Saul Kripke Memorial Conference[1]
8 May 2023

Kripke on Gödel Incompleteness

Daniel Isaacson
Oxford University

Contents

1 Introduction

Saul Kripke's second published paper (after his epoch making "Completeness Theorem in Modal Logic") was on Gödel incompleteness. It was written and published while he was an undergraduate, with the title " 'Flexible' predicates of formal number theory". His next major work on incompleteness was presented in a lecture

[1]I am grateful to Romina Birman, Yale Weiss, and Anandi Hattiangadi for inviting me to join in remembering and honouring Saul Kripke in this memorial conference. I knew Saul from my senior year as an undergraduate at Harvard, 1966-67, when I took his courses Phil 151 and Phil 143 in which he lectured on the topics that became *Naming and Necessity*, and later we were together at the Rockefeller University for two years and then in Oxford where Saul twice held a college visiting fellowship for a whole academic year and visited from time to time for shorter periods.

in Oxford in 1978 under the title, "A model-theoretic proof of Gödel's theorem". Kripke continued to develop insights into the phenomenon of incompleteness in the forty years after this, and in the last decade four papers of his on incompleteness have been published. (I'm sure we owe those later publications to the industry and expertise of the Kripke Center.) In my talk today I will survey these six contributions to understanding the incompleteness of formal systems.

2 " 'Flexible' predicates of formal number theory" (1962)

Kripke was 19 when he submitted " 'Flexible' predicates of formal number theory" for publication, and it was published in *Proceedings of the American Mathematical Society* in 1962. In this short paper (four pages) Kripke extended Gödel incompleteness from sentences to predicates. It displays extraordinary mastery of Kleene's *Introduction to Metamathematics*, the bible of the subject at that time (a comparable accomplishment to Kreisel's mastery of Hilbert and Bernays, *Grundlagen der Mathematik* as an undergraduate). In a note written much later, Kripke remarks that as an undergraduate at Harvard he was being discouraged from publishing his many further results on modal and intuitionistic logic (by Quine and Dreben, I take it he meant), and says that he published this paper, "almost to show that I could do something else." This paper remains relevant to current research, as in the work of Joel Hamkins.

For a given a base theory **F**, Kripke calls an m-place predicate $P_n(x_1, \ldots, x_m)$ in the language of **F** *flexible* for m-place Σ_n-formulas iff for every Σ_n-formula $Q(x_1, \ldots, x_m)$ with m-free variables, the sentence $\forall x_1 \ldots \forall x_m(P_n(x_1, \ldots, x_m) \leftrightarrow Q(x_1, \ldots, x_m))$ is consistent with **F**, i.e. $P_n(x_1, \ldots, x_m)$ can be consistently interpreted to have any Σ_n-definable extension. In this paper Kripke proved the existence of flexible predicates. Mostowski had obtained similar results at that time. In the note I mentioned earlier, Kripke reports that "When I met Andrzej Mostowski in 1962 at a conference on modal and many-valued logics held in Helsinki and told him of my proof, he said that I should have remarked that it was an 'essential improvement' over his result, because he could not get his result for arbitrary systems containing the theory R, while my version does so." Kripke also notes that, "Mostowski's argument was much longer than mine."

Joel Hamkins cites this paper of Kripke's in a number of his papers, in particular, "The modal logic of arithmetic potentialism and the universal algorithm", where he generalizes it to a uniform version: There is a computable sequence of formulas $\sigma_n(x)$ for $n \geq 2$, with σ_n having complexity Σ_n, such that for any model of arithmetic M

and any sequence of formulas ϕ_n coded in M, with ϕ_n of complexity Σ_n, there is an end-extension M^* of M with $M^* \vDash \forall x(\sigma_n(x) \leftrightarrow \phi_n(x)$ for all $n \geq 2$ (Theorem 22(2), p. 16). Hamkins obtains this result by use of Woodin's universal algorithm that "there is a Turing machine program that can in principle enumerate any desired finite sequence of numbers, if only it is run in the right universe; and furthermore, in any model of arithmetic, one can realize any desired further extension of the enumerated sequence by moving to a taller model of arithmetic end-extending the previous one" (p. 10)[2].

3 Model-theoretic proof of the incompleteness of arithmetic (1978)

Kripke spent the academic year 1977-78 in Oxford as a Visiting Fellow at All Souls College. On 2 February 1978 he gave a lecture at the Mathematical Institute on "A model-theoretic proof of Gödel incompleteness". The lecture began at 5.00 pm and finished at 7.30 pm, by which time much of the original audience had left. One of those who stayed to the end was Joseph Quinsey, a graduate student in the Mathematical Institute Logic Group who was inspired by that lecture to write his D.Phil. thesis on *Applications of Kripke's Notion of Fulfilment*, supervised by Dana Scott and Robin Gandy. Another who stayed to the end and talked with Kripke afterwards was Jeff Paris, who Kripke had asked to attend.

The context in which Kripke carried out this work was the then recent model-theoretic proof of the incompleteness of Peano Arithmetic obtained by Jeff Paris and Leo Harrington "A mathematical incompleteness in Peano Arithmetic" published in the *Handbook of Mathematical Logic* in which they showed by a model theoretic proof that a variant of finite Ramsey's Theorem cannot be proved in PA. For his own result Kripke introduced his powerful technique of fulfillability, which applies to systems weaker and stronger than PA, as well as to PA itself. Kripke in a 1982 joint publication with Simon Kochen, "Non-standard models of Peano Arithmetic" [5], referred to this work briefly (in Section VII (d)), but otherwise never published it, and so far as I'm aware there is no definitive account of this result in print. Here's a brief sketch of it[3].

Kripke's notion of fulfillability is defined in terms of the following game.

Definition 1 (game G) *G is a two-player game, played with a sentence A in the language of arithmetic and a strictly increasing sequence σ, which may be either finite*

[2] I am grateful to Joel Hamkins for information on his use of Kripke's flexible predicates

[3] I am grateful to Jeff Paris, Joseph Quinsey, and especially to Alex Wilkie for their help to me in understanding this result

or infinite. Since every formula in the language of arithmetic with one or more unbounded quantifiers is provably equivalent in PA to a formula in prenex normal form with no adjacent like quantifiers, we will take A to be in this form, and that, if need be by inserting an initial vacuous universal quantifier and/or a vacuous existential quantifier at the end of the quantifier prefix, we will assume that A is either Σ_0 (has no unbounded quantifiers) or is Π_{2k} for $k \geq 1$, i.e. $\forall x_1 \exists y_1 \ldots \forall x_k \exists y_k B(x_1, y_1 \ldots, x_k, y_k)$ where $B(x_1, y_1 \ldots, x_k, y_k)$ is Σ_0. Player I picks values for the universal quantifiers and Player II picks values for the existential quantifiers, according to the rules of the game; thus for A as described, there are $2k$ moves in a game. The rules of the game are as follows: Player I picks an element $\sigma(i)$ of σ, with the constraint that if σ is finite, $i < l(\sigma)$, and then chooses a number $m_1 < \sigma(i)$ by which to instantiate $\forall x_1$. Player II then chooses a number $n_1 < \sigma(i+1)$ by which to instantiate $\exists y_1$. If there are further quantifiers in the prefix, Player I then picks an element $\sigma(j)$ of σ such that $j < l(\sigma)$ if σ is finite and chooses a number $m_2 < \sigma(j)$ by which to instantiate $\forall x_2$. Player II then chooses a number $n_2 < \sigma(max(i,j)+1)$ by which to instantiate $\exists y_2$. The game continues in this way for each further pair of quantifiers $\forall x_i \exists y_i$, and ends when numbers have been picked for all $2k$ quantifiers in the prefix of A. At each round of the game, for $\sigma(i_1), \ldots \sigma(i_r)$ the elements of σ chosen as bounds by Player I in this and previous rounds, Player II picks a number $< \sigma(max(i_1, \ldots i_r)+1)$. Player II wins if the Σ_0-matrix of A instantiated with the chosen numbers, $B(m_1, n_1, \ldots, m_k, n_k)$ is true. Otherwise Player I wins. If A is Σ_0, i.e. has no unbounded quantifiers, Player II wins if A is true.

Definition 2 (fulfillability of a sentence by a sequence) *For A a sentence in the language of arithmetic that is either Σ_0 or Π_{2k} and σ is a strictly increasing sequence, we say that σ fulfills A if and only if*

(1) A is Σ_0 and A is true, or

(2) Player II has a winning strategy in the game G (as specified in Definition 1) played with A and σ.

Lemma 1 *Fulfillability of a given sentence A by a given finite sequence σ is expressible by a Σ_0-sentence (i.e. which contains no unbounded quantifiers).*

Proof: We give the idea of the proof by considering the case of A as the formula

(1) $\forall x_1 \exists y_1 \forall x_2 \exists y_2 B(x_1, y_1, x_2, y_2)$

Then a finite sequence σ fulfills A if and only if the Σ_0-sentence

(2) $(\forall i < l(\sigma))(\forall x_1 < \sigma(i))(\exists y_1 < \sigma(i+1))(\forall j < l(\sigma))(\forall x_2 < \sigma(j))$
$(\exists y_2 < \sigma(max(i,j)+1)B(x_1, y_1, x_2, y_2)$ is true.

Lemma 2 *Fulfillability of a sentence A by an infinite strictly increasing sequence σ*

is expressible by a Π_1-sentence in the language of arithmetic.

Proof: Fulfillability by an infinite sequence corresponds to fulfillability by a finite sequence except without a natural number as the length of the sequence. Deleting the bounds by the length of σ in the formula expressing fulfillability by a finite sequence, given in the proof of Lemma 1, yields

$$\forall i(\forall x_1 < \sigma(i))(\exists y_1 < \sigma(i{+}1))\forall j(\forall x_2 < \sigma(j))(\exists y_2 < \sigma(max(i,j){+}1))B(x_1, y_1, x_2, y_2)$$

Lemma 3 *A sentence A in the language of arithmetic is true iff some infinite sequence fulfills A.*

Proof: For clarity we give the argument in terms of $A = \forall x_1 \exists y_1 \forall x_2 \exists y_2 B(x_1, y_1, x_2, y_2)$

(i) It's easy to see that if some infinite sequence fulfills A, then A is true:

(ii) To show that if A is true then some infinite sequence fulfills A:

(1) Assume that A is true. Then there exist Skolem functions $f_1(x_1)$ and $f_2(x_1, x_2)$ such that $\forall x_1 \forall x_2 A(x_1, f_1(x_1), x_2, f_2(x_1, x_2))$ is true.

(2) The following recursive definition generates an infinite strictly increasing sequence σ:

$\sigma(1) = k$ for some number $k > 0$

$\sigma(i+1) = max\{f_1(x_1)+1, f_2(x_1, x_2)+1, x_1 + x_2, x_1 \cdot x_2 : x_1 < \sigma(i) \wedge x_2 < \sigma(i)\}$

(3) Note that by the condition that $\sigma(i+1)$ is closed under addition and multiplication of numbers $< \sigma(i)$, σ is strictly increasing (even if, for example, there is only one y_1 and one y_2 such that $\forall x_1 \exists y_1 \forall x_2 \exists y_2 B(x_1, y_1, x_2, y_2)$.

Definition 3 (a finite sequence is nice) *We call a finite sequence nice iff its initial element exceeds its length.*

Definition 4 (a sentence is nicely n-fulfillable) *A sentence A is nicely n-fulfillable by a finite sequence σ (or a finite sequence σ nicely n-fulfills A) iff some nice sequence σ of length n fulfills A.*

Lemma 4 *If A is true, then for every n, there is a nice sequence σ of length n such A is nicely n-fulfilled by σ.*

Proof: Since A is true, by Lemma 3 some infinite sequence σ fulfills A. For any n, choose a finite portion of σ by beginning with an element of σ bigger than n and then include the next $n-1$ elements. The resulting finite monotone increasing sequence of length n nicely fulfills A. The proof of Lemma 4 can be formalized in PA, so

Lemma 5 *For each sentence A in the language of arithmetic,* $PA \vdash (A \rightarrow \forall x \exists \sigma (l(\sigma) = x \wedge \sigma$ *nicely x-fulfills* $A))$.

Kripke then constructed a true Π_2-sentence which is unprovable in Peano Arithmetic. The informal meaning of the Kripke sentence is

$$\forall x \exists \sigma (l(\sigma) = x \wedge \sigma \text{ nicely x-fulfills } A_1 \wedge \ldots \wedge A_x)$$

where $A_1, \ldots, A_x$ are the first x axioms of PA in some given primitive recursive enumeration. However, this is not a well-formed sentence since the index k of A_k is not a variable, so cannot be quantified into. What this informal sentence is saying has strictly to be formulated in arithmetized syntax.

Theorem 6 (model-theoretic proof that PA is incomplete) *For K an arithmetized formulation of the Kripke sentence given above, on the assumption that PA is consistent there is a non-standard model of PA in which K is false, so $PA \nvdash K$, and on the assumption that PA is Σ_2-sound, $PA \nvdash \neg K$.*

Proof:

(i) Proof that $PA \nvdash K$.

(1) Assume $PA \vdash K$.

(2) Let $\mathfrak{M}$ be a non-standard model of PA, which exists by the assumption that PA is consistent and the compactness theorem for first-order logic. (This is the one place in the proof where we require the hypothesis that PA is consistent.) By assumption (1) $\mathfrak{M} \models K$.

(3) Hence for each $\ulcorner A_1 \wedge \ldots \wedge A_x \urcorner$ for A_i axioms of PA is true in $\mathfrak{M}$, and so by Lemma 5, $\mathfrak{M} \models \forall x \exists \sigma (l(\sigma) = x$ and σ nicely x-fulfills $A_1 \wedge \ldots \wedge A_x)$.

(4) Let z be a non-standard number in $\mathfrak{M}$. By (3) there is a nice sequence τ such that $l(\tau) = z$ which nicely z-fulfills $A_1 \wedge \ldots \wedge A_z$ in $\mathfrak{M}$.

(5) By the niceness of τ, $\tau(1) > z$, so since τ is strictly increasing, all the members in τ are non-standard. By the least number principle in $\mathfrak{M}$, we may take $\tau(z)$ to be minimal among all $\sigma(z)$ for nice sequences σ that nicely z-fulfill $A_1 \wedge \ldots \wedge A_z$ in $\mathfrak{M}$.

(6) We specify a submodel $\mathfrak{M}^*$ of $\mathfrak{M}$ in terms of τ as follows: $\mathfrak{M}^* =_{df}$ $\{a \in \mathfrak{M} :$ for some n $\in \mathfrak{N}\ a \leq \tau(n)\}$. Since by niceness of τ, $z < \tau(1)$, so $z \in \mathfrak{M}^*$. By the stipulation of $\sigma(i+1)$ in the proof of Lemma 3(ii), $\mathfrak{M}^*$ is closed under $+$ and $\cdot$. We will show that $\mathfrak{M}^*$ is a model of $\ulcorner A_1 \wedge \ldots \wedge A_z \urcorner$.

(7) Suppose $PA \vdash (A_1 \wedge \ldots \wedge A_z \equiv \forall x_1 \exists y_1 \forall x_2 \exists y_2 B(x_1, y_1, x_2, y_2)$ [This supposition doesn't make literal sense for non-standard z, and this condition must be expressed in arithmetized syntax.] From (4)) we have that τ z-fulfills $\forall x_1 \exists y_1 \forall x_2 \exists y_2 B(x_1, y_1, x_2, y_2)$

in $\mathfrak{M}$. Then by Lemma 1, $\mathfrak{M} \models (\forall i < l(\tau))(\forall x_1 < \tau(i))(\exists y_1 < \tau(i+1))(\forall j < l(\tau))(\forall x_2 < \tau(j))(\exists y_2 < \tau(max(i,j)+1)B(x_1, y_1, x_2, y_2))$.

(8) Restricting the scope of the bounded universal quantifiers $(\forall i < l(\tau))$ and $(\forall j < l(\tau))$ in this last formula to $i \in \mathfrak{N}$ and $j \in \mathfrak{N}$, we have that
$\mathfrak{M} \models (\forall i \in \mathfrak{N})(\forall x_1 < \tau(i))(\exists y_1 < \tau(i+1))(\forall j \in \mathfrak{N})(\forall x_2 < \tau(j))(\exists y_2 < \tau(max(i,j)+1)B(x_1, y_1, x_2, y_2))$

(9) Thus by (6), $\mathfrak{M}^* \models \forall x_1 \exists y_1 \forall x_2 \exists y_2 B(x_1, y_1, x_2, y_2)$, and so by (7) $\mathfrak{M}^* \models A_1 \wedge \ldots \wedge A_z$.

(10) Since $A_1 \wedge \ldots \wedge A_z$ is the conjunction of all the axioms A_i of PA such that $i \leq z$ and z is a non-standard number in $\mathfrak{M}^*$, for every $n \in \mathfrak{N}$, $n < z$. Hence the infinitely many axioms of PA are all included in this conjunction. So $\mathfrak{M}^*$ is a model of PA.

(11) We now show that $\mathfrak{M}^* \not\models K$. Suppose
$\mathfrak{M}^* \models \forall x \exists \sigma (l(\sigma) = x \wedge \sigma$ nicely x-fulfills $A_1 \wedge \ldots \wedge A_x)$. The quantifier $\forall x$ can be instantiated by z in $\mathfrak{M}^*$, so $\mathfrak{M}^* \models \exists \sigma (l(\sigma) = z \wedge \sigma$ nicely z-fulfills $A_1 \wedge \ldots \wedge A_z)$.

(12) Let ρ be a nice sequence of length z that nicely z-fulfills $A_1 \wedge \ldots \wedge A_z$ in $\mathfrak{M}^*$. The statement that ρ nicely z-fulfills $A_1 \wedge \ldots \wedge A_z$ is Σ_0. Hence since $\mathfrak{M}^*$ is a submodel of $\mathfrak{M}$, ρ nicely z-fulfills $A_1 \wedge \ldots \wedge A_z$ in $\mathfrak{M}$.

(13) Since ρ instantiates an existential quantifier in $\mathfrak{M}^*$, its elements are in $\mathfrak{M}^*$, so in particular $\rho(z) \in \mathfrak{M}^*$, which means that for some $k \in \mathfrak{N}$, $\rho(z) < \tau(k)$. Since k is a standard number and z is a non-standard number, $k < z$ and since τ in $\mathfrak{M}$ is strictly increasing, $\tau(k) < \tau(z)$, so in $\mathfrak{M}$, $\rho(z) < \tau(z)$. But as stipulated at (5), $\tau(z)$ is minimal. This contradiction has been derived from the supposition that $\mathfrak{M}^* \models K$, so $\mathfrak{M}^* \not\models K$. Since we have at (10) that $\mathfrak{M}^*$ is a model of PA, $PA \nvdash K$.

Proof that $PA \nvdash \neg K$.

(14) We have by Lemma 5 that for each natural number k and $A_1 \wedge \ldots \wedge A_k$ the conjunction of the first k axioms of PA,
$PA \vdash \forall x \exists \sigma (l(\sigma) = x \wedge \sigma$ nicely x-fulfills $A_1 \wedge \ldots \wedge A_k)$.

(15) Then for each natural number k by $\forall$-elimination
$PA \vdash \exists \sigma (l(\sigma) = k \wedge \sigma$ nicely k-fulfills $A_1 \wedge \ldots \wedge A_k)$.

(16) By the assumption that PA is Σ_2-sound, PA is Σ_1-sound, so for each natural number k, $\exists \sigma (l(\sigma) = k \wedge \sigma$ nicely k-fulfills $A_1 \wedge \ldots \wedge A_k)$ is true, and so
$\forall x \exists \sigma (l(\sigma) = x \wedge \sigma$ nicely x-fulfills $A_1 \wedge \ldots \wedge A_x)$, i.e. K, is true.

(17) Since $\neg K$ is Σ_2, by the assumption that PA is Σ_2-sound, $PA \nvdash \neg K$. ◀

4 "The road to Gödel" (2014)

This paper was published in 2014 in a volume edited by Jonathan Berg that was based on the proceedings of a conference, "Naming, Necessity, and More", held at the University of Haifa in 1999, in honour of Saul Kripke on the occasion of his being awarded an honorary doctorate. At that conference Kripke gave a lecture, "The Road to Gödel", and he lectured under this title a number of other times, including in Utrecht and in Oxford in February 2001.

Kripke sets out two things he wants to do in this paper: "first, to present the Gödel theorem as almost the inevitable result of a historic line of thought. I don't mean that it *did* happen that way; I mean that it *could* have, and perhaps *should* have [. . .]. Second, I want to show that the Gödel statement, the one Gödel proves to be undecidable in the first incompleteness theorem, makes a fairly intelligible assertion that can actually be stated" (p. 223).

Kripke argues that the Gödel incompleteness theorem is an inevitable outgrowth of the inconsistency of the naive (unrestricted) comprehension principle, and indeed a special case of it, and that this is the best way to understand Gödel incompleteness, rather than as an analogue of the Liar Paradox (which Gödel suggested in his publication of the result).

Kripke notes that "It is a matter of pure first-order logic that the unrestricted comprehension axiom schema is inconsistent. Russell already realized this in his example of the barber who shaves all and only those who do not shave themselves– . . . the interpretation of the epsilon relation is irrelevant" (p. 228).

Kripke then derives from this fact a non-constructive proof of the incompleteness of formal systems of arithmetic containing plus and times, i.e. there exists is a true unprovable sentence in the language of arithmetic, without finding such a sentence (pp. 232-235). He then goes on to show how consideration in terms of paradox rather than pure logic can yield a constructive proof of incompleteness. Rather than using Russell's paradox, or the Liar, Kripke focuses on Kurt Grelling's heterological paradox, which he quotes from Quine ("The ways of paradox"):

> The adjective 'short' is short; the adjective 'English' is English; the adjective 'adjectival' is adjectival; the adjective 'polysyllabic' is polysyllabic. Each of these adjectives is, in Grelling's terminology, autological: each is true of itself. Other adjectives are heterological; thus 'long', which is not a long adjective; 'German', which is not a German adjective; 'monosyllabic', which is not a monosyllabic one. Grelling's paradox arises from the query: Is the adjective 'heterological' an autological or a heterological one?

Kripke gives the following account of how to construct the Gödel sentence from Grelling's paradox:

> Suppose now we try to imitate the 'heterological' paradox, only replacing satisfaction (or 'true of') by 'provability of'. Replacing adjectives, or adjectival phrases, by formulae with one free variable (Gödel's 'class signs' – remember that we could even fix the free variable as x_1), a class sign $A(x_1)$ is naturally called 'provable of' a number n if $A(0^n)$ is provable, or alternatively, as we have seen, if $(\exists x_1)(x_1 = 0^n \wedge A(x_1))$ is provable. Suppose that we identify formulae with their Gödel numbers. Then a particular formula $Pr(x, y)$ with two free variables says that x is provable of y. $\neg Pr(x_1, x_1)$ is a class sign (in Gödel's sense) that says that a formula is unprovable of itself. It itself has a particular Gödel number n, and $\neg Pr(0^n, 0^n)$ is simply a way of saying:
>
> 'Unprovable of itself' is unprovable of itself.
>
> This is precisely the statement G constructed by Gödel. Thus the basic statement G can be called Gödel's form of the 'heterological' paradox, and to the present writer its content is clearer than if it is regarded in terms of the Liar paradox." (pp. 238-239)

5 "Gödel's theorem and direct self-reference" (2021)

In this short paper (5 pages) published online in *The Review of Symbolic Logic* in 2021, Kripke demonstrated that contrary to the usual view, stemming from Gödel (1931), direct self reference need not be contradictory, and it is possible to prove the first incompleteness theorem with a nonstandard Gödel numbering "allowing a statement to contain a numeral designating its own Gödel number." (p. 2). The trick is the following:

> Let the 'original' Gödel numbering be Gödel's own prime power product numbering, except that the smallest prime used is 3, so that Gödel numbers are always odd. In the 'new' numbering, all Gödel numbers coincide with the 'original', except that for each n, the formula
>
> $$\exists x_1(x_1 = 0^{(2k_n)} \wedge A_n(x_1)),$$
>
> gets the Gödel number $2k_n$, where k_n is the original Gödel number of $\exists x_1(x_1 = 0^{(n)} \wedge A_n(x_1))$. The 'new' numbering allows a formula to contain a numeral designating its Gödel number, and in that sense it is a self-referential Gödel numbering.

In this self-referential Gödel numbering, every formula $A_n(x_1)$ has an 'instance' $\exists x_1(x_1 = 0^{(2k_n)} \wedge A_n(x_1))$ asserting that its own Gödel number satisfies $A_n(x_1)$. The Gödel incompleteness theorem is the special case where $A_n(x_1)$ is unprovability in the system. (pp. 2-3)

6 "Mathematical incompleteness results in first-order Peano Arithmetic: a revisionist view of the early history" (2022)

[*History and Philosophy of Logic* 43 (2022), pp. 175-182]. When Jeff Paris and Leo Harrington published their paper, "A mathematical incompleteness in Peano Arithmetic" in the *Handbook of Mathematical Logic*, the editor of the *Handbook*, Jon Barwise, declared that "Since 1931, the year Gödel's Incompleteness Theorems were published, mathematicians have been looking for a strictly mathematical example of an incompleteness in first-order Peano arithmetic, one which is mathematically simple and interesting and does not require the numerical coding of notions from logic", and claimed that the Paris-Harrington sentence is the first such. In this paper Kripke takes exception to this claim, incontrovertibly, it seems to me, citing Gentzen's proof that transfinite induction of order type ϵ_0 is not derivable in Peano Arithmetic. This does not quite strictly meet Barwise's criterion, since the Gentzen theorem requires coding of ordinals $< \epsilon_0$. However, Goodstein's theorem, obviates that objection by giving a purely number-theoretic formulation of Gentzen's result with no coding, though of course to see the truth of the Goodstein sentence does require coding of base ω notations of ordinals less than ϵ_0. But equally the formulation of the finite Ramsey's Theorem and its variant in the language of PA requires coding of finite sets of numbers. Kripke also mentions the work of" Matiyasevich (1970), building on earlier work by Davis, Putnam, and Robinson, showing that in any consistent recursively axiomatized system (in which some weak theory such as R is interpretable), we can effectively find a Diophantine equation that has no solution, but where this fact cannot be proved in the system." (p. 177).

7 "The collapse of the Hilbert program: a variation on the Gödelian theme" (2022)

This paper was published in *The Bulletin of Symbolic Logic* in 2022. Kripke had earlier published an extended abstract of an Invited Special Talk on this topic given at the 2008 winter meeting of the ASL. The usual view is that the Hilbert programme was shown to be unrealizable by Gödel incompleteness. In this paper Kripke shows that the impossibility of Hilbert's programme was internal to the programme itself. This new insight comes from Kripke's demonstration that realization of the Hilbert programme for Π_2-sentences is expressible by a Π^0_2-sentence which is about all Π^0_2-sentences, and that this self-application is contradictory.

The key idea of Hilbert's programme is the ϵ-substitution method, by which provable formulas of the form $\exists x A(x)$ are replaced by $A(\epsilon x A(x))$, where $\epsilon x A(x)$ denotes any true instance of $A(x)$ and an arbitrary object of there is none, axiomatized by the schema $A(t) \rightarrow A(\epsilon x A(x)))$ for all terms t in the language. If the Hilbert programme were realized for a given formal system S, it would establish not only the consistency of S but also its Σ_1-soundness: by the ϵ-substitution method, if $S \vdash \exists x A(x)$, for $A(x)$ a Σ_0-formula, then $A(\epsilon x A(x))$ is true. But it would do more than this. It would establish Π_2-soundness, since if $S \vdash \forall x \exists y B(x, y)$, then for all n, $S \vdash \exists y B(\overline{n}, y)$, so by Σ_1-soundness of S, $\forall x \exists y B(x, y)$ is true. Kripke's key insight here is that the realization of Hilbert's programme in (a weak, i.e. finitary subsystem of) a system S for Π_2-sentences in the language of T can itself by expressed by a Π_2-sentence in the language of S, call it K, and that the realizability of the Hilbert programme for S is tantamount to $S \vdash K$. Kripke derives a contradiction from the supposition that $S \vdash K$, and establishes thereby that Hilbert's programme is not realizable.

The realization of the Hilbert programme for Π_2-sentences provable in a system S means that:

(1) For any Π_2-sentence $\forall x \exists y A(x, y)$, if $S \vdash \forall x \exists y A(x, y)$, then for every x there is some number n such that $S \vdash A(x, n)$. We can formalize this statement in arithmetized syntax by expressing the following predicate and relation:

(2) $B(x)$ to express "x is the Gödel number of a proof in S of a Π_2-sentence $\forall x \exists y A(x, y)$"

(3) $C(x, y)$ to express "there is a number $n < y$ such that y is the Gödel number of a proof of $A(x, n)$"

(4) $B(x)$ and $C(x, y)$ are expressible by Σ_0-formulas.

(5) Statement (1) can be expressed in arithmetized syntax of S by
$\forall x \exists y (B(x) \rightarrow C(x, y))$, which by (4) is a Π_2-sentence.

(6) Note that (5) does not literally express (1) since the x in $C(x, y)$ is not an arbitrary number but the Gödel number of a proof of a given Π_2-sentence. However, there is no loss of generality in this formalization since the point of the argument is to show that (1) is not provable, and if a literal formalization of (1) were provable, (5) would be provable by universal instantiation.

(7) Note also that the condition in (3) that $n < y$ reflects the fact that the numeral for

n occurs in the proof whose Gödel number is y (so it's an artefact of arithmetization).

(8) Suppose that a is the Gödel number of a proof in S of $\forall x \exists y(B(x) \to C(x,y))$.

(9) Application of the ϵ-substitution method and (8) yields a number b and a proof in S of $(B(\overline{a}) \to C(\overline{a}, \overline{b}))$

(10) By the least number principle, we may take b to be the least y such that $(B(\overline{a}) \to C(\overline{a}, \overline{y}))$ has a proof in S.

(11) By (5) and (8), a is the Gödel number of a proof of a Π_2-sentence, so by (2), (4), and Σ_0 completeness of any theory in which arithmetization of syntax can be carried out, $S \vdash B(\overline{a})$.

(12) Hence by (9), $S \vdash C(\overline{a}, \overline{b})$

(13) Then by (3), there is a number $n < b$ such that b is a proof of $(B(\overline{a}) \to C(\overline{a}, \overline{n}))$.

(14) This contradicts and therefore refutes supposition (10), which was on the basis of supposition (8), so there is no proof in S of $\forall x \exists y(B(x) \to C(x,y))$, which establishes that the Hilbert programme cannot be realized (pp. 423-424). This result transforms our understanding of Hilbert's programme.

Two remarks about the failure of Hilbert's programme:

(1) What failed was the Hilbert programme as originally conceived, i.e. looking for proofs of the consistency of infinitary branches of mathematics by finitary means, where finitary mathematics is a subpart of infinitary mathematics. Gentzen, working on Hilbert's programme under supervision from Bernays and Hilbert, gave a consistency proof for Peano Arithmetic using ϵ_0-transfinite induction applied to finitary mathematics, which is not provable in PA, but is constructive though not finitary (as Gödel later noted), which made the development of Hilbert's programme tenable, and it continued as proof theory, one of the four main branches of mathematical logic, to the present day (as shown by the recently published textbook on proof theory [13]).

(2) While Kripke showed, without question, that the impossibility of realizing Hilbert's programme as originally conceived by Hilbert does not depend on Gödel's incompleteness theorems—in particular the second incompleteness theorem—but is intrinsic to the programme itself, we should remark, as Kripke does (p. 424) that his proof of this fact relies on the arithmetization of syntax by which Gödel proved his incompletenesss theorems. At the same time one can say that arithmetization of syntax itself was implicit in Hilbert's programme, as when he said in his 1925 lecture "On the infinite", "a formalized proof, like a numeral, is a concrete and sur-

veyable object." ([4], p. 383), but it was only Gödel who realized the game-changing implications of this insight.

The structure of Kripke's proof is reminiscent of his model-theoretic proof of incompleteness of arithmetic, which Kripke remarks about in footnote 7 of this paper: "I myself arrived at the present result through a circuitous route. I had already found a purely model-theoretic version of the Gödel theorem and realized that it could also be carried out syntactically, using appropriate finite approximations and semantic tableaux. But then I saw that the ladder could be kicked away and that, formulated in detail, the result the Hilbertians were attempting to obtain in fact implies its own impossibility."

If I may be permitted an entirely personal remark, I am very struck by an acknowledgment by Kripke related to this last remark: "I am indebted to Burton Dreben for his insistence that the Hilbert program or approach (Ansatz) was not merely to prove the consistency of mathematics by finitary means, but was a specific program for interpreting proofs. Thus, as Dreben emphasized, it is a kind of constructive model theory." (p. 424). From time to time over many years I heard Saul express resentment at Dreben as having been unsupportive or even discouraging of his work on modal logic when he arrived at Harvard as an undergraduate. That Saul writes with such warmth and appreciation of having learned something from an idea that was quite central to Dreben's thinking is very striking to me, and—given my own affection for Dreben, who was my undergraduate tutor, and my affection for Saul—it's gratifying to me to see this rapprochement.

8 Conclusion

These six contributions by Saul Kripke to understanding incompleteness of formal systems of arithmetic and other systems, are very rich. They are also very diverse, and don't, as such, constitute a research programme. Rather each one gives us a gem of new understanding of the phenomenon of incompleteness, though that said, there is a very striking connection between the model-theoretic proof of incompleteness and the proof-theoretic internal refutation of Hilbert's programme, as Kripke notes. Both establish Π_2-incompleteness, and both make central use of the least number principle.

These contributions differ one from another in the extent to which they are mathematical or philosophical. I would classify as primarily mathematical the paper on flexible predicates, the model-theoretic proof of incompleteness, and the paper on Gödel's theorem and direct self-reference. The paper on the collapse of the Hilbert programme obtains a result that's highly significant mathematically and philosophically. The "revisionist" view of the early history of incompleteness results is mainly philosophical. The mathematical results in these contributions display tremendous ingenuity. What all these contributions have in common is the uniqueness of their viewpoint, inviting us to think differently about each of these topics.

References

[1] Kurt Gödel, "Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I", *Monatshefte für Mathematik und Physik* 38 (1931), pp. 173-198; English translation by Jean van Heijenoort with Gödel's revisions "On formally undecidable propositions of *Principia Mathematica* and related systems I", in Solomon Feferman *et al.* (eds), *Kurt Gödel Collected Works Volume I: Publications 1929-1936*, Oxford University Press, 1986, pp. 144-195.

[2] Warren Goldfarb, "Herbrand's thoerem and the incompleteness of arithmetic", *Iyyun, A Jerusalem Philosophical Quarterly* 39 (January 1990), pp. 45-64.

[3] Mattias Granberg Olsson, *A Model-Theoretic Proof of Gödel's Theorem: Kripke's Notion of Fulfilment*, MSc thesis, Självständiga Arbeten I Matematik, Department of Mathematics, Stockholm University, 2017

[4] David Hilbert, "Über das Unendliche", *Mathematische Annalen* 95 (1926), pp. 161-190; English translation by Stefan Bauer-Mengelberg, "On the infinite", in Jean van Heijenoort (ed.), *From Frege to Gödel: a source book in mathematical logic 1879-1931* Harvard University Press, 1967, pp. 369-392.

[5] Simon Kochen and Saul Kripke, "Non-standard models of Peano Arithemtic", *Logic and Algorithmic*, an international Symposium in honour of Ernst Specker, Zürich, February 1980, published in Monographie de l'Enseignmement Mathématique No 30, Université de Genève, 1982, pp. 275-295; republished with correction to "Note added in proof" in *L'Enseignement Mathématique* second series Vol 28 Nos 3-4 (July-December 1982), pp. 211-231.

[6] Saul Kripke, " 'Flexible' predicates of formal number theory", *Proceedings of the American Mathematical Society* Vol. 13 No 4 (August 1962), pp. 647-650.

[7] Saul Kripke, "A model-theoretic proof of Gödel's theorem", lecture in the Mathematical Institute Oxford University, 2 February 1978.

[8] Saul Kripke, Abstract of invited Special Talk: "The collapse of the Hilbert program: why a system cannot prove its own 1-consistency", *The Bulletin of Symbolic Logic* Vol 15 No 2 (June 2009), pp. 229-231.

[9] Saul Kripke, "The road to Gödel", Jonathan Berg (ed.), *Naming and Necessity and More*, Palgrave Macmillan, 2014, pp. 223-241.

[10] Saul Kripke, "Gödel's theorem and direct self-reference", *The Review of Symbolic Logic* published online 2 December 2021, 5 pp.

[11] Saul Kripke, "The collapse of the Hilbert program: a variation on the Gödelian theme", *The Bulletin of Symbolic Logic* Vol 28 No 3 (September 2022), pp. 413-426.

[12] Saul Kripke, "Mathematical incompleteness results in first-order Peano Arithmetic: a revisionist view of the early history", *History and Philosophy of Logic* Vol 43 No 2 (2022), pp. 175-182.

[13] Paolo Mancosu, Sergio Galvan, Richard Zach, *An Introduction to Proof Theory: Normalization, Cut-Elimination, and Consistency Proofs*, Oxford University Press, 2021.

[14] Jeff Paris and Leo Harrington, "A mathematical incompleteness in Peano Arithmetic", Jon Barwise (ed.), *Handbook of Mathematical Logic*, North-Holland Publishing Company, Amsterdam, 1977, pp. 1133-1142.

[15] Hilary Putnam, "Nonstandard models and Kripke's proof of the Gödel theorem", *Notre Dame Journal of Formal Logic* Vol 41 No 1 (2000), pp. 53-58.

[16] Joseph Quinsey, *Applications of Kripke's notion of fulfilment*, D.Phil. thesis Oxford University, April 1980; transcribed into LaTex and posted online April 2019 at `https://doi.org/10.48550/arXiv.1904.10540`