Notre Dame Journal of Formal Logic
Volume 41, Number 1, 2000

NONSTANDARD MODELS AND KRIPKE'S PROOF OF THE GÖDEL THEOREM

HILARY PUTNAM

Abstract This lecture, given at Beijing University in 1984, presents a remarkable (previously unpublished) proof of the Gödel Incompleteness Theorem due to Kripke. Today we know purely algebraic techniques that can be used to give direct proofs of the existence of nonstandard models in a style with which ordinary mathematicians feel perfectly comfortable—techniques that do not even require knowledge of the Completeness Theorem or even require that logic itself be axiomatized. Kripke used these techniques to establish incompleteness by means that could, in principle, have been understood by nineteenth-century mathematicians. The proof exhibits a statement of number theory—one which is not at all "self referring"—and constructs two models, in one of which it is true and in the other of which it is false, thereby establishing "undecidability" (independence).

Our subject will be elementary number theory, that is, the theory of the nonnegative integers as formalized in standard (first-order) quantificational logic. As primitives we will simply take the primitive recursive predicates. (Alternatively, we could have just taken $+$ and $\cdot$, or any other set of basic predicates or functions from which all the primitive recursive predicates can be defined.) As axioms we will take the standard first-order version of the Peano Axioms.

One of the surprising facts which was not noticed by nineteenth-century mathematicians, but which was observed after Gödel proved the completeness of first-order logic in 1930, is the fact that Peano Arithmetic has *nonstandard models*. By a nonstandard model is meant a model in which, in addition to the integers 0, 1, 2, ... there are also (viewed from the outside) "Infinite integers". I say "viewed from the outside" because within Peano Arithmetic itself there is no way to single out these nonstandard elements or even to say they exist: It is only in the set-theoretic mathematical language in which we prove the existence of the nonstandard model that we can say

Received June 5, 2001; printed July 15, 2002
2001 Mathematics Subject Classification: Primary, 03; Secondary, 03C62,03H15
Keywords: models of arithmetic, nonstandard models of arithmetic

that there are "foreign elements" in the model, elements other than the "real" natural numbers.

These models exist not only for elementary number theory but for mathematics as a whole (as represented by, say, Zermelo-Fraenkel set theory). In the metalanguage in which we talk about such a model we can distinguish between "finite" and "infinite" real numbers, even though within the object language it is a theorem that "all real numbers are finite". And there are reciprocals of infinite real numbers in the model, that is, *infinitesmals*. Robinson showed, in fact, that by using such models one can carry out Leibniz's dream of a true calculus of infinitesmals, and the resulting branch of mathematics, which has been called Nonstandard Analysis, already has significant applications in many areas—to the theory of Lie groups, to the study of Brownian motion, to Lebesgue integration, and so on. What gives the subject its power is that, because the "infinite numbers" all belong to a model for the standard (finite) integers and real numbers, a model within which they are not distinguished by any predicate of the language from the standard numbers, we are guaranteed from the start that they will obey all the laws that standard numbers obey.

Strangely enough, it is only in recent years that various workers, the most famous being Paris and Harrington, have begun to use the existence of nonstandard models to give independence proofs in number-theory itself (see [1]). The very existence of independent (or "undecidable") propositions of elementary number theory was proved by Gödel in 1934 by syntactic, not model theoretic, means. The proposition proved independent by Paris and Harrington is a statement of graph theory (a strengthened version of Ramsey's Theorem). What I shall show you is a simplified successor to the Paris and Harrington theorem, a successor which proves the independence of something resembling Gödel's undecidable proposition *by purely model theoretic means*.

Because this theorem does not aim at establishing the independence of a statement which is nearly as complicated as the proposition Paris and Harrington wished to prove independent, the proof is much simpler than theirs. Also, because the independence proof is semantic rather than syntactic, we do not need to arithmetize the property "x is a proof of y", as Gödel did for his proof. We do not need the famous predicate $Bew(x)$ (x is the Gödel number of a theorem), or the famous self-referring sentence which is true if and only if its own Gödel number is not the Gödel number of a theorem. In short, I am going to show you a *different* proof of the Gödel theorem, not just a different *version* of Gödel's proof.

As the title of this paper indicates, the honor of inventing this proof belongs to Kripke. He has not published it yet, in part because he is still investigating what can be done using the "bounded ultrapower construction" which appears in the most constructive version of the proof. The version I am going to show you is a quicker and less constructive version that is also due to Kripke.

I mentioned at the beginning of this paper that the existence of nonstandard models was first observed as a corollary to Gödel's 1930 work on the completeness of first-order logic. But today we know techniques which are purely algebraic rather than logical—the techniques of ultrapower construction—which can be used to give direct proofs of the existence of nonstandard models in a style with which mathematicians who are not trained logicians feel perfectly comfortable. In fact, these techniques do not even require knowledge of the Gödel Completeness Theorem, or even require that logic itself be axiomatized. In short, the proof I am about to show you is one that establishes independence by means that could, in principle, have been understood by nineteenth-century mathematicians. We will exhibit a statement of number theory—one which is not at all "self referring"—and construct two models, in one of which it is true and in the other of which it is false, thereby establishing "undecidability" (independence).

Consider a finite monotone-increasing series s of natural numbers, say $182, 267, 349, 518, \ldots, 3987654345$. And consider a formula A of number theory, say $(x)(Ey)Rxy$ (with primitive recursive R). (In what follows I shall identify the series s with its Kleene Gödel number where convenient.) I shall say s fulfills A if the second player (the "defending player") has a winning strategy in the game I shall describe.

1. The Game G

The game is played as follows. The first player (the "attacking player") picks a number less than the length of the given sequence s, say 3. The sequence s is examined to determine the third place in the sequence (to determine "$(s)_2$", in the Kleene notation, since the members of a sequence with Gödel number s are $(s)_0, (s)_1, (s)_2, \ldots$ in this notation). The same player (the attacker) now picks a number less than this number (less than 349, in the case of this example). Let us suppose he picks 17. We assume the number picked by the first player was less than the length of the sequence (otherwise the first player has lost). If so, the second player (the "defending player") gets to look at the next number in the sequence (at $(s)_3$ or 518, in the case of the example). He must pick a number less than this number (less than $(s)_{n+1}$, if the first player picked the place $(s)_n$). Let us suppose he picks 56. We now evaluate the statement $R(17, 56)$ (the statement $R(n, m)$, where n is the number picked by the first player and m is the number picked by the second player). Since R is primitive recursive, this can be done effectively. If the statement is *true* the defending player has won; if *false* the attacking player has won.

The statement that a sequence s fulfils this statement A (that there is a winning strategy for the defending player) can itself be written out in number theory, as follows:

(I) $(i \leq \text{length}(s) - 1)(n \leq (s)_{i-1})(Em \leq (s)_i)Rnm$

Similarly, if we are given a statement A with four, or six, or however many alternating quantifiers in the prefix, we can define "s fulfills A" to mean that there is a winning strategy for the defending player in a game which is played very much like the game **G**: a game in which the attacking player gets to choose a new place in the sequence, each time it is his turn to play. The attacking player must also choose a number less than $(s)_{i-1}$ where $(s)_{i-1}$ is the number in the position he chose in the sequence. (N.B. The number in the ith position is called '$(s)_{i-1}$' and not '$(s)_i$' because Kleene—whose notation I am employing—calls the first position '$(s)_0$' and not '$(s)_1$'.) Each time he plays, the attacking player has to choose a place which is to the right of the place in the sequence he chose before (unless it is his first turn to play) and not the last place in the sequence (unless he has no legal alternative, in which case he loses), and the defending player must then pick a number less than $(s)_i$ (less than the number in the next place in the sequence). The game ends when as many numbers have been chosen as there are quantifiers in the prefix of the formula. (We assume all formulas are prenex, and that quantifiers alternate *universal, existential, universal, existential, ...*) The numbers chosen are then substituted for the variables in the matrix of the formula A in order (first number chosen for x_1, second number chosen for x_2, and so on, where x_1 is the variable in the first argument place, x_2 the variable in the second argument place, and so on). The resulting primitive recursive statement is evaluated and, as before, the defending player wins if the statement is true and the first (attacking) player wins if the statement is false. Once again, for any fixed formula A we can easily express the statement that s fulfills A arithmetically (primitive recursively in s). And for any fixed recursively enumerable sequence of formulas $A_1, A_2 \ldots$, the statement that s fulfills A_n can be expressed as a primitive recursive relation between s and n, say $\mathit{Fulfills}(s, A_n)$. Note that we can also speak (by an obvious extension) of an ordinary infinite monotone increasing sequence "fulfilling" a formula (this means that if one picks any number less than a given number in the sequence to be the value of the first universal quantifier, it is always possible to pick a number less than the next place to the right in the sequence to be a value for the succeeding existential quantifier, so that, no matter what number less than the number in an arbitrarily selected place still farther to the right in the sequence one picks for the *next* universal quantifier . . ., it is possible to pick a number less than the number in the place in the sequence immediately to the right of the last "universal quantifier place" chosen for the last existential quantifier so that the statement A comes out true). And note that a statement which is fulfilled by an infinite monotone increasing sequence is *true*. (Since the restriction that one must pick numbers as values for the universal quantifiers which are bounded by the numbers in the sequence is, in effect, no restriction on the "attacking player" at all—the numbers in the sequence get arbitrarily large, so he can pick any number he wants by going out far enough in the sequence!)

Henceforth, we shall confine attention to sequences with the following two properties (call them *good sequences*):

1. The first number in the sequence is larger than the length of the sequence.
2. Each number in the sequence after the first is larger than the square of the number before. (This is to ensure that the sum and product of numbers $\leq (s)_i$ are $\leq (s)_{i+1}$.)

Finally (this is the last of the preliminaries!) let $P_1, P_2, P_3, \ldots$ be the axioms of Peano Arithmetic.

We will say that a statement is n-fulfillable if there is a good sequence of length n which fulfills the statement. The following is the statement which we shall show to be independent of Peano Arithmetic:

(II) For every n and every m, the conjunction of the first m axioms of Peano Arithmetic is n-fulfillable.

or (this is easily seen to be equivalent)

(III) For every n, the conjunction of the first n axioms is n-fulfillable.

What does this actually say? Well, if for "n-fulfillable" we substitute "fulfilled by an increasing infinite sequence", (III) is the statement that Peano Arithmetic (or whatever consistent extension we take P to be) is *true*. Of course, truth is not expressible in P itself. So (III) is a kind of weak substitute for the statement that Peano Arithmetic is a true theory. In fact, "n-fulfillable" is a Σ_1 property, so the above is only a Π_2 sentence. What it says, however, is that Peano Arithmetic has a weak kind of *correctness*. From here on I shall only outline Kripke's proof; the details are not hard to verify.

First of all, we observe that if a formula A is true, then so is the statement that, for every n, A is n-fulfillable. This is true because we can take any number we please larger than n for $(s)_0$ and then choose a number which is larger than the maximum values assumed by the "Skolem functions" corresponding to the existential quantifiers in A (as the arguments of those functions range through numbers $\leq (s)_i$), and also larger than the square of $(s)_i$, to be $(s)_{i+1}$, where $i = 0, 1, \ldots, n-1$. This choice guarantees that a suitable value for the next existential quantifier in the formula can always be found without going more than one place to the right of the place chosen by the "attacking" player when he picked a number for the preceding universal quantifier. Moreover, for any fixed formula A, this argument can be formalized in Peano Arithmetic, that is,

(IV) It is a theorem of P that if A, then A is n-fulfillable. (For $n = 1, 2, 3, \ldots$).

Obviously, since it is a theorem of P that *if P_i then P_i is n-fulfillable*, and P_i is itself an axiom of P, then (for each i, n) it is a theorem of P that *P_i is n-fulfillable*. So if P is Σ_2-sound (if all Σ_2-statements implied by P are true), then it is *not a* theorem of P that *there exists an n such that the conjunction of $P_1, P_2, \ldots$ up to P_n is not n-fulfilled*. In other words, the *negation* of (III) is not provable in P unless P fails to be Σ_2-sound (N.B. Σ_2-soundness—truth of all Σ_2-statements implied by P—is weaker than Gödel's hypothesis of "omega consistency".)

Our problem is to show by model theoretic means that (III) (which is a true statement if P is Σ_1-sound, since for each n, *the conjunction of $P_1, P_2, \ldots,$ up to P_n is n-fulfillable* is a Σ_1-consequence of the conjunction (P_1 & P_2 & $\cdots$ & P_n), and hence of P) is not a theorem of P. This means we must construct a model in which (III) is false. Obviously the standard model will not do since (III) is true in the standard model. So we must construct a nonstandard model!

We know (by the ultrapower technique) how to construct nonstandard models. So let us construct one—construct *any* nonstandard model, say M. In M the statement 'The conjunction of $P_1, P_2, \ldots, P_n$ is n-fulfillable' is true for the "standard" numbers $n = 0, 1, 2, 3, \ldots$ because for each of *these* numbers the statement is a theorem and M is a *model*. *But a statement expressible in the language of P cannot be true of the finite numbers in M and false of all the infinite ones!* (For, by a theorem of P, there is a least number of which the statement is not true, if there are any numbers at all of which it is not true. But there cannot be a least infinite integer in M, because if k is an infinite integer in M so is $k - 1$.) So there must be at least one infinite integer in M of which this statement is also true—an infinite N such that in M *the conjunction of the first N-axioms of P is N-fulfillable*. Hence there must be a "Gödel number" S (also an infinite integer, as we shall see), which is the Gödel number of a "sequence" (in the sense of the model M) of length N which fulfills P_N.

Consider the ordinary infinite sequence (this is not an object of the model M, but what model theorists call an "external" object) $(S)_0, (S)_1, (S)_2, \ldots$ (The members of this sequence are all infinite integers from the model M, since S is a "good" sequence and so even $(S)_0$ is larger than the infinite integer N.). Since S fulfills $P_1, P_2, \ldots$, it is easy to verify that this "external" sequence also fulfills these statements, that is, this external sequence fulfills each axiom of P. Now let H be the submodel of M which contains all the members of M which are smaller than a member of the external sequence. The external sequence is "good", so H is closed under "$+$" and "$\cdot$", and the external sequence we constructed is cofinal with H. But each axiom of

P is fulfilled by an ω-sequence which is cofinal with the structure H, and this means that each axiom of P is *true* in H! (The argument that a formula which is fulfilled by an ω-sequence which is cofinal with the integers is true in the integers carries over to any structure!) So M is actually a model of P.

Now, let us assume that we carried out this construction choosing as S the *smallest* Gödel number of a sequence of length N which N-fulfills P_N. S, considered as an "infinite integer", will be larger than every number in the sequence S (the Gödel number of a sequence in the Kleene system is always larger than every number in the sequence), and, by the construction of H, every number in H is smaller than some number in S. So S itself (considered as an "integer") is not in H. Is there any Gödel number of a sequence of length N which N-fulfills P_N in H? The answer is "No". For the statement *X N-fulfills P_N* is a Σ_1 statement, and Σ_1 statements "persist upward": if they are true in a substructure, they are true in the bigger structure. So if this statement were true in H, then it would also be true (of the very same X) in the original structure M. But then X would be a Gödel number smaller than S with the property of S (being the Gödel number of a sequence of length N which . . .)—contrary to the choice of S as the *smallest* integer with the property. So H contains no Gödel number which is a "witness" to the statement "There is an X which N-fulfills P_N", that is, this statement is *not true in H*. Hence (III) is not true in H. We have succeeded in producing a model in which (III) is false!

One last remark: if P is any consistent *finitely axiomatizable* extension of Peano Arithmetic, then if A is the conjunction of the axioms of P, *A implies $(n)(A$ is n-fulfillable$)$* is a theorem of Peano Arithmetic, and hence $(n)(A$ *is n-fulfillable*$)$ is a theorem of A. So, if we let T_A be the theory each of whose axioms $P_1, P_2, \ldots$ is just A, the statement $(n)(P_n$ *is n-fulfillable*$)$ is just (up to logical equivalence) $(n)(A$ *is n-fulfillable*$)$, and this is a theorem of T_A. But we just showed this is undecidable in a consistent recursively enumerable extension of Peano Arithmetic; hence *Peano Arithmetic has no consistent finitely axiomatizable extensions*.

References

[1] Paris, J., and L. Harrington, "A mathematical incompleteness in Peano Arithmetic," pp. 1133–42 in *Handbook of Mathematical Logic*, edited by J. Barwise, North-Holland Publishing Co., 1977. Zbl 0443.03001. MR 56:15351.

Acknowledgments

This paper is developed from a lecture to the Department of Computer Science at Peking University, June 1984. I have decided to publish this lecture at this time because Kripke's proof is *still* unpublished.

Department of Philosophy
Harvard University
Emerson Hall 209
Cambridge MA 02138
hputnam@fas.harvard.edu