My route to arithmetization

Solomon Feferman

For Per Lindström

I had the pleasure of renewing my acquaintance with Per Lindström at the meeting of the Seventh Scandinavian Logic Symposium, held in Uppsala in August 1996. There at lunch one day, Per said he had long been curious about the development of some of the ideas in my paper [1960] on the arithmetization of metamathematics. In particular, I had used the construction of a non-standard definition π^* of the set of axioms of P (Peano Arithmetic) to show that P+$\{\neg\ \mathrm{Con}_\pi\}$ is interpretable in P, where π is a standard definition of the axioms of P and Con_π expresses the consistency of P via that presentation. Per pointed out that there is a simple "two-line" proof of this interpretability result which does not require the use of such formulas as π^*, and he wondered whether I had been aware of that. In fact, his proof had never occurred to me, and if it had at the time, it is possible I would never have been led to the use of non-natural definitions. Per regarded this as a happy accident, since subsequent work by him and others on interpretability made essential use of such definitions. In our conversation, I enlarged a bit on the background to my work on arithmetization, and when I was invited to contribute to this special issue of *Theoria* dedicated to Lindström, it seemed a natural choice to use the occasion to fill out the story. One caveat, though: the following is drawn largely from memory, not always reliable, supplemented by consultation of the 1960 paper and the 1957 dissertation from which it was drawn.

The circumstances leading to this work began in the period 1953-55 when I was enduring an enforced hiatus from my graduate studies in mathematics at U. C. Berkeley, as a draftee in the U. S. Army. Prior to my military service, I had been working on a thesis problem in model theory under the direction of Alfred Tarski, namely to establish the decidability of the theory of ordinals under addition. I had made only partial progress on that problem, by reducing it to the weak second-order theory of the ordinals with the $<$ relation, accomplished via a new notion of generalized power of a theory which was shown to preserve decidability. That work did not become part of my thesis, but the basic idea was eventually combined in a fruitful way with work of Bob Vaught on sentences preserved under cartesian products, to produce the fundamental Feferman-Vaught paper [1957] on generalized products of theories[1].

My Army posting had been in the Signal Corps at Fort Monmouth, New Jersey. This was at the height of the Cold War, though fortunately no hot wars were in progress; the U. S involvement in Korea was safely past, and its engagement in Vietnam was still some years away. At Fort Monmouth I was assigned to a math unit which primarily researched "kill" percentages by Nike missile batteries against hypothetical enemy attacks. The percentages we calculated under different hypotheses always turned out to be somewhat less than 100%, so (according to those results) in the event of a massed attack on, say, New York or Washington, at least one atom-bomb-tipped missile would be sure to get through, and that would be the end of those cities. Needless to say, the results of this research left me somewhat uneasy. To divert my mind and try to make up for the lost time in my education, in the few spare hours of leisure that I had during evenings and weekends I set out to master Kleene's *Introduction to Metamathematics*. That gave me a much better understanding of recursive function theory and Gödel's incompleteness theorems than I had previously obtained from my courses in Berkeley, where we were heavily into model theory, set theory and algebraic logic. Besides expanding my vistas in logic, Kleene's concern with foundational matters struck a responsive chord in me, and I began to think of a shift in research directions. Serendipitously, during this period I received in the mail from Alonzo Church a request which I can still visualize; it was meticulously handwritten in black and red ink on a postcard with wavy, straight and double-straight underlines. The request was to review a paper by Hao Wang for *The Journal of Symbolic Logic* on an arithmetized version

[1] Meanwhile, the decision problem for the ordinals under $+$ was solved in the affirmative by Andrzej Ehrenfeucht in 1956, via the use of back-and-forth game methods.

of Gödel's completeness theorem[2]. It was my work on this review that led me straight into questions about the arithmetization of metamathematics.

The main result of the paper in question, Wang [1951], was that if T is a recursively axiomatized first-order theory, then T has a formal model (i.e., can be interpreted) in P together with the formal statement Con_T that T is consistent. This was an extension of the corresponding result for finitely axiomatized T in Hilbert-Bernays [1939]. However, Wang's arguments were rather sketchy and his statements insufficiently precise according to Tarskian standards. I invested a great deal of effort in supplying the details of a modified proof in order to convince myself of the correctness of Wang's result. This went via a formalization of the well-known Henkin [1949] simplified proof of completeness of first-order logic[3] rather than the original Gödel proof, which Wang had followed. However, I remained troubled by the ambiguity of associating such statements Con_T with recursively axiomatized T in general, and found that this question had not been adequately addressed in the literature.

By the time I returned from the army to Berkeley in the fall of 1955, I decided to devote myself fully to the study of consistency statements and the arithmetization of metamathematics in some generality. It happened that Tarski was on sabbatical in Europe that year, and while he continued nominally to serve as my advisor, Leon Henkin agreed to give active help with my pursuit of this direction of research. Though his own major interests were in model theory and algebraic logic, he offered me a willing ear and a great deal of encouragement, and with the prod of weekly meetings I soon made significant progress.[4] The problem as I took it was this: given that there are many non-equivalent ways in which a recursive set can be defined arithmetically, in what sense can we speak of *the* formal statement Con_T of consistency of an *arbitrary* recursively axiomatized theory T? And if, as I expected, there is in general no canonical choice of such a statement, what conditions have to be imposed on the definition of T in arithmetic in order to

[2]I don't know what led Church to me, as we had not previously corresponded, nor was I known for any expertise in this area. Perhaps my name had been suggested to him by Dana Scott, who had become Church's student after leaving Berkeley.

[3]As simplified still further in Hasenjaeger [1953].

[4]Actually, Henkin's interests were not entirely disconnected from mine: besides my use in the arithmetized version of his simplified proof of the completeness theorem, he had raised a few years earlier the nice question as to the provability of sentences which express their own provability. The positive solution to that problem by Löb [1955] was definitely relevant to my concerns, as discussed below.

obtain general statements of metatheorems involving Con_T such as Gödel's second incompleteness theorem and the Hilbert-Bernays-Wang arithmetized completeness theorem? I had various bits and pieces of work to go on:

1° Gödel's own formulation in [1931] of his second incompleteness theorem referred to arbitrary extensions of a base system P by a primitive recursive class T of axioms; using the fact that every primitive recursive class is binumerable ('entscheidungsdefinit' in Gödel's terminology) in P, he constructed a statement Con_T from such a binumeration to express the consistency of T.[5] In terms of this, Gödel stated, but did not prove his second incompleteness theorem for arbitrary consistent such T, simply giving a brief outline which rested on formalizing the proof of the first part of the first incompleteness theorem.

2° While the idea of Gödel's proof was convincing, carrying out the details was another matter. This was first done in Hilbert-Bernays [1939], and there only for a very special case, namely Peano Arithmetic P (in two versions Z and Z_μ). As an intermediate step in the argument, they made use of certain *derivability conditions* on a formal definition of provability $\mathrm{Prov}_\mathrm{P}(\mathrm{x})$ in the system in question. The third derivability condition is used to show the formal expression of the statement that for the system P of arithmetic

$$\text{if } \mathrm{P} \vdash \phi \text{ then } \mathrm{P} \vdash \mathrm{Prov}_\mathrm{P}(\overline{\phi}), \tag{1}$$

where $\overline{\phi}$ is the numeral of the Gödel number of ϕ. The obvious statement that would be needed for the statement of Gödel's second incompleteness theorem more generally would be

$$\text{if } \mathrm{T} \vdash \phi \text{ then } \mathrm{P} \vdash \mathrm{Prov}_\mathrm{T}(\overline{\phi}). \tag{2}$$

3° Though it could be recognized by inspection that the formula $\mathrm{Prov}_\mathrm{P}(\mathrm{x})$ expresses provability from P in the intended way, what was missing

[5] Here and below I am freely substituting the notation from my 1960 paper for that in Gödel's and other works on arithmetization, in order to explain the problems in a uniform way. Gödel used 'κ' for the class T and wrote 'Wid(κ)' where I write 'Con_T'. Initially, Gödel took for P a form of simple type theory over the natural numbers at lowest type. He modified the base system to a version of Peano Arithmetic in a brief note in 1932 (cf. Gödel [1986], pp. 234-237).

from Gödel's formulation is the step from the *class* T to the formula $\mathrm{Prov}_{\mathrm{T}}(\mathrm{x})$ via a definition of membership in T in arithmetic. The alternatives here were either:

(i) to say that this definition is produced via the *proof* of his Theorem V that every primitive recursive relation is binumerable in the base system P, following a specific step-by-step generation of the characteristic function of T by the schemata for primitive recursive functions, or

(ii) to ignore T *per se* and simply assume that one is dealing with a formula $\mathrm{Prov}_{\mathrm{T}}(\mathrm{x})$ satisfying the Hilbert-Bernays derivability conditions, or the later, more elegant derivability conditions of Löb [1955], or

(iii) to say that one simply takes an arbitrary binumeration of T in P.

I countenanced taking the route (i) as the basis for a precise formulation of Gödel's theorem for primitive recursively axiomatized extensions of arithmetic by explaining what it means to be given the sequence of defining equations for a primitive recursive function, and then canonically following out such a sequence for the formal definition of the characteristic function of T. Moreover, one could extend this to arbitrary recursively axiomatized T or even T with a recursively enumerable set of axioms by applying Craig's theorem [1953] that any such T can be axiomatized by a primitive recursive set of sentences T′. But this line did not seem fully satisfactory to me both because of the artificiality of the Craig construction and because it did not seem to be relevant to the proper general formulation of the arithmetized completeness theorem, à la Hilbert-Bernays-Wang. Indeed, I recognized that the third derivability condition was irrelevant to the latter, for which all one would need would be a numeration of T in P. That was also one reason for not following the derivability condition approach (ii). Another reason was that given any putative definition $\mathrm{Prov}_{\mathrm{T}}(\mathrm{x})$ of provability from T, one would *still* have to check that this meets the derivability conditions, and that might take non-trivial work in each case. Finally, as to route (iii) it did not seem likely that construction of $\mathrm{Prov}_{\mathrm{T}}(\mathrm{x})$ from an arbitrary binumeration of T would be sufficient to obtain Gödel's second incompleteness theorem, though the proof of that would somehow have to be established by a counter-example.

My solution to these dilemmas was to deal with formulas $\mathrm{Prov}_{\tau}(\mathrm{x})$ which incorporate a standard representation of derivability in the first-order predicate calculus, and to leave open only how the set T of non-logical axioms might be defined by a formula $\tau(\mathrm{x})$. In this way, we could hope to say what conditions on the *way* τ defines T would be needed for each metatheorem employing formal provability predicates. In particular, I introduced the classes of PR- and RE- formulas, which correspond respectively to canonical primitive recursive definitions in arithmetic and existential quantifications of such, since every recursively enumerable set is definable by the latter kind of formula.[6] These classes of formulas are readily distinguishable by their syntactic form, especially in a language which has a symbol for each primitive recursive function directly available. In a language with just $+$ and $\cdot$, by Gödel's work with the β-function, one can replace the RE-formulas by the even syntactically simpler class of formulas in prenex form using existential and bounded universal quantifiers.[7] On the other hand, one can consider whether $\tau(\mathrm{x})$ numerates or binumerates a given set T without taking the form of τ into account. The two come together of course in the fact that the recursively enumerable sets are exactly those numerable in P, and the recursive sets are exactly the same as those binumerable in P (and also the same in much weaker systems); moreover, one can choose an RE-formula to serve as the numeration, resp. binumeration of the given set in each case.

My first main results were then as follows:

(I) If T is a consistent extension of P and $\tau(\mathrm{x})$ is an RE formula which numerates the axioms of T in P, then Con_τ is not provable in T.

(II) If T is any system and $\tau(\mathrm{x})$ is *any* formula which numerates the axioms of T in P, then T is (relatively) interpretable in $\mathrm{P}+\{\mathrm{Con}_\tau\}$.

(III) One can construct a binumeration π^* of P in P such that $\mathrm{P} \vdash \mathrm{Con}_{\pi^*}$.

(IV) More generally, for any reflexive[8] extension T of P, on can construct a binumeration τ^* of T in T such that $\mathrm{P} \vdash \mathrm{Con}_{\tau^*}$.

[6] The RE-formulas are nowadays referred to as the Σ_1 formulas in arithmetic. Moreover, the system Σ_1-IA (the subsystem of P in which the induction axiom is restricted to Σ_1 formulas) is nowadays generally taken as a base for the Gödel theorems (cf. Smorynski [1977]).

[7] By the work on Hilbert's 10th Problem, we could take these nowadays to be in the even simpler form of purely existential formulas; that work has been shown to be formalizable in the system Σ_1-IA of ftn. 7 and even weaker systems; cf. Hájek and Pudlák [1993], p. 97.

[8] A system T is called *reflexive* if it proves the consistency of every finite subsystem of T, and *essentially reflexive* if every consistent extension S of T in the same language as T is

The first of these results, (I), was my general formulation of Gödel's incompleteness theorem, while (II) was my formulation of the arithmetized completeness theorem. The result (III) was necessary to show that the route (iii) above could, indeed, not be followed.[9] For its proof I simply took π^* to name the "longest" initial segment of P which is consistent, and similarly for the proof of (IV). While these served as cautionary counter-examples, their real value turned out to be their use in obtaining some interesting new results about (relative) interpretability, to begin with:

(V) For π an RE numeration of P in P one has that $\mathrm{P}+\{\neg\mathrm{Con}_\pi\}$ is interpretable in P.

The proof involved the following steps, all in P, using the formula π^* of (III):

1) Since $\neg\mathrm{Con}_\pi$ is an RE formula, it is provable that it implies $\mathrm{Prov}_\mathrm{Q}(\overline{\neg\mathrm{Con}_\pi})$ hence also $\mathrm{Prov}_{\pi^*}(\overline{\neg\mathrm{Con}_\pi})$.

2) So then by Con_{π^*}, $\neg\mathrm{Con}_\pi \rightarrow \neg\mathrm{Prov}_{\pi^*}(\overline{\mathrm{Con}_\pi})$, i.e. $\mathrm{Prov}_{\pi^*}(\overline{\mathrm{Con}_\pi}) \rightarrow \mathrm{Con}_\pi$.

3) Also $\mathrm{Prov}_{\pi^*}(\overline{\mathrm{Con}_\pi}) \rightarrow \mathrm{Prov}_\pi(\overline{\mathrm{Con}_\pi})$; but we know that Con_π implies $\neg\mathrm{Prov}_\pi(\overline{\mathrm{Con}_\pi})$.

4) Hence $\mathrm{Prov}_{\pi^*}(\overline{\mathrm{Con}_\pi}) \rightarrow \neg\mathrm{Con}_\pi$.

---

reflexive. Mostowski [1952] had proved that P is essentially reflexive. Using Gödel's second incompleteness theorem, it follows that any consistent extension T in the same language as P is not finitely axiomatizable. However, there are well-known finitely axiomatized extensions of P in stronger languages.

[9]Here is another "happy accident", at least assuming the 1-consistency of P. In the result (I) it makes no difference whether we deal with RE numerations in P or in any subsystem of P which contains Robinson's finite axiom system Q. Now what I overlooked in my 1960 paper was the following more general result:

THEOREM. If T is an extension of P and τ is *any* numeration of T in S, where S is a finite subset of P then T does not prove Con_τ.

The proof is as follows. Let $\tau'(\mathrm{x})$ be the formula $\mathrm{Prov}_\mathrm{S}(\overline{\tau(\mathrm{num}(\mathrm{x}))})$. Then τ' is RE and numerates T in P, so T does not prove $\mathrm{Con}_{\tau'}$ by (I). But P proves the reflection principle for S, so P proves $\tau'(\mathrm{x}) \rightarrow \tau(\mathrm{x})$, hence also $\mathrm{Con}_\tau \rightarrow \mathrm{Con}_{\tau'}$. Thus also T does not prove Con_τ.

What this shows is that if one deals with definitions (in the sense of numerations or binumerations) τ of T in a finite subsystem of P such as Q, we don't need to make any syntactic restrictions on the form of τ to obtain a generalization of Gödel's second incompleteness theorem. But we still have to first know the result (I) as given above.

5) It follows from 2) and 4) that $\neg\mathrm{Prov}_{\pi^*}(\overline{\mathrm{Con}_\pi})$, i.e. that $\mathrm{Con}_{\pi'}$ holds, where $\pi'(\mathrm{x}) = \pi^*(\mathrm{x}) \vee \mathrm{x} = \overline{\neg\mathrm{Con}_\pi}$.

6) Since all this is done in P, we have $\mathrm{P} \vdash \mathrm{Con}_{\pi'}$ and that tells us that $\mathrm{P}+\{\neg\mathrm{Con}_\pi\}$ is interpretable in P by (II).

Now, by comparison, here is the "two line proof" that I garnered from Per Lindström:

a) Let $\mathrm{T}=\mathrm{P}+\{\neg\mathrm{Con}_\pi\}$ and let $\tau(\mathrm{x})$ be $\pi(\mathrm{x})\vee$ $\mathrm{x} = \overline{\neg\mathrm{Con}_\pi}$; then since $\mathrm{P} \vdash \mathrm{Con}_\pi \rightarrow \mathrm{Con}_\tau$, we have that T is interpretable in $\mathrm{P}+\{\mathrm{Con}_\pi\}$ by (II).

b) Of course T is interpretable in itself, i.e. in $\mathrm{P}+\{\neg\mathrm{Con}_\pi\}$.

So we finish with a "disjunctive" interpretation of T in P, depending formally on whether Con_π holds or not.[10]

Returning to the paper [1960], after (V) the next main application of non-standard definitions for interpretability results came in the proof of Orey's Compactness Theorem:

(VI) Suppose S and T are recursively enumerable sets of axioms and that S is a reflexive extension of P. Then T is interpretable in S if each finite subset A of T is interpretable in S.

This is proved by first noting that $\mathrm{S} \vdash \mathrm{Con}_\mathrm{A}$ for each finite subset A of T. Then, as in (IV) or (V), one defines $\tau^*(\mathrm{x})$ which binumerates T in S and which is such that $\mathrm{S} \vdash \mathrm{Con}_{\tau^*}$. The interpretability of T in S then follows from (II) with P replaced by S.

I don't remember when I first learned of Orey's theorem. It was certainly after I presented the results of my dissertation in a talk at the Summer Institute for Symbolic Logic held at Cornell University in the summer of 1957 (a watershed conference in our subject). Steven Orey attended the meeting and it is my recollection that I met him there for the first time. I also believe that it was only after the meeting that he communicated this and

[10] I don't know who first observed this; at any rate, though the idea is clear in two lines, it takes a few more lines to fill in the details. It should be noted that this way of proving the result generalizes to a stronger result than my way of proving (V), since it doesn't depend on the reflexivity of P.

some related results concerning arithmetization to me, and which he kindly allowed me to include in my paper [1960].[11] Then, also in 1960, he, Kreisel and I collaborated on a paper on 1-consistency and faithful interpretations. I recently tried to confirm the historical sequence of interactions with Orey, only to learn the sad news that he had died a few years ago.

The foregoing by no means outlines the full content of [1960]; among other things there were further results (which I shall not go into here) about relative consistency statements that proved to be useful in subsequent extensions of this work. But after 1960 I myself did not actively continue my work on applications of arithmetization to questions of interpretability (though see the next paragraph). Instead, I soon took up the subject of transfinite recursive progressions of r.e. theories, as a reconstruction and extension of Turing's work on ordinal logics. My work on arithmetization was also crucial to the precise formulation of the notions and results in the resulting paper [1962], but non-standard definitions of axiom systems and the relation of interpretability played no role there.[12] However, this did not mean the end of my interest in those aspects of the subject. I was pleased to see further applications and extensive ramifications being made in the following years by Orey, Jeroslow, Hájek, Hájková, Guaspari, Lindström and Visser, among others[13]), but I did not follow that work in detail.

There are two codas to the preceding. The first is the following announcement made in footnote 3 of my 1960 paper, about which I have often been asked:

> "Work is underway on a monograph, being written in collaboration with Richard Montague, on the method of arithmetization and some of its applications. It is intended that this monograph should contain a completely detailed presentation of the work discussed here as well as of the research of Montague in [his dis-

[11] Further evidence for this is an abstract of Orey's on relative interpretations which appeared in the *J. Symbolic Logic* 24 (1959), 281-282.

[12] For the history of work on ordinal logics and progressions of theories, see my paper [1988]. It may be thought that non-standard definitions of axiom systems were also used in an essential way for the completeness results in [1962]; rather, these come about indirectly via axiom systems attached to non-standard ordinal notations. The definitions $\tau_a(x)$ of the axiom systems $T_a (a \in O)$ in a progression of theories are there defined in a uniform RE way.

[13] Cf. the list of references below as well as Ch. III.4 of Hájek and Pudlák [1993], and the historical notes thereto.

sertation [1957]]."

That monograph never appeared, though an agreement for its publication had been made early on with the North-Holland Publishing Co. and, in the following years, Montague and I invested considerable effort in its preparation. Somewhere in my files are drafts of chapters for about half the volume, which I have not looked at in years. What happened was that our paths steadily diverged in the 1960s and our thoughts and energies became largely directed elsewhere. Nevertheless, this was a commitment which we wished devoutly to be fulfilled, and we continued to work on it sporadically, frequently prodding each other to take the next step. But after Montague's awful murder in 1970, I no longer had any heart for the project. Moreover, in the meantime, research by others had overtaken us and would have had to be taken into account in order to remain up to date. In particular, Montague's dissertation work on reflexivity of various systems of set theory, first reported on publicly at the 1957 Cornell conference and then outlined in his paper [1961], was superseded by the paper of Kreisel and Levy [1968] on applications of partial truth definitions to non-axiomatizability of various systems of arithmetic and set theory by statements of bounded complexity.

The second coda has to do with the role and influence of Georg Kreisel, first as mentor, then as the closest of colleagues for some fifteen years, before the rupture in our relationship in the early 1970s. The personal aspects of this have been deftly described by my wife, Anita Burdman Feferman, in her recollections [1996]. Here, I shall only go into the particulars relevant to the present story. I first met Kreisel during the period in early 1956 when I was well into the research for my hoped-for dissertation; Kreisel happened to be visiting Berkeley for a month or so at that time. Our initial personal contact was magical for me: I had hardly to begin explaining what I had done and what I was in the process of working on, to see that Kreisel understood immediately, and that it related to things he had thought about and to a whole body of literature in which he was completely at home. His positive reception of my ideas confirmed my views of the significance of what I was up to, and added to my determination to make this work my thesis, despite Tarski's reservations. The problem for Tarski was that its content was out of the mainstream of his interests, and this was compounded by the difficulties of give-and-take communication while he was on sabbatical in Europe. (What was not a concern for him, fortunately, was that I had previously demonstrated that I could meet his exacting standards of rigor and clarity of exposition.) In addition to the active encouragement and regular monitoring of the work given to me by Henkin, mentioned above, the boost provided by Kreisel's quick appreciation was psychologically crucial at that agonizing time. In addition, Kreisel opened up a new world to me through his interests in constructivity, predicativity and proof theory, interests which I was naturally attracted to and which would come to dominate my own subsequent work.

Even so there was a fly in the ointment of this new stage in my development as a logician. Much to my surprise, in the public discussion following my lecture on consistency statements and interpretability of theories at the 1957 Cornell Summer Institute, Kreisel turned on me and made light of my criticisms of Wang's formulation of his results which had been the starting point of my research, suggesting that there were *no* genuine ambiguities involved.[14] All I could say in response was that Wang should speak for himself; Wang demurred. But that was not the end of the matter. Like a father who keeps bringing up one's lapses that one doesn't want to hear, in the following years Kreisel kept disparaging the use of non-natural consistency statements, and stressing the need for an account of "canonical" presentations of axiomatic systems in order to obtain a proper generalization of Gödel's second incompleteness theorem. Naturally this rankled, but I had to grant some justice to his goading. Still it was not until 1980 that I gave serious consideration to the matter, not just to meet Kreisel's entreatments but because of the way it emerged from other work that I had been up to. My solution, in the paper [1982] and, in an improved form in [1989], was to treat formal systems as they are actually presented to us in practice through the finite inductive generation of various syntactic categories of objects, operations on them, and relations between them.[15] Besides addressing the conceptual issue of finding the "right" framework for the second incompleteness theorem and other results of the same character in the formalization of metamathematics, this work was conceived of as having potential pedagogical and practical value, the latter via the more currently fashionable pursuit of computer implemen-

[14]To the contrary, for example, Bill Hanf subsequently corroborated my conjecture in the 1960 paper (ftn. 23, p. 78) about the falsity of Wang's statement that the consistency of every decidable theory is provable in arithmetic. In fact, in the very interesting paper Hanf [1965], using a novel construction, he produced a finitely axiomatizable decidable system for which that fails.

[15]The reader is referred to the papers [1982,1989] for the relationship to earlier work, particularly of Post, Smullyan and Jeroslow.

tation of a wide variety of logical systems.[16] The irony of all this is that if I had addressed the question of canonical presentations of formal systems from the outset rather than concentrating on what was problematic in the literature and thence what roles non-canonical statements might play, the "happy accident" to which Per Lindström referred above might never have occurred.

Acknowledgments. I wish to thank Anita Burdman Feferman and Karl-Georg Niebergall for their helpful comments on a draft of this paper.

Departments of Mathematics and Philosophy
Stanford University

References

W. Craig [1953] On axiomatizability within a system, *J. Symbolic Logic 18*, 30-32.

A. B. Feferman [1996] Kreisel on the telephone: an appreciation, in *Kreiseliana. About and around Georg Kreisel*, P. Odifreddi, ed., (A. K. Peters, Wellesley, MA), 43-49.

S. Feferman [1957] *Formal Consistency Proofs and Interpretability of Theories*, Dissertation, Univ. of California at Berkeley.

S. Feferman [1960] Arithmetization of metamathematics in a general setting, *Fundamenta Mathematicae 49*, 35-92.

S. Feferman [1962] Transfinite recursive progressions of axiomatic theories, *J. Symbolic Logic 27*, 259-316.

S. Feferman [1982] Inductively presented systems and the formalization of metamathematics, in *Logic Colloquium '80*, D. van Dalen, et al, eds. (North-Holland, Amsterdam), 95-128.

[16]The solution in [1989] goes far beyond my [1960] in the generality of formal systems covered; to quote from the introduction op. cit., it covers "propositional and predicate calculi of various kinds (classical, intuitionistic, many-valued, modal, temporal [etc., etc.]) and styles (Hilbert, Gentzen natural deduction, Gentzen sequential, linear, etc.), as well as equational calculi, lambda calculi, combinatory calculi, [etc., etc.]."

S. Feferman [1988] Turing in the land of O(z), in *The Universal Turing Machine. A Half-century Survey*, R. Herken, ed. (Oxford Univ. Press, Oxford), 113-147.

S. Feferman [1989] Finitary inductively presented logics, in *Logic Colloquium '88*, R. Ferro, et al, eds. (North-Holland, Amsterdam), 191-220. [Reprinted in *What is a Logical System?*, D. Gabbay, ed. (Clarendon Press, Oxford, 1994), 297-328].

S. Feferman, G. Kreisel and S. Orey [1960] 1-consistency and faithful interpretations, *Archiv für Mathematische Logik und Grundlagenforschung 5*, 52-63.

S. Feferman and R. L. Vaught [1959] The first order properties of products of algebraic systems, *Fundamenta Mathematicae 47*, 57-103.

K. Gödel [1931] Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I, *Monatshefte für Mathematik und Physik 38*, 172-198. [Reprinted with translation, in Gödel [1986], 144-195.]

K. Gödel [1986] *Collected works, Vol. I. Publications 1929-1936*, S. Feferman, et al, eds. (Oxford University Press, New York).

D. Guaspari [1979] Partially conservative extensions of arithmetic, *Transactions American Math. Soc. 254*, 47-68.

P. Hájek and M. Hájková [1972] On interpretability in theories containing arithmetic, *Fundamenta Mathematicae 76*, 131-137.

P. Hájek and P. Pudlák [1993] *Metamathematics of First Order Arithmetic*, (Springer, Berlin).

M. Hájková [1971] The lattice of bi-numerations of arithmetic, *Commentationes Mathematicae Univ. Carolinae 12*, Part I: 81-104, Part II: 281-306.

W. Hanf [1965] Model-theoretic methods in the study of elementary logic, in *Symposium on the Theory of Models*, J. Addison, et al, eds. (North-Holland, Amsterdam).

G. Hasenjaeger [1953] Eine Bemerkung zu Henkin's Beweis für die Vollständigkeit des Prädikatenkalküls der ersten Stufe, *J. Symbolic Logic 18*, 42-48.

L. Henkin [1949] The completeness of the first-order functional calculus, *J. Symbolic Logic 14*, 159-166.

D. Hilbert and P. Bernays [1939], *Grundlagen der Mathematik, Vol. II*, (Springer, Berlin). [Second edition 1970]

R. G. Jeroslow [1971] Consistency statements in formal theories, *Fundamenta Mathematicae 72*, 17-40.

S. C. Kleene [1952] *Introduction to Metamathematics*, (North-Holland, Amsterdam).

G. Kreisel and A. Levy [1968] Reflection principles and their use for establishing the complexity of axiomatic systems, *Archiv für Mathematische Logik und Grundlagenforschung 5*, 17-40.

P. Lindström [1979] Some results on interpretability, in *Proceedings of the 5th Scandinavian Logic Symposium*, Jensen et al, eds. (Aalborg), 329-361.

P. Lindström [1984] On certain lattices of degrees of interpretability, *Notre Dame J. of Formal Logic 25*, 127-140.

M. H. Löb [1955] Solution of a problem of Leon Henkin, *J. Symbolic Logic 20*, 115-118.

R. Montague [1957] *Contributions to the Axiomatic Foundations of Set Theory*, Ph. D. Dissertation, Univ. of Calif., Berkeley.

R. Montague [1961] Semantical closure and non-finite axiomatizability I, in *Infinitistic Methods*, (Pergamon Press, Oxford), 45-69.

A. Mostowski [1952] On models of axiomatic systems, *Fundamenta Mathematicae 39*, 133-158.

S. Orey [1961] Relative interpretations, *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik 7*, 146-153.

C. Smorynski [1977] The incompleteness theorems, in *Handbook of Mathematical Logic*, J. Barwise, ed. (North-Holland, Amsterdam), 821-865.

H. Wang [1951] Arithmetical models for formal systems, *Methodos 3*, 217-232.

A. Visser [1990] Interpretability logic, in P. Petkov, ed., *Mathematical Logic*, (Plenum Press, New York), 175-208.