case, that formulas which express false propositions are unprovable in the system, i.e.

(2) false formulas are unprovable.

Now the formula A cannot be false, because by (1) that would mean that it is not unprovable, contradicting (2). But A can be true, provided it is unprovable. Indeed this must be the case. For assuming that A is provable, by (1) A is false, and hence by (2) unprovable. By (intuitive) reductio ad absurdum, this gives that A is unprovable, whereupon by (1) also A is true. Thus the system is incomplete in the sense that it fails to afford a proof of every formula which is true under the interpretation (if (2) is so, or if at least the particular formula A is unprovable if false).

The negation $\neg$A of the formula is also unprovable. For A is true; hence $\neg$A is false; and by (2), $\neg$A is unprovable. So the system is incomplete also in the simple sense defined metamathematically in the last section (if (2) is so, or if at least the particular formulas A and $\neg$A are each unprovable if false).

The above is of course only a preliminary heuristic account of Gödel's reasoning. Because of the nature of this intuitive argument, which skirts so close to and yet misses a paradox, it is important that the strictly finitary metamathematical proof of Gödel's theorem should be appreciated. When this metamathematical proof is examined in full detail, it is seen to be of the nature of ordinary mathematics. In fact, if we chose to make our metamathematics a part of number theory (now informal rather than formal number theory) by talking about the indices in the enumeration, and if we ignore the interpretations of the object system (now a system of numbers), the theorem becomes a proposition of ordinary elementary number theory. Its proof, while exceedingly long and tedious in these terms, is not open to any objection which would not equally involve parts of traditional mathematics which have been held most secure.

We can give the rigorous metamathematical proof now, by borrowing one lemma from results of the next two chapters. Our numbering of the lemmas and theorems corresponds to the logical order.

In making use of the idea of enumerating the formal objects, practical considerations dictate that the indices of formal objects should be correlated to the objects by as simple a rule as possible. We can modify the above heuristic argument (inessentially) by using, rather than an enumeration in the usual sense, an enumeration with gaps in the natural numbers, i.e. a correlation of distinct natural numbers to the distinct formal objects, not all of the natural numbers being used in the correlation. We call this a Gödel numbering, and the correlated number of a formal object its Gödel number. (Sometimes separate Gödel numberings are given of the formal symbols, of the formal expressions, and of the finite sequences of formal expressions. If that is done, then when one speaks of a number as the Gödel number of a symbol, or of an expression, or of a sequence of expressions, in each case a different correlation is being referred to.)

Relative to any specified Gödel numbering, for any n which is the Gödel number of a formula, let "A_n" designate the formula. (For other n's, we need not define A_n.) We may write this formula A_n also as "$\mathrm{A}_n(a)$", showing the free variable a for use with our substitution notation (§ 18).

LEMMA 21. *There is a Gödel numbering of the formal objects such that the predicates $A(a, b)$ and $B(a, c)$ defined as follows are numeralwise expressible* (§ 41) *in the formal system.*

$A(a, b)$: a is the Gödel number of a formula (namely $\mathrm{A}_a(a)$), and b is the Gödel number of a proof of the formula $\mathrm{A}_a(\mathbf{a})$.

$B(a, c)$: a is the Gödel number of a formula (namely $\mathrm{A}_a(a)$), and c is the Gödel number of a proof of the formula $\neg\mathrm{A}_a(\mathbf{a})$.

Now let $\mathrm{A}(a, b)$ and $\mathrm{B}(a, c)$ be particular formulas which numeralwise express the predicates $A(a, b)$ and $B(a, c)$, respectively, for the Gödel numbering given by the lemma. The two formulas $\mathrm{A}(a, b)$ and $\mathrm{B}(a, c)$ could actually be exhibited, after we have the proof of the lemma (to be completed in § 52).

Consider the formula $\forall b \neg \mathrm{A}(a, b)$ which contains a and no other variable free. This formula has a Gödel number, call it p, and is then the same as the formula which we have designated "$\mathrm{A}_p(a)$". Now consider the formula $\mathrm{A}_p(\mathbf{p})$, i.e.

$$\mathrm{A}_p(\mathbf{p}): \qquad \forall b \neg \mathrm{A}(\mathbf{p}, b),$$

which contains no variable free. Note that we have used Cantor's diagonal method in substituting the numeral $\mathbf{p}$ for a in $\mathrm{A}_p(a)$ to obtain this formula.

To relate this to the preliminary heuristic outline, we can interpret the formula $\mathrm{A}_p(\mathbf{p})$ from our perspective of the Gödel numbering as expressing the proposition that $\mathrm{A}_p(\mathbf{p})$ is unprovable, i.e. it is a formula A which asserts its own unprovability.

In the metamathematical argument, the assumptions of the heuristic argument that the system should not allow the proof of either of the formulas A or $\neg$ A if false will be replaced by metamathematical equivalents. For the unprovability of A if false, this equivalent will be the (simple) consistency of the system (§ 28). For the unprovability of $\neg$A if false, we shall need a stronger condition called 'ω-consistency' which we shall now define.

The formal system (or a system with similar formation rules) is said to be *ω-consistent*, if for no variable x and formula A(x) are all of the following true:

$$\vdash \mathrm{A}(0), \quad \vdash \mathrm{A}(1), \quad \vdash \mathrm{A}(2), \ldots; \quad \vdash \neg\forall x \mathrm{A}(x)$$

(or in other words if not both $\vdash \mathrm{A}(\mathbf{n})$ for every natural number n and $\vdash \neg\forall x\mathrm{A}(x)$). In the contrary case that for some x and A(x) all of A(0), A(1), A(2), ... and also $\neg\forall x\mathrm{A}(x)$ are provable, the system is *ω-inconsistent.*

Note that ω-consistency implies simple consistency. For if A be any provable formula containing no free variables, writing it as "A(x)" where x is a variable, all of A(0), A(1), A(2), ... are provable (under our substitution notation § 18, each of these is simply A itself); and hence if the system is ω-consistent, $\neg\forall x\mathrm{A}(x)$ is an example of an unprovable formula (cf. § 28).

THEOREM 28. *If the number-theoretic formal system is (simply) consistent, then not $\vdash \mathrm{A}_p(\mathbf{p})$; and if the system is ω-consistent, then not $\vdash \neg\mathrm{A}_p(\mathbf{p})$.*

Thus, if the system is ω-consistent, then it is (simply) incomplete, with $A_p(p)$ *as an example of an undecidable formula.* (Gödel's theorem, in the original form.)

PROOF that, if the system is consistent, then not $\vdash A_p(p)$. Suppose (for intuitive reductio ad absurdum) that $\vdash A_p(p)$, i.e. suppose that $A_p(p)$ is provable. Then there is a proof of it; let the Gödel number of this proof be k. Then $A(p, k)$ is true. Hence, since $A(a, b)$ was introduced under the lemma as a formula which numeralwise expresses $A(a, b)$, $\vdash A(p, k)$. By ∃-introd., $\vdash \exists b A(p, b)$. Thence by *83a, $\vdash \neg\forall b \neg A(p, b)$. But this is $\vdash \neg A_p(p)$. This, with our assumption that $\vdash A_p(p)$, contradicts the hypothesis that the system is consistent. Therefore by reductio ad absurdum, not $\vdash A_p(p)$, as was to be shown. (We could also have contradicted the consistency by using ∀-elim. to infer $\vdash \neg A(p, k)$ from $\vdash A_p(p)$.)

PROOF that, if the system is ω-consistent (and hence also consistent), then not $\vdash \neg A_p(p)$. By the consistency and the first part of the theorem, $A_p(p)$ is not provable. Hence each of the natural numbers 0, 1, 2, ... is not the Gödel number of a proof of $A_p(p)$; i.e. $A(p, 0)$, $A(p, 1)$, $A(p, 2)$, ... are all false. Hence, since $A(a, b)$ numeralwise expresses $A(a, b)$, $\vdash \neg A(p, 0)$, $\vdash \neg A(p, 1)$, $\vdash \neg A(p, 2)$, ... By the ω-consistency, then not $\vdash \neg\forall b \neg A(p, b)$. But this is not $\vdash \neg A_p(p)$, which was to be shown.

We have given the original Gödel form of the theorem first, as the proof is intuitively simpler and follows the heuristic outline. Rosser 1936 has shown, however, that by using a slightly more complicated example of an undecidable formula, the hypothesis of ω-consistency can be dispensed with, and the incompleteness proved from the (simple) consistency alone. Consider the formula $\forall b[\neg A(a, b) \vee \exists c(c \leq b \,\&\, B(a, c))]$. This has a Gödel number, call it q. Now consider the formula $A_q(q)$, i.e.

$$A_q(q): \qquad \forall b[\neg A(q, b) \vee \exists c(c \leq b \,\&\, B(q, c))].$$

We can interpret the formula $A_q(q)$ from our perspective of the Gödel numbering as asserting that to any proof of $A_q(q)$ there exists a proof of $\neg A_q(q)$ with an equal or smaller Gödel number, which under the hypothesis of simple consistency implies that $A_q(q)$ is unprovable.

THEOREM 29. *If the number-theoretic formal system is (simply) consistent, then neither* $\vdash A_q(q)$ *nor* $\vdash \neg A_q(q)$; *i.e. if the system is consistent, then it is (simply) incomplete, with* $A_q(q)$ *as an undecidable formula.* (Rosser's form of Gödel's theorem.)