An Elucidation of Kurt Gödel's "On Formally Undecidable Propositions"

by Paul Obrecht

15 December 2000

(Winner of the 2001 St. John's College Mathematics Essay Prize)

This paper is an attempt to elucidate the important concepts in Kurt Gödel's 1931 paper, "On Formally Undecidable Propositions in *Principia Mathematica* and Related Systems I." It retains some of the technical considerations of Gödel's paper in order to explore the subtleties that can only be understood in terms of Gödel's notation, but attempts not to be inaccessible. We begin with Alfred Whitehead and Bertrand Russell's effort to systematize formal logic (in *Principia Mathematica*), to which Gödel's paper is a response.

A WORD ON FORMALISM

Whitehead and Russell's goal in *Principia Mathematica* was to propose a set of logical axioms and a purely mechanical method for deriving the consequences of these axioms. 'Mechanical' in this sense means that the rules of inference, i.e., the logic itself, can be codified as a set of valid symbolic manipulations in order to strip the system of all meaning, while still leaving open the possibility of determining whether or not a given sequence of steps is logically valid. As a result of making axiomatic all the rules of inference within the system, the consequences of the axioms are in some senses self-generating. That is, given a set of primitive signs, axioms, and generalized rules of inference, no human agent is required to accept a proof as plausible or not. The validity of a given proof simply falls out from the valid manipulations of the symbols and primitive signs. As a result, such a logical system depends in no way on the specific content of any given proposition. Such a system is called 'formalistic' because it attempts to derive the form of logic independent of its content.

From the ancient point of view in which reason is seen as a uniquely human faculty, the formalistic approach to logic deprives it of any connection with the human mind. But this approach removes the pitfalls with which the human faculty of reason is fraught: invalid leaps of reasoning and willing acceptance of untrue but plausible propositions based on coincidences in the content of the propositions. It does not seem that Whitehead and Russell, or any logical formalists for that matter, claim that the logical systems they derive correspond in any way to the operation of the human mind. In fact, *because* they do not correspond to the operation of the human mind, their validity is said to be absolutely decidable. Or so it was thought.

GÖDEL AND THE SYSTEM P

In order to provide a suitable groundwork for the proof, Gödel chooses what he calls the system P as a context. In his words, "P is essentially the system obtained when the logic of [*Principia Mathematica*] is superposed upon the Peano axioms" (90). In a footnote, Gödel says that the inclusion of the Peano axioms makes the proof more workable and is dispensable in principle. At the beginning of Section 2, Gödel lists the primitive signs of the system: "$\sim$" (not), "$\vee$" (or), "Π" (for all), "0"(zero), "f" (the successor of), "(", ")" (parentheses), and variables of type 1, 2, 3, ...

The Notion of 'Primitives'

These primitives should be addressed in some detail. A large part of Russell and Whitehead's project was to banish self-reference from set theory because of the numerous paradoxes it introduces. This goal led them to devise the hierarchy of variables and functions of 'types'. Sets become ill-defined and paradoxical when they contain themselves or other sets of equal rank on the hierarchy of types. The sole possibility for working around such paradoxes, thought Russell and Whitehead, is in explicitly defining a 'set' only to contain variables and other sets of lower type.

Why is self-reference such a problem? Some examples are in order. The following are self-referential sets (that is, they contain themselves): the set of all sets, the set of all things except Kurt Gödel, and so on. At first glance, these may seem like trivial cases worthy only to be ignored. But a problem arises when we consider W, the set of well-defined (nonself-referential) sets. If we say that W is well-defined, then it does not contain itself. But then W cannot be well-defined, since it is not a member of the set of well-defined sets. Because the formalization of logic is inherently a self-referential task (reasoning about reasoning), extreme care must be taken, and the existence of such trivial cases leading directly to paradox suggests that nothing can ever be known about logic as a concept independent of specific instances of its use. In order to get off the ground at all, Russell and Whitehead had to define sets as they did. Gödel designates primitive n-type variables only to follow Russell and Whitehead's system.

The other primitive signs are more straightforward. We have seen that all other logical connectives can (or must) be defined in terms of $\sim$ and $\vee$, so leaving these two as primitives is no hindrance. Similarly, all other quantifiers can be defined in terms of Π; again, no hindrance. All numbers can be derived through applying the successor function an arbitrary number of times to 0, so these two concepts suffice to determine all numbers. Lastly, the parentheses determine simply through their existence in a proposition the order of operations. Their function is so self-evident and primary that it need not be defined in terms of anything else.

The notion of primitives in a formal system is a bit curious. How can such important notions, on which all other notions in the system are based, remain undefined? The answer to this question stems from the formalist philosophy. Since the goal of formalism is to extract all meaning from logic, we would find a contradiction from the outset if we ascribed meaning to certain symbols. Thus the primitives function as indicators of valid and invalid symbol manipulation, and what must be preserved through any sequence of valid symbol manipulations is that, in the end, the statements we derive must be *interpretable*; that is, the symbols, their manipulations, and statements containing them must correspond to meaningful concepts. Yet we do not want these concepts to cloud our certainty at any step in the process, so we withhold them until the end. Thus the formal system provides certain meaningless primitives that serve in some senses only to solve the 'bootstrap' problem of providing something by which all other symbolic manipulations can be defined. For example, knowing nothing about what $\sim$ and f 'mean', Gödel's Axiom schema I-1 tells us that $\sim (fx_1 = 0)$. We can interpret this statement to mean that 'the successor of no natural number is zero', but this interpretation comes after the fact; Axiom schema I-1 simply postulates a valid symbol manipulation that in the end will correspond to a meaning, but whose logical consequences can be determined without any dependence on the 'meaning' of $\sim$, f, x_1, or 0.

The Axioms

Much of the notational complexity in Gödel's paper is necessary only because of the particular axiomatic system chosen as a framework for the proof. Gödel states explicitly in the introduction to his paper that the incompleteness demonstrated for the system of *Principia Mathematica* (*PM*) is not at all unique to the system *PM*. "This situation is not in any way due to the special nature of the systems that have been set up but holds for wide class of formal systems" (88). The proof itself holds rigorously only for the axiomatic system *PM*, but by arguing through analogy, Gödel claims that nothing in his Theorem VI depends uniquely on the particular set of axioms chosen. Rather, he claims, any consistent formalized set of axioms meeting certain conditions will exhibit the same kind of undecidability. It appears that Gödel intended to write a follow up paper generalizing his results for any system, but the prompt acceptance of his argument made such plans unnecessary.

So Gödel goes on to list the axioms of the system P. It is noteworthy that Axiom schemata I and II concern only straightforward symbol manipulation, functionally prescribing certain traits of the primitives. For instance, I-2 can be interpreted to mean that two numbers must be equal if their successors are equal, yet we must remember that I-2 merely postulates a symbolic consequence of the f function. I-3 is essentially a notation whose interpretation corresponds to the validity of proof by induction: If a function holds for zero and, for any number, that it holds for this number follows from the fact that it holds for the previous number, then it universally holds for all numbers. This is grammatically awkward, but the notation unambiguously postulates a symbolic consequence of Π, f, and variables of consecutive types written in a certain order. Axiom schemata II provide similar statements delineating functional attributes of the symbols $\vee$ and $\supset$ (note that $\supset$ can be defined in terms of the primitives $\sim$ and $\vee$).

Axiom schemata III give us reason to pause. It might appear at first glance that these two axioms are metalogical, that is, that they define what it means for a formula to be interpreted as true for a generalized class of variables. Given Gödel's reliance on recursiveness as the only infallible method of proof, we should rightly suspect that any valid proof of generalization should be recursive, which Axiom schemata III-1 and 2 clearly are not. On reflection, though, these statements are seen merely to prescribe consequences of a generalization. That is to say, if the formula a holds for all v, it follows that $\mathrm{Subst}\, a\binom{v}{c}$ holds. Similarly with III-2: if either b or a hold for all v, then either b holds or $v\Pi(a)$ holds (for specific conditions of a and b). Note that these two axioms in combination allow us to restate $v\Pi(b \vee a)$ without using the Π notation, carrying out only symbolic manipulations: $v\Pi(b \vee a) \supset b \vee \mathrm{Subst}\, a\binom{v}{c}$.

Axiom schemata IV and V are far more specific than the others. Axiom schemata IV asserts that a u exists such that the generalization of v in u is equivalent to some formula a that does not contain u as a free variable. This axiom seems somewhat nebulous, but seems to say that there is at least one generalizable formula in the system P and that, once again, that formula is expressible without using the Π notation. Axiom schemata V, $x_1\Pi(x_2(x_1) \equiv y_2(x_1)) \supset x_2 = y_2$, in Gödel's words, "states that a class is completely determined by its elements" (92). We now insert a brief parenthetical consideration.

Bottom-up Reasoning

This type of reasoning is abundant in formalist philosophy. Without recourse to meaning, asserting that two statements are equivalent, or stating that two sets are equal, must rely on the end result. We could try to intuit a rule that determines membership in a given class, but using such reasoning leads to ambiguities that can derail any formal inquiry. For example, consider the case of determining the set of all the trustworthy people in a given population. Devising a set of overarching, top-down rules against which each potential member of the set must be measured introduces innumerable problems. We find ourselves reasoning as follows: "All trustworthy people are honest and have no secret loyalties. They also have the means and are eager to reciprocate beneficence." Person X at times demonstrates his honesty, and he has recently reciprocated a kind act. It is impossible to determine if he has secret loyalties. Is he trustworthy?" The only way to attain surety, at least in a formalistic sense, is to reason from the bottom up: to decide in the case of each person independently whether or not he is trustworthy. The set of trustworthy people is then simply determined in sum by who is deemed to be trustworthy. It follows that we can, if we choose, determine generalized membership requirements of the set by inferring them from characteristics of its members, rather than undertake the hopeless task of determining who matches the arbitrary prescribed requirements for membership in the set. Note that the possibility of inferring the general membership rule is entirely for the sake of human curiosity; in Gödel's system (and in any formal system) the set is no less well-defined if we consider only its elements as a definition. In fact, as the above example suggests, the set is unambiguously defined only if it is defined by its elements.

This is not the only example of such reasoning. We use the same sort of reasoning when we claim that two logical statements are equivalent because their truth tables are identical. On a side note, it can *only* be because one does not intuitively accept such reasoning that he would hesitate to accept Taylor's assertion that

$$\cos x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!}$$

even though every point that satisfies one satisfies the other. Such intuitive hunches can be seen clearly to cloud the process of human reasoning, highlighting the need for a formal and abstract system of logic if we are to understand what it is, independent of specific instances of its use. That even formalized, concrete, self-generating systems such as *PM* reach limits in reasoning about themselves presents a wholly different problem, which will be addressed at the appropriate time.

Immediate Consequence

The only other concept defined by Gödel before his introduction of the system of "Gödel numbers" is that of *immediate consequence*. This concept will be necessary in the proof of Theorem VI. It is simply defined: "A formula c is called an *immediate consequence* of a and b if a is the formula $(\sim b) \vee c$, and it is called an *immediate consequence* of a if it is the formula $v\Pi(a)$, where v denotes any variable" (92). In other words, $(\sim b \vee c).b \supset c$ in the first case, and $a \supset v\Pi(a)$ in the second. Here we have a notational definition of a metamathematical concept, yet defining it in terms of the notation makes it absolutely unambiguous. It is clear that a proper definition of *immediate consequence* is essential to any concept we would form of *proof*, because any formal proof must proceed from step to step in an unambiguous manner, and each step must be an immediate consequence of the prior step. Gödel gives a more exact definition of this relation in his list of 45 definitions; his Theorem V, the last pit stop on the road to undecidability, depends heavily on this concept.

And that's it. Gödel's proof of the existence of undecidable theorems follows directly from these few primitives, axioms, and rules of inference, which are supplied by *PM* itself. Russell and Whitehead probably never imagined that the raw materials for proving the limitations of their own system, and all of axiomatic logic, were embedded from the beginning in the very notions that define it. In order to make the limitations apparent and demonstrable, though, Gödel had to make one fundamental modification:

GÖDEL NUMBERS

Gödel's genius is most apparent here. The problem to which the system of Gödel numbering is the solution is a tricky one. In order to rigorously demonstrate anything about provability, completeness, or decidability, Gödel must be able to make statements about metamathematical concepts from within the axiomatic system under investigation. But *PM* deals only with the step-by-step processes involved in formal deduction and proof. There is no way for the system to ask questions or to make statements about itself, which is in line with Russell and Whitehead's goal of removing self-referential aspects from the formalization of logic. Yet by designing a system in which every possible statement is assigned a unique number, Gödel is able to develop a means by which metamathematical notions "become notions about natural numbers or sequences of them; therefore they can (at least in part) be expressed by the symbols of the system *PM* itself" (88). Namely, by this method, Gödel is able to define the notions "formula," "proof array," and "provable formula" as consequences of the axioms of *PM*. Interpreting logical primitives as natural numbers is valid because the nature of undefined primitives allows for any interpretation, given certain constraints. *PM* demonstrates that the natural numbers meet these constraints.

Gödel's Method

The symbols are assigned to the odd numbers ≤ 13, and variables of type n are each assigned a number p^n where p is a prime number greater than 13. Every possible finite arrangement of primitive signs in *PM* is thus assigned a unique finite sequence of natural numbers. (The rules of deduction are manifestly able to weed out the nonsensical ones.) Yet it is not sufficient to stop here, because it is not possible to perform the mathematical operations corresponding to metamathematical analysis on sequences of numbers. Gödel must now map these finite sequences of natural numbers onto the natural numbers again in order that every possible statement in *PM* corresponds to a unique number. Knowing that every natural number has a unique prime factorization, he composes the number corresponding to each possible statement in *PM* by raising the sequence previously determined to the powers of the ordered primes. For example, "(x_1)" corresponds to the sequence $\{11, 17, 13\}$, which corresponds in turn to the natural number $2^{11} \cdot 3^{17} \cdot 5^{13} \approx 3.23 \times 10^{20}$. (This approximation may be misleading, for the natural number to which "(x_1)" corresponds is precisely determined by this method. The approximation is given only to indicate the staggering magnitude of the numbers obtained for even the simplest sequence of primitives.)

As a result of this method, proof arrays correspond to finite sequences of natural numbers, and merely by using rules of arithmetic, it can be determined whether, for example, a statement is an immediate consequence of the previous statement, or if it is an axiom, or if it is a provable formula. This claim results from Gödel's highly technical definitions of the relations R and R' on page 92. In short, he denotes by $\Phi(a)$ the natural number assigned to a primitive sign, or sequence of primitive signs, a. For example, if a_1 is (x_1), then $\Phi(a_1) = 2^{11} \cdot 3^{17} \cdot 5^{13}$. Now, given a particular set of as, we wish to define a relation R between them. Such relations could be almost anything, but would most usefully correspond to the metamathematical notions of interest in this proof. Gödel associates the relation $R'(x_1, x_2, \ldots, x_n)$ with $R(a_1, a_2, \ldots, a_n)$ if and only if every x_i (natural number) corresponds to an a_i (sequence of primitive signs) according to the rule $x_i = \Phi(a_i)$ for all $1 < i < n$.

This begs to be clarified. We have a logical relationship R between statements a. This relationship can be defined verbally, as in "sequence of formulas" or "provable formula." We want to ensure that, after the transition to natural numbers, the logical relationship can *still* be expressed between the numbers x_i that correspond to the statements a_i. Of course, it is antithetical to express R' verbally, because the goal of making the transition to natural numbers is to express R' as an arithmetic relationship that can be analyzed within the system *PM*. This is where $\Phi(a)$ comes in. Gödel says that if R holds, and $x_i = \Phi(a_i)$ for all $1 < i < n$, then R' can be associated with R. R' is thus an arithmetic relation between natural numbers, or classes of natural numbers, that corresponds to the logical relationship R. Gödel expresses the *interpretation* of certain useful arithmetic relations R' in small capitals: SEQUENCE OF FORMULAS or PROVABLE FORMULA.

The first question in the reader's mind should be this: What guarantees that *every* logical statement R corresponds to an R'? Relying superficially on the words of the above paragraph, the writer would respond, "the function $\Phi(a)$, of course." This is not a satisfactory answer, because we know nothing about such a function, other than that it must exist in order to prove what we want to prove; we know nothing about its qualities. Yet it possesses a quality fundamental to the nature of the proof, and by extension, to all formal axiomatic systems.

Recursiveness

In the end, we will see that if $\Phi(a)$ can be shown to be recursive, then there will be no problem drawing an absolute correspondence between logical statements and the mathematical statements that correspond to them. Gödel first defines what it means for a number-theoretic function $\varphi(x_1, \ldots, x_n)$ to be *recursively defined in terms of* two other number theoretic functions $\psi(x_1, \ldots, x_{n-1})$ and $\mu(x_1, \ldots, x_{n+1})$. This definition in principle closely mirrors that of Axiom schemata I-3. In short, $\varphi(0)$ must equal some bootstrap function ψ, and $\varphi(k+1)$ must equal some function $\mu(k, \varphi(k))$ determined by the previous values of both k and $\varphi(k)$. This is a special case of what it means for a function to be *recursive*: a function φ is recursive if it is itself definable by a finite sequence of functions $\varphi_1, \varphi_2, \ldots, \varphi_n$ that ends with φ, each of which intermediary functions φ_k is recursively defined by the criteria above. Our relation R, then, is said to be recursive if $R(x_1, \ldots, x_n) \sim [\varphi(x_1, \ldots, x_n) = 0]$. This is the gem. Essentially, we now have means to guarantee that our function R is indeed recursive, by requiring that it be logically equivalent to a recursive function φ, such that φ, as a function of the same natural numbers of which R is a relation, is equal to zero. It is essential that this is an equivalence: Any recursive function φ can now be said to *represent* some logical relationship R, and conversely, every recursive relationship R can be represented as a function φ of natural numbers.

This equivalence demonstrates the importance of recursiveness as a fundamental notion in logic. Aside from being the only mechanical method of proof (i.e., a formalistic method relying in no way on the meaning of given propositions), it demonstrates the full extent to which a logical system can reason about itself. Any concept that can be defined recursively becomes raw material upon which the system can operate, in theory, so even metamathematical notions fall within reach of the formal logic. Seen in this light, the Gödel numbering system becomes merely a way to phrase recursive metamathematical notions in the language of the logical system. The real transaction occurs in recursiveness. Essentially, Gödel's Theorem V says just this, with the addition of the notions of PROVABLE FORMULA, Bew(x), and SUBST $Sb(a\,{}^{b}_{c})$.

After the definition of *recursive*, Gödel proposes four theorems that follow from the definition. In short, these theorems demonstrate that the following metamathematical notions can be shown to be recursive:

$\overline{R}$	R does *not* hold,
$\vee$	or,
$=$	equals,
$(\mathrm{E}x)(y)$	there exists an x such that y holds,
(x)	for all x,
$\varepsilon x(y)$	the least x such that y holds

Having demonstrated the recursiveness of all necessary metamathematical relations in *PM*, and having demonstrated that every recursive metamathematical relation R has an associated arithmetic relationship R', Gödel proceeds to define 45 arithmetic operations, all of which are recursive, which operate on the numbers corresponding to metamathematical statements in the Gödel numbering system. Without spending an inordinate amount of energy delineating the definitions, we note that the method of defining these arithmetic operations makes abundantly clear the formalistic notion of definition, for each relation is defined only by symbolic manipulation. For example, the definition of Neg(x) consists of appending the number associated with the relation $\sim$ to the front of formula x. In order to ensure that the entire formula is 'negated', the operation associated with enclosing a formula in parentheses is performed on x prior to the 'negation'. There is no meaning here, no interpretation of the notion of negation. The concept is defined only by the interaction of the primitive signs, or more accurately, of the numbers associated with the primitive signs.

The definitions proceed methodically from simple concepts to complex, and over the course of the list, we see the emergence of recursive arithmetic operations corresponding to what we would consider very elusive concepts in a non-formal system: AXIOM, IMMEDIATE CONSEQUENCE, PROOF ARRAY, and so on. The culmination of the list is Bew(x) (x IS A PROVABLE FORMULA), which by all appearances is defined recursively, though the results of the proof will demonstrate otherwise.

THEOREM V AND ω-CONSISTENCY

As mentioned above, the content of Theorem V has been suggested earlier in the paper less rigorously, and without two necessary specifications that will allow Gödel to prove the existence of undecidable propositions. Essentially, Theorem V proves that the existence of a recursive relation $R(x_1, \ldots, x_n)$ implies the existence of an equivalent recursive arithmetic operation. The arithmetic operation says that the relation r, for whose n free variables we substitute the numerals corresponding to the arguments $(x_1, \ldots, x_n)$, is provable. Theorem V demonstrates that the converse holds as well, that is, that the negation of R implies that the negation of the above arithmetic relation is provable.

However, foreseeing the results of Theorem VI, Gödel must define a concept called ω-consistency in order to achieve the sort of contradiction necessary for the *reductio ad absurdum* proof. Given a class κ of formulas, he defines the set $\mathrm{Flg}(\kappa)$: "The smallest set of FORMULAS that contains all FORMULAS of κ and all AXIOMS and is closed under the relation IMMEDIATE CONSEQUENCE. κ is ω-consistent if there is no CLASS SIGN a such that $(n)[Sb(a\,{}^{v}_{Z(n)})\ \varepsilon\ \mathrm{Flg}(\kappa)]\ \&\ [\mathrm{Neg}(v\ \mathrm{Gen}\ a)]\ \varepsilon\ \mathrm{Flg}(\kappa)$, where v is the FREE VARIABLE of the CLASS SIGN a" (98). In other words, κ is ω-*in*consistent if it implies the following:

1. a holds for any particular numerical substitution into v,
2. in general, a holds for no v.

These two statements are not directly contradictory, but no one can reasonably claim that they can both hold at the same time. It is not the case, as in inductive reasoning, that we can prove individual cases without being able to establish a general rule. Instead, ω-inconsistency says that a holds for every particular substitution, yet a expressly *does not hold* in general. Theorem VI says that a particular substitution into a particular formula a will prove the ω-inconsistency of a class of FORMULAS assumed to be ω-consistent. This is the sort of direct contradiction necessary for a constructive proof. That Theorem VI is proved for an ω-consistent class of FORMULAS is a subtle distinction with a far-reaching implication.

The assumption of an ω-consistent κ generalizes the results of the proof to apply to any finite extension of the axioms of *PM*. That is, regardless of the number of (expressly noncontradictory) formulas in κ, we will always arrive at this contradiction, because $\mathrm{Flg}(\kappa)$ contains every possible statement in *PM and* every possible implication of κ. The FORMULAS in κ meet the criteria for axioms because (1) by supposition κ contains no contradictory statements, and because (2) the set of consequences of κ is finite. Statement (2) is true because κ is closed under the relation IMMEDIATE CONSEQUENCE. In this light, they differ from potential axioms in no way. This prevents us from simply asserting $17\,\mathrm{Gen}\,r$ or its negation to be a new axiom. Despite doing so, we would still find another unprovable formula, and to assert that to be an axiom would yet produce another unprovable formula, *ad infinitum*. But before discussing the implications of Theorem VI any further, we should look at the proof itself.

THEOREM VI

In Equation (8.1) Gödel defines the relation $Q(x, y) \equiv \overline{x\ B_\kappa\ [Sb(y\,{}^{19}_{Z(y)})]}$. He points out that the relation Q must be recursive because everything on the right-hand side is recursive. Since Q is recursive, there is an associated arithmetic relation q that, with appropriate substitutions, is PROVABLE. But the definition of Q gives it curious properties. Q holds for x and y (associates x and y as a pair) if x is *not* the proof array of the formula obtained by substituting the numeral representing y into itself.

This is, from one point of view, nothing but an underhanded trick to find a loophole in the restrictions in *PM* due to the theory of types. In other words, it would be invalid to substitute y into itself, because the resulting statement would be self-referential. But substituting the *numeral corresponding to* y into itself is not unlike handing y a 3x5 card with the words "statement y" written on it. y simply holds this card up and points to it whenever it encounters $Z(y)$ in the formula resulting from the substitution. Self-reference is thus completely removed from the system; y refers to an external representative entity in order to make the appropriate substitutions. This is the genius in the Gödel numbering system: layers and layers of *representations*. In this way, Gödel is able to refer to a logical relation by its associated arithmetic relationship, and in turn, is able to refer to the arithmetic relationship by the *numeral* associated with it. Because the arithmetic relations are nothing but numbers, they can be ordered, and each is represented by a *numeral*, the unique symbol corresponding to each number. Here we see the similarity with Cantor's diagonal proof. We have an ordered list of statements $\mathrm{Flg}(\kappa)$ supposed complete and containing every possible consequence of κ, yet we can find a consequence of κ that *cannot* be a consequence without rendering it inconsistent.

We have seen that the relation $Q(x, y)$ was chosen in order to bypass the problem of self-reference. Its associated arithmetic relation q has two FREE VARIABLES, representing x and y. Gödel carefully chooses to define p and r such that, after the appropriate substitutions, he arrives at a FORMULA with no free variables, $17\,\mathrm{Gen}\,r$. This statement says, "For all FREE VARIABLES 17, the substitution of the NUMERAL representing $17\,\mathrm{Gen}\,q$ into q holds." $17\,\mathrm{Gen}\,r$ is thus the statement q bound in specific ways at its two free variables; our logical relation Q has become a SENTENTIAL FORMULA, which in theory, can be proved or disproved. That is, we should be able to show without contradiction that it is either a member of $\mathrm{Flg}(\kappa)$ or not. That either possibility leads to contradiction (implying the ω-inconsistency of κ) is the crux of Theorem VI.

The definition of $Q(x, y)$ set out in the beginning says something else as well. Q relates its arguments in such a way that $17\,\mathrm{Gen}\,r$ is the statement whose interpretation says "I am not provable in κ if κ is consistent." If κ is consistent, the statement is true, and therefore not provable. But if the statement can be proved, then the system must be inconsistent (which means that it must contain axioms that contradict each other). And this paradox is an unambiguous logical consequence of the axioms of PM. Dismissing the method of arriving at such a truth as underhandedly finding a loophole in *PM*, though, is to miss the significance of Gödel's proof, for it shows that diligent precautions to prevent self-reference in a formal system can always be circumvented. In other words, self-reference is inescapable, and realizing this truth suggests a great deal about the notions of truth and provability. It suggests that provability is a weaker notion than truth, and rigorously demonstrates that there are true statements which *cannot* be proved, because self-reference cannot be avoided in any formal system, no matter how carefully it is constructed. Any system that is complex enough to represent a robust logical system is complex enough to reason about itself.