Notre Dame Journal of Formal Logic
Volume 22, Number 4, October 1981

Fifty Years of Self-Reference in Arithmetic

C. SMORYŃSKI

It is now fifty years since Hans Hahn first presented an abstract of the then unknown Kurt Gödel to the Vienna Academy of Sciences. The rest, as it is said, is history. Much of this history is well-known and I do not propose to repeat the usual platitudes. On any golden anniversary, however, it is natural to look back and I am not one to rebel against nature. On this occasion I will sing the hitherto unsung song of diagonalisation. While self-reference is one of the more outstanding features of Gödel's work and self-reference in arithmetic has had some notable success, this success is neither so widely known nor so great that my message should bore the average reader.

One of the great curiosities of my topic is how long it took (perhaps better: is taking) for the subject to develop. Even the most obvious and central fact—the Diagonalisation Theorem—seems to have had difficulty in surfacing. It is not to be found in many of the basic textbooks (e.g., Kleene [20], Mendelson [25], Shoenfield [38], Bell and Machover [1], and Manin [24]) and it is only stated in its most rudimentary form in most others (e.g., Boolos and Jeffrey [3], Enderton [4], and Monk [26]). The two most substantial expositions of Incompleteness Theory (Mostowski [28] and Stegmüller [47]) offer no explicit statement of the Diagonalisation Theorem in any form. Indeed, it is only in a recent more advanced exposition (Boolos [2]) that the full Diagonalisation Theorem has finally graced the pages of a book. Yet, diagonalisation in arithmetic is fifty years old and was stated in proper generality in print in 1962 [27]—long before most of the available textbooks were written.

Perhaps, before writing another word, I should outline the history of the Diagonalisation Theorem. While I have not made an exhaustive search of the literature, I can report that a cursory examination of the more important papers yields the following development:

Received August 25, 1980

1931: Gödel [9] produces a specific instance of self-reference: There is a sentence ϕ such that $\vdash \phi \leftrightarrow \neg Pr(\ulcorner \phi \urcorner)$.

1939: Rosser [35] makes explicit the Diagonalisation Lemma for sentences: For any ψv with only v free, there is a sentence ϕ such that $\vdash \phi \leftrightarrow \psi(\ulcorner \phi \urcorner)$. This comes a year after Kleene [19] has stated and proved the Recursion Theorem in full generality.

1960: Ehrenfeucht and Feferman [5] prove a free-variable form of the theorem: For $\psi v_0 v_1$ with only v_0, v_1 free, there is a formula ϕv with only v free such that $\vdash \phi v \leftrightarrow \psi(\ulcorner \phi \dot{v} \urcorner, v)$. ($\dot{v}$ is Feferman's dot notation, [6].) However, they only explicitly state the conclusion: For all $x \in \omega$, $\vdash \phi \bar{x} \leftrightarrow \psi(\ulcorner \phi \bar{x} \urcorner, \bar{x})$.

1962: Montague [27] states the final result—the Diagonalisation Theorem: For $\psi v_0 v$ with only v_0, v free, there is a formula ϕv with only v free (and with no occurrence of v_0) such that $\vdash \phi v \leftrightarrow \psi(\ulcorner \phi v_0 \urcorner, v)$. This is precisely the form analogous to Kleene's original formulation of the Recursion Theorem back in 1938.

Concomitant with this slow unfolding of the fundamental theorem of the subject was, of course, a slow realisation of its utility and a lack of depth in its application. The reasons for this are sociological rather than mathematical. My pet theory is that development was hindered by diffidence in the face of philosophical issues. Be that as it may, the growth of recursion theory also had a substantial negative effect. First, originating partly from Gödel's work, the rapid development of recursive function theory under Kleene drew attention away from the linguistic aspects of the Incompleteness Theorems. (Freudian adherents to the Diffidence Theory would even submit the thesis that this offered a "safe" outlet for one's interest in incompleteness and undecidability.) Second, the relatively easy use of nonrecursive r.e. sets, creative sets, and pairs of effectively inseparable r.e. sets in deriving basic results of Incompleteness Theory misled many researchers into believing the definitive work to have been done—that all interesting questions were many-one reducible to one's favourite pair of effectively inseparable r.e. sets.

The celebrant who chooses to extol the glories of five decades of arithmetical egotism must decide upon the most virtuous path to follow. In view of the popularity of the above-cited misinformation, I have made the perverse decision to concentrate on a few usable general fixed points and some of their applications. Before discussing the positive side to my choice, I will first document my fall from grace.

Unquestionably the most solid accomplishment to report on in any survey of "fifty years of self-reference in arithmetic" is the modal analysis of the definability of fixed points. Having taken part in this analysis myself, I am fully aware of its beauty and would criticise most severely any scoundrel who would contrive to survey "fifty years of self-reference in arithmetic" without discussing it. I do not discuss it here. To temper any judgment of hypocrisy, I hastily note that I have already published one exposition of this aspect of self-reference [41] and that I am slated to write another [45]. I also refer the reader to [12] and [11] for further information. Indeed, I note that via [11] and, to some extent, [44] the modal material has become directly relevant to the material I have chosen to discuss.

As I have already said, I shall concentrate on a few usable general fixed points. Together with some of their applications and a few comparative remarks, I hope to demonstrate these fixed points to be not only no less powerful than the *usual*[1] recursion theoretic tools, but perhaps even more powerful. The study of self-referential statements has reached a stage where we can recognise certain fixed points to transcend their individuality and embody unifying concepts analogous to those unifying recursion theoretic concepts embodied by particular r.e. sets. The Shepherdson fixed point, to cite the most outstanding example, codifies something like the utility of effective inseparability. The exact correspondence between fixed points and recursion theoretic concepts is a long way off. Indeed, at the present time, we do not even have weak transfer principles.[2] It follows that my discussion will be only slightly better than anecdotal. But it should at least illustrate the possibilities for the future as well as document the highlights of the past.

In each of Sections 1 and 2, below, I offer five fixed points for the reader's consideration. Those of the first section are specific sentences, the stark simplicity of which belies my promise of generality. They are introduced to provide skeletons on which later to hang the flesh of this generality. Their presentation is also pedagogical. They show a chronologically long development of three simple ideas: (i) simulation of the Liar Paradox, (ii) comparison of witnesses, and (iii) hierarchical generalisation (i.e., relativisation to a partial truth definition). This development is only slightly nonlinear and the discussion coheres well. More difficult to describe is the less coherent Section 2. For one thing, the new twists not being fully compatible, the exposition is extremely nonlinear. For another, its first fixed point belongs halfway between Sections 1 and 2, thus wreaking havoc on any universal observation I choose to make on the dichotomy between the fixed points of these two sections. Nonetheless, if we agree to ignore this troublesome counterexample, we can distinguish the fixed points of Section 2 from those of Section 1 by their greater flexibility: they have parameters.

Following the definition of each fixed point are: (i) a statement of its provability, refutability, or such (eventually in parametric terms), and (ii) one or more sample applications. Since I wish to stress the applicability of the fixed points, I give proofs only of the latter. The simplicity of these proofs should lend credence to my claim. The other proofs, though not too complex, would be distracting. The reader desirous of them is referred to [42] and [43], where additional applications are also to be found.

Among those topics not included, I should emphasise the following: (i) the modal theory, (ii) analyses and applications of fixed points not based on the so-called provability predicate, and (iii) applications of diagonalisation requiring more than the Ehrenfeucht-Feferman version. I have already discussed the first of these. With respect to (ii), I refer the reader to Manevitz and Stavi [23] for a beautiful example of such an analysis and to Švejdar [48] for a bountiful collection of examples of the latter. As for the full power of diagonalisation, I note that: (a) the weak Rosser and Ehrenfeucht-Feferman versions suffice for a great many applications, and (b) applications of Montague's full theorem are usually of the flavour of applications of the Recursion Theorem and tend to yield expected results. A very nice exception is the Uniformisation Lemma (I.4.2) of Leivant [22].

1 No-frills fixed points Consider the following five fixed points:

I. (Gödel, 1931 [9])
$\phi \leftrightarrow Pr(\ulcorner \neg\phi \urcorner)$

II. (Rosser, 1936 [34])
$\phi \leftrightarrow .Pr(\ulcorner \neg\phi \urcorner) \preccurlyeq Pr(\ulcorner \phi \urcorner)$

III. (Kreisel and Levy, 1968 [21])
$\phi \leftrightarrow Pr_\Gamma(\ulcorner \neg\phi \urcorner)$

IV. (Kent, 1973 [18])
$\phi \leftrightarrow .Pr_\Gamma(\ulcorner \neg\phi \urcorner) \preccurlyeq Pr_\Gamma(\ulcorner \phi \urcorner)$

V. (Guaspari 1976, as modified by Solovay, Hájek, and Smoryński; cf. [10], [15], and [43])
$\phi \leftrightarrow .Pr_{\Gamma_1}(\ulcorner \neg\phi \urcorner) \preccurlyeq Pr_{\Gamma_2}(\ulcorner \phi \urcorner)$.

Even without an explanation of the notation, the reader can see here a simple evolution of ideas. To achieve this effect, I have had to cheat a bit. Even the novice should recognise that Gödel and Rosser actually used the dual Π_1 forms of their fixed points: Letting, e.g., $\psi = \neg\phi$ in fixed point *I*, one obtains the familiar $\psi \leftrightarrow \neg Pr(\ulcorner \psi \urcorner)$. The *cognoscente* should note the missing fixed point of Scott [36] and how its inclusion would ruin the forward progression of my cute little development. However, not even the expert should immediately see the relation between Guaspari's original fixed points and their modern descendant *V*. The history is revisionist; but the conceptual ordering is natural.

Notation: $Pr(\cdot)$ denotes a standard Σ_1 provability predicate for a given consistent r.e. arithmetic theory T, say $T \supseteq PA$. Occasionally, to avoid ambiguity I write $Pr_T(\cdot)$. Con_T will denote $\neg Pr_T(\ulcorner 0 = 1 \urcorner)$. There are two witness comparison formulas,

$$\exists v\psi v \preccurlyeq \exists v\chi v\colon \exists v_0[\psi v_0 \wedge \forall v_1 < v_0 \neg\chi v_1]$$
$$\exists v\psi v \prec \exists v\chi v\colon \exists v_0[\psi v_0 \wedge \forall v_1 \leqslant v_0 \neg\chi v_1].$$

Γ denotes a class of formulas (Σ_n, Π_n, etc.) which possesses a truth definition in arithmetic: There is some $Tr_\Gamma(\cdot) \in \Gamma$ such that, for all formulas $\phi v_0 \ldots v_{n-1} \in \Gamma$,

$$PA \vdash \phi v_0 \ldots v_{n-1} \leftrightarrow Tr_\Gamma(\ulcorner \phi \dot{v}_0 \ldots \dot{v}_{n-1} \urcorner).$$

$Pr_\Gamma(\cdot)$ is just the relativisation of $Pr(\cdot)$ to $Tr_\Gamma(\cdot)$:

$$Pr_\Gamma(v) \leftrightarrow \exists v_0 [Tr_\Gamma(v_0) \wedge Pr(v_0 \dot{\rightarrow} v)],$$

where $\dot{\rightarrow}$ represents implication. In comparing witnesses involving $Pr_\Gamma(\cdot)$, however, one rewrites the formula thus,

$$Pr_\Gamma(v)\colon \exists v_1 \exists v_0 < v_1 [Tr_\Gamma(v_0) \wedge Prov(v_1, v_0 \dot{\rightarrow} v)],$$

where $Pr(v)$ is $\exists v_1 Prov(v_1, v)$.

Bearing this in mind, let us take a close look at each of our five fixed points.

Fixed Point I. If $T \vdash \phi \leftrightarrow Pr(\ulcorner \neg\phi \urcorner)$, then

i.a. $T \nvdash \neg\phi$
b. if T is sufficiently sound, $T \nvdash \phi$
ii. $T + Con_T \vdash \neg Pr(\ulcorner \neg\phi \urcorner)$
iii. $T \vdash \neg\phi \leftrightarrow Con_T$.

The additional assumption required in i.b is known as Σ_1-soundness and will be encountered again later. Its use here is usually regarded as a weakness of this fixed point. Conclusions ii and iii, however, vindicate this choice. Conclusion ii, which is merely the formalisation of i.a, yields immediately the Second Incompleteness Theorem: $T \nvdash Con_T$. Conclusion iii is also of interest. First, it justifies the familiar reference to *the* sentence $\neg\phi$ asserting its own unprovability. Second, it is the prototype for the prettiest result of the modal theory—the explicit definability of all modal fixed points (cf. my surveys cited earlier).

To bypass the additional assumption of i.b, Rosser introduced the witness comparison:

Fixed Point II. If $T \vdash \phi \leftrightarrow .Pr(\ulcorner \neg\phi \urcorner) \preccurlyeq Pr(\ulcorner \phi \urcorner)$, then

i. $T \nvdash \phi, \neg\phi$
ii. $T + Con_T \vdash \neg Pr(\ulcorner \phi \urcorner), \neg Pr(\ulcorner \neg\phi \urcorner)$
iii.a. $T + Con_T \vdash \neg\phi$
b. $T + \neg\phi \nvdash Con_T$.

While we have gained independence without any additional assumption, we have lost the aesthetically pleasing explicit definition. The traditional analysis of $Pr(\cdot)$ is, in fact, demonstrably insufficient to settle such a simple question as the uniqueness or nonuniqueness (up to provable equivalence) of fixed points of the form II. I refer the reader to Guaspari and Solovay [12] for further information on this matter. The simple failure, iii.b., of $\neg Con_T$ to serve as an explicit definition of ϕ follows from ii ($T + Con_T \vdash \neg Pr(\ulcorner \phi \urcorner)$) and the Second Incompleteness Theorem for $T + \neg\phi$ ($Con_{T+\neg\phi} \leftrightarrow \neg Pr(\ulcorner \phi \urcorner)$).

While hierarchical generalisations are often pedestrian in themselves, their applications can be novel. Fixed points III and V exemplify this.

Fixed Point III. If $T \vdash \phi \leftrightarrow Pr_\Gamma(\ulcorner \neg\phi \urcorner)$, then

i. For all sentences $\gamma \in \Gamma$ such that $T + \gamma$ is consistent, $T + \gamma \nvdash \neg\phi$.
ii. $T + RFN(T) \vdash \neg\phi$.

RFN(T) is a sort of generalisation of consistency known as the (proof theoretic) *Reflection schema*: For ϕ containing only v free,

$$\forall v[Pr(\ulcorner \phi\dot{v} \urcorner) \rightarrow \phi v].$$

The point to i and ii is given by the following theorem of Kreisel and Levy [21]:

Theorem *No consistent set of axioms of bounded quantifier complexity when added to T will yield all consequences of $T + RFN(T)$.*

Proof: Let Δ be a set of sentences of bounded complexity such that $T + \Delta$ is consistent. Choose n such that $\Delta \subseteq \Sigma_n$ and let $\Gamma = \Sigma_n$ and ϕ be as in III. By i and ii,

$$T + \Delta \nvdash \neg\phi,\ T + RFN(T) \vdash \neg\phi;$$

whence Δ does not yield all consequences of $RFN(T)$ over T.

Sample Corollary *PA and ZF are not finitely axiomatisable.*

The corollary follows from the observation that, in their respective languages, *PA* and *ZF* prove the Reflection schema for the predicate calculus.[3] It follows that neither *PA* nor *ZF* have axiomatisations of bounded quantifier complexity over the predicate calculus–in particular, they are not finitely axiomatised.

This is not the quickest proof of the nonfinite axiomatisability of *PA* but it is certainly the most revealing one. Further examples are given in Kreisel and Levy [21].

Fixed Point IV. If $T \vdash \phi \leftrightarrow .Pr_\Gamma(\ulcorner\neg\phi\urcorner) \preccurlyeq Pr_\Gamma(\ulcorner\phi\urcorner)$, then

i. For all sentences $\gamma \in \Gamma$ such that $T + \gamma$ is consistent,
   $T + \gamma \nvdash \phi, \neg\phi$
ii. $T + RFN(T) \vdash \neg\phi$.

As in the step from I to II, that from III to IV takes us from mere nonrefutability to independence (cf. Kent [18] for an application[4]).

Fixed Point V. If $T \vdash \phi \leftrightarrow .Pr_{\Gamma_1}(\ulcorner\neg\phi\urcorner) \preccurlyeq Pr_{\Gamma_2}(\ulcorner\phi\urcorner)$, then

i.a. For all sentences $\gamma_1 \in \Gamma_1$ such that $T + \gamma_1$ is consistent,
   $T + \gamma_1 \nvdash \neg\phi$
b. For all sentences $\gamma_2 \in \Gamma_2$ such that $T + \gamma_2$ is consistent,
   $T + \gamma_2 \nvdash \phi$
ii. $T + RFN(T) \vdash \neg\phi$.

The step from IV to V is not as silly as it seems. By using different Γ_1 and Γ_2, one can control the complexity of ϕ. For example, choosing $\Gamma_1 = \Sigma_1$, $\Gamma_2 = \Pi_1$, one gets $\phi \in \Sigma_1$. The importance of this is brought out by the following definition and result of Guaspari [10].

Definition A sentence ψ is Γ-*conservative over* T (or, Γ-*con over* T) if, for all sentences $\gamma \in \Gamma$,

$$T + \psi \vdash \gamma \Rightarrow T \vdash \gamma.$$

Theorem *There is a* $\phi \in \Sigma_1$ *such that*

i. ϕ *is* Π_1-*con over* T
ii. $\neg\phi$ *is* Σ_1-*con over* T.

Proof: These are just the contrapositions of i.a and i.b above, respectively, for $\Gamma_1 = \Sigma_1, \Gamma_2 = \Pi_1$.

While the Recursion Theorem can be invoked to yield the existence of such sentences as ϕ (at least for PA and ZF, if not yet for GB) there is no known construction involving only effective inseparability. For the sets

$A = \{\phi: \phi$ is not Π_1-con over $PA\}$
$B = \{\phi: \neg\phi$ is not Σ_1-con over $PA\}$

can be complete Σ_2 (cf. Solovay [46] and Quinsey [33], respectively). Thus, the fact that their reductions, $A \preccurlyeq B$ and $B \prec A$, separate the refutable from the provable sentences leads to no contradiction.

The use of hierarchical generalisation, even though we are constructing a Σ_1 formula, is necessary here. Fixed point II demonstrably fails to yield the theorem and, moreover, there is a convincing sense (cf. Guaspari [11]) in which our most advanced nonhierarchically generalised fixed points (B and D of the next section) cannot be used to construct sentences satisfying the theorem.

We will meet Π_1-con sentences again in the next section. For more on them and their more general Γ-con associates cf. [10], [15], [43], and [46].

*2 **Deluxe fixed points*** We now move on to less subtle, more sublime variations on the self-referential theme. Again we will consider five fixed points:

A. (Mostowski, 1961 [29])
$\phi \leftrightarrow .\bigvee\!\!\bigvee_i Pr_{T_i}(\ulcorner\neg\phi\urcorner) \preccurlyeq \bigvee\!\!\bigvee_i Pr_{T_i}(\ulcorner\phi\urcorner)$

B. (Shepherdson, 1960 [37])
$\phi \leftrightarrow .(Pr(\ulcorner\neg\phi\urcorner) \vee \psi) \preccurlyeq (Pr(\ulcorner\phi\urcorner) \vee \chi)$

C. (Smoryński, 1976 [42])
$\phi \leftrightarrow .(\bigvee\!\!\bigvee_i Pr_{T_i}(\ulcorner\neg\phi\urcorner) \vee \psi) \preccurlyeq (\bigvee\!\!\bigvee_i Pr_{T_i}(\ulcorner\phi\urcorner) \vee \chi)$

D. (Smoryński, 1976 [42])
$\phi \leftrightarrow :\theta_0 \vee .\theta_1 \wedge [(Pr(\ulcorner\neg\phi\urcorner) \vee \psi) \preccurlyeq (Pr(\ulcorner\phi\urcorner) \vee \chi)]$

E. (Guaspari, 1976, as modified by Solovay, Hájek, and Smoryński, cf. [10] and [43])
$\phi \leftrightarrow .(Pr_{\Gamma_1}(\ulcorner\neg\phi\urcorner) \vee \psi) \preccurlyeq (Pr_{\Gamma_2}(\ulcorner\phi\urcorner) \vee \chi)$.

This time, as should be evident even without an explanation of the notation, there is no directed evolution of ideas. Three distinct and apparently not fully compatible modifications of fixed points II and V of the last section are made.

Fixed Point A. Let $T_0, T_1, \ldots$ be a uniformly r.e. sequence of consistent extensions of PA and let

$$PA \vdash \phi \leftrightarrow .\bigvee\!\!\bigvee_i Pr_{T_i}(\ulcorner\neg\phi\urcorner) \preccurlyeq \bigvee\!\!\bigvee_i Pr_{T_i}(\ulcorner\phi\urcorner).$$

Then for each n, $T_n \nvdash \phi, \neg\phi$.

By $\bigvee\!\!\bigvee_i Pr_{T_i}(\cdot)$, I mean simply any Σ_1 formula defining the union of the sets of theorems of the theories T_n.

While the success of this diagonalisation is fairly unremarkable, it should be noted that this success doesn't go very far. Fixed points I and III of the previous section are also easily uniformised; but fixed points IV and V are not. Briefly, the customary infinite descent occasioned by the provability or refutability of ϕ breaks down in the latter cases: for a generalisation of IV,

$$PA \vdash \phi \leftrightarrow . \bigvee_i Pr_{T_i+\Gamma}(\ulcorner \neg\phi \urcorner) \preccurlyeq \bigvee_i Pr_{T_i+\Gamma}(\ulcorner \phi \urcorner),$$

one would like to go from a proof k_0 that (say) $T_{n_0} + \gamma_0 \vdash \neg\phi$, with $T_{n_0} + \gamma_0$ consistent, to a smaller proof $k_1 < k_0$ that $T_{n_1} + \gamma_1 \vdash \phi$, with $T_{n_1} + \gamma_1$ consistent. Unfortunately, one only gets $T_{n_0} + \gamma_1$ consistent and the descent stops.

The single most widely used fixed point has thus far been the following rather curious one:

Fixed Point B. Let $\psi, \chi \in \Sigma_1$ and let

$$T \vdash \phi \leftrightarrow .(Pr(\ulcorner \neg\phi \urcorner) \vee \psi) \preccurlyeq (Pr(\ulcorner \phi \urcorner) \vee \chi).$$

Then:

i.a. $T \vdash \phi$ iff $\mathbf{N} \models \psi \preccurlyeq \chi$
   iff $\mathbf{N} \models \phi$

b. $T \vdash \neg\phi$ iff $\mathbf{N} \models \chi \prec \psi$

ii.a. $T + Con_T \vdash Pr(\ulcorner \phi \urcorner) \leftrightarrow .\psi \preccurlyeq \chi$

b. $T + Con_T \vdash Pr(\ulcorner \neg\phi \urcorner) \leftrightarrow .\chi \prec \psi$

iii. $T + Con_T \vdash \phi \leftrightarrow .\psi \preccurlyeq \chi$.

In witness comparisons, disjunctions $\exists v_0 \theta v_0 \vee \exists v_1 \rho v_1$ are assumed rewritten with one quantifier: $\exists v(\theta v \vee \rho v)$. In applications, the disjuncts ψ, χ can occur vacuously. In such cases, the reader can assume ψ, χ to be suppressed instances of $\exists v(\neg v = v)$ or any similar false refutable Σ_1 sentence.

A remarkable feature of this fixed point is that it is not always independent; that, in fact, one has incredible control over its provability and refutability. The original and most straightforward application of it merely tallies the instances of provability and refutability.

Definition Let ϕv have only v free and let $X \subseteq \omega$. We say that ϕv *defines* X if, for all $x \in \omega$, $x \in X$ iff $\mathbf{N} \models \phi \bar{x}$. We say that ϕv *semirepresents* X in T if, for all $x \in \omega$, $x \in X$ iff $T \vdash \phi \bar{x}$.

The first application, due to Shepherdson [37], is the following:

Theorem *Let X, Y be disjoint r.e. sets. There is a formula $\phi v \in \Sigma_1$ such that*

i. *ϕv defines X*
ii. *ϕv semirepresents X in T*
iii. *$\neg\phi v$ semirepresents Y in T.*

Proof: Let ψv, $\chi v \in \Sigma_1$ define X, Y, respectively, and choose ϕv such that, for all $x \in \omega$,

$$T \vdash \phi \bar{x} \leftrightarrow .(Pr(\ulcorner \phi \bar{x} \urcorner) \vee \psi \bar{x}) \preccurlyeq (Pr(\ulcorner \phi \bar{x} \urcorner) \vee \chi \bar{x}).$$

Since X, Y are disjoint, for any $x \in \omega$,

$$\mathbf{N} \models (\psi \preccurlyeq \chi)\bar{x} \text{ iff } \mathbf{N} \models \psi\bar{x} \text{ iff } x \in X$$
$$\mathbf{N} \models (\chi \prec \psi)\bar{x} \text{ iff } \mathbf{N} \models \chi\bar{x} \text{ iff } x \in Y.$$

Given this, the theorem follows immediately from the above criteria on the truth, provability, and refutability of ϕ.

This theorem provides another excellent counterexample to the contention that the subject is a footnote to the existence of a pair of effectively inseparable r.e. sets. Using a creative set Ehrenfeucht and Feferman [5] constructed a formula $\phi \in \Sigma_1$ satisfying ii. A pair of effectively inseparable r.e. sets achieved both ii and iii for Putnam and Smullyan [32]. Again using a pair of such sets, Hájková and Hájek [16] were able further to satisfy the correctness condition i, but only for a more complicated ϕ. Eventually, I obtained the full result recursion theoretically, but using a configuration of two pairs of effectively inseparable r.e. sets and a generalised completeness property of them (cf. [40]).

Metamathematical applications of Shepherdson's fixed point abound. The following, due independently to Friedman [7], Jensen and Ehrenfeucht [17], and Guaspari [10], is particularly popular:

Theorem *The following are equivalent*:

i. (Σ_1-*Disjunction Property*). *For all sentences* $\sigma_0, \sigma_1 \in \Sigma_1$,
$T \vdash \sigma_0 \vee \sigma_1 \Rightarrow T \vdash \sigma_0$ *or* $T \vdash \sigma_1$

ii. (Σ_1-*soundness*). *For all sentences* $\sigma \in \Sigma_1$,
$T \vdash \sigma \Rightarrow \mathbf{N} \models \sigma$.

Proof: ii $\Rightarrow$ i is an immediate corollary of the provability of all true Σ_1 sentences.

i $\Rightarrow$ ii. One argues by contraposition. First, choose $\psi \in \Sigma_1$ to be a false theorem of T and choose $\phi \in \Sigma_1$ such that

$$T \vdash \phi \leftrightarrow .(Pr(\ulcorner \neg\phi \urcorner) \vee \psi) \preccurlyeq Pr(\ulcorner \phi \urcorner).$$

The refutation of i will be given by the further choices

$$\sigma_0 = \phi = (Pr(\ulcorner \neg\phi \urcorner) \vee \psi) \preccurlyeq Pr(\ulcorner \phi \urcorner)$$
$$\sigma_1 = Pr(\ulcorner \phi \urcorner) \prec (Pr(\ulcorner \neg\phi \urcorner) \vee \psi).$$

$T \vdash \sigma_0 \vee \sigma_1$: Note that for any θ, χ,

$$T \vdash \theta \rightarrow (\theta \preccurlyeq \chi .\vee. \chi \prec \theta).$$

In particular, $T \vdash \psi \rightarrow \sigma_0 \vee \sigma_1$. But $T \vdash \psi$, whence $T \vdash \sigma_0 \vee \sigma_1$.

$T \nvdash \sigma_0, \sigma_1$: By the parametric determination of the provability and refutability of ϕ, ϕ is independent of T. Since $\sigma_0 = \phi$, $T \nvdash \sigma_0$. But $T \vdash \sigma_1 \leftrightarrow \neg\phi$, since we have just seen $T \vdash \sigma_0 \vee \sigma_1$ and, as is even more easily seen, $T \vdash \neg(\sigma_0 \wedge \sigma_1)$. Thus $T \nvdash \sigma_1$.

These two applications have depended only on the actual parametric determination of the provability and refutability of ϕ. The next two applications depend on the formalisations ii and iii, above, of this determination. The first is actually a pair of related applications due to Harrington and Friedman (cf. [8]).

Theorem *Let $\theta \in \Pi_1$. There are $\sigma \in \Sigma_1$, $\pi \in \Pi_1$ such that*

i. $T + Con_T \vdash \theta \leftrightarrow Con_{T+\sigma}$
ii. $T + Con_T \vdash \theta \leftrightarrow Con_{T+\pi}$.

Proof: i. Choose $\sigma = \phi$ such that

$$T \vdash \phi \leftrightarrow .Pr(\ulcorner \neg\phi \urcorner) \preccurlyeq (Pr(\ulcorner \phi \urcorner) \vee \chi),$$

where $\chi = \neg\theta$. Then

$$\begin{aligned} T + Con_T &\vdash Pr(\ulcorner \neg\phi \urcorner) \leftrightarrow \chi \\ &\vdash \neg Pr(\ulcorner \neg\phi \urcorner) \leftrightarrow \theta \\ &\vdash \theta \leftrightarrow Con_{T+\phi}. \end{aligned}$$

ii. Choose $\pi = \neg\phi$ where ϕ is such that

$$T \vdash \phi \leftrightarrow .(Pr(\ulcorner \neg\phi \urcorner) \vee \psi) \preccurlyeq Pr(\ulcorner \phi \urcorner)$$

and $\psi = \neg\theta$.

My favourite application of this fixed point is the following from Smoryński [42]:

Theorem *The following are equivalent*:

i. *Con_T is Σ_1-con over T*
ii. (Σ_1*-soundness*). *For all sentences $\sigma \in \Sigma_1$, $T \vdash \sigma \Rightarrow \mathbf{N} \models \sigma$.*

Proof: ii $\Rightarrow$ i is an immediate consequence of the provability of all true Σ_1 sentences.

i $\Rightarrow$ ii. One argues by contraposition. Let $\psi \in \Sigma_1$ be a false theorem of T and choose $\phi \in \Sigma_1$ such that

$$T \vdash \phi \leftrightarrow .(Pr(\ulcorner \neg\phi \urcorner) \vee \psi) \preccurlyeq Pr(\ulcorner \phi \urcorner).$$

By the basic calculation of ϕ,

$$T \nvdash \phi. \tag{1}$$

By the formalisation,

$$T + Con_T \vdash \phi \leftrightarrow \psi,$$

and, by assumption, $T \vdash \psi$, whence

$$T + Con_T \vdash \phi. \tag{2}$$

(1) and (2) violate i.

While Mostowski's uniformisation of Rosser's basic fixed point II and Shepherdson's adjunction of side-formulas thereto occurred nearly simultaneously, it took over a decade for these refinements to be combined:

Fixed Point C. Let $T_0, T_1, \ldots$ be a uniformly r.e. sequence of consistent extensions of PA, let $\psi, \chi \in \Sigma_1$ be sentences, and let

$$PA \vdash \phi \leftrightarrow .(\bigvee_i Pr_{T_i}(\ulcorner \neg\phi \urcorner) \vee \psi) \preccurlyeq (\bigvee_i Pr(\ulcorner \phi \urcorner) \vee \chi).$$

Then:

i. $\exists n(T_n \vdash \phi)$ iff $\forall n(T_n \vdash \phi)$ iff $\mathbf{N} \models \psi \preccurlyeq \chi$
ii. $\exists n(T_n \vdash \neg\phi)$ iff $\forall n(T_n \vdash \neg\phi)$ iff $\mathbf{N} \models \chi \prec \psi$.

With fixed point C, applications of fixed point B can be uniformised. The most interesting example of such a uniformisation is the following:

Theorem *Let $T_0, T_1, \ldots$ be a uniformly r.e sequence of consistent extensions of PA and let X, Y be disjoint r.e. sets. There is a $\phi v \in \Sigma_1$ with only v free such that*

i. *ϕv defines X*
ii. *ϕv semirepresents X in each T_n*
iii. *$\neg\phi v$ semirepresents Y in each T_n.*

The proof is straightforward and I omit it. As one might guess from the presence of condition i, this theorem has not yet been given a proof via the usual pair of effectively inseparable r.e. sets, although two nested pairs will yield the result (cf. Smoryński [40]). Incidentally, conclusion i is no longer remarkable—we can always force it to hold trivially by adjoining a new sufficiently sound theory T_{-1} (e.g., $T_{-1} = PA$) to the enumeration.

Where fixed points A and C were introduced for the sake of uniformity, the following fixed point was introduced for the sake of nonuniformity.

Fixed Point D. Let $\psi, \chi \in \Sigma_1$, let θ_0, θ_1 be arbitrary, and let

$$T \vdash \phi \leftrightarrow .\theta_0 \vee \theta_1 \wedge [(Pr(\ulcorner \neg\phi \urcorner) \vee \psi) \preccurlyeq (Pr(\ulcorner \phi \urcorner) \vee \chi)].$$

Then:

i. $T \vdash \phi$ iff $T \vdash \theta_0$ or $[T \vdash \theta_0 \vee \theta_1$ and $\mathbf{N} \models \psi \preccurlyeq \chi]$
ii. $T \vdash \neg\phi$ iff $T \vdash \neg\theta_0$ and $[T \vdash \neg\theta_0 \wedge \neg\theta_1$ or $\mathbf{N} \models \chi \prec \psi]$.

A sample application from Smoryński [42] is the following:

Theorem *Let $T_0 \subsetneq T_1$ be consistent r.e. extensions of PA and let $X_0 \subseteq X_1$ be r.e. sets. There is a formula ϕv with only v free such that*

i. *ϕv semirepresents X_0 in T_0*
ii. *ϕv semirepresents X_1 in T_1.*

Proof: We will apply the basic calculation to both theories in a single fixed point. To this end, let $\psi_i v \in \Sigma_1$ define X_i (for $i = 0, 1$) and let $\Phi_i(\ulcorner \phi \dot{v} \urcorner)$ be the Shepherdson component:

$$\Phi_i(\ulcorner \phi \dot{v} \urcorner)\colon Pr_{T_i}(\ulcorner \neg\phi \dot{v} \urcorner) \vee \psi_i v. \preccurlyeq .Pr_{T_i}(\ulcorner \phi \dot{v} \urcorner).$$

Choose ϕv such that, for all $x \in \omega$,

$$PA \vdash \phi \bar{x} \leftrightarrow [\theta_0 \bar{x} \wedge \Phi_0(\ulcorner \phi \bar{x} \urcorner). \vee .\theta_1 \wedge \Phi_1(\ulcorner \phi \bar{x} \urcorner)],$$

where θ_1 is a sentence provable in T_1 but not in T_0 and $\theta_0 v$ is a formula to be specified later.

For T_0 we have

$$\begin{aligned} T_0 \vdash \phi\bar{x} \text{ iff } & T_0 \vdash \theta_1 \wedge \Phi_1(\ulcorner\phi\bar{x}\urcorner) \\ & \text{or: } T_0 \vdash \theta_0\bar{x} \vee \theta_1 \wedge \Phi_1(\ulcorner\phi\bar{x}\urcorner) \text{ and } \mathbf{N} \models \psi_0\bar{x} \\ \text{iff } & T_0 \vdash \theta_0\bar{x} \vee \theta_1 \wedge \Phi_1(\ulcorner\phi\bar{x}\urcorner) \text{ and } x \in X_0. \end{aligned}$$

Thus we get

$$T_0 \vdash \phi\bar{x} \text{ iff } x \in X_0$$

provided only that

$$\theta_0 v \text{ semirepresents some superset of } X_0 \text{ in } T_0. \tag{1}$$

For T_1 we have

$$\begin{aligned} T_1 \vdash \phi\bar{x} \text{ iff } & T_1 \vdash \theta_0\bar{x} \wedge \Phi_0(\ulcorner\phi\bar{x}\urcorner) \\ & \text{or: } T_1 \vdash \theta_0\bar{x} \wedge \Phi_0(\ulcorner\phi\bar{x}\urcorner) \vee \theta_1 \text{ and } \mathbf{N} \models \psi_1\bar{x} \\ \text{iff } & T_1 \vdash \theta_0\bar{x} \wedge \Phi_0(\ulcorner\phi\bar{x}\urcorner) \text{ or } x \in X_1. \end{aligned}$$

Thus we get

$$T_1 \vdash \phi\bar{x} \text{ iff } x \in X_1$$

provided only that

$$\theta_0 v \text{ semirepresents a subset of } X_1 \text{ in } T_1. \tag{2}$$

To satisfy conditions (1) and (2), let $\theta_0 v$ uniformly semirepresent X_0 (say) in T_0 and T_1.

The reduction of the nonuniform case–specifically, the use of the uniform semirepresentation of an interpolant between X_0 and X_1 in T_0 and T_1–is the key recursion theoretic difficulty. Assuming the existence of a *recursive* interpolant, the result can be proven via a pair of effectively inseparable r.e. sets (cf., di Paola [31]). Without this condition, one can use two nested pairs of such sets (cf., Smoryński [40]).

Finally, we come to:

Fixed Point E. Let $\psi, \chi \in \Sigma_1$ and let

$$T \vdash \phi \leftrightarrow .(Pr_{\Gamma_1}(\ulcorner\neg\phi\urcorner) \vee \psi) \preccurlyeq (Pr_{\Gamma_2}(\ulcorner\phi\urcorner) \vee \chi).$$

Then:

i.a. $T \vdash \phi$ iff $\mathbf{N} \models \psi \preccurlyeq \chi$

b. $T \vdash \neg\phi$ iff $\mathbf{N} \models \chi \prec \psi$

ii. ϕ is $\check{\Gamma}_1$-con and $\neg\phi$ is $\check{\Gamma}_2$-con over T iff $\mathbf{N} \models \neg\psi \wedge \neg\chi$.

A few words of explanation are probably called for. Γ_1 and Γ_2 are as in Section 1. $\check{\Gamma}$ is the dual of Γ, i.e., the set of negations of formulas in Γ. One could take Γ to be empty, in which case the most reasonable meaning of $\check{\Gamma}$-conservation is consistency. (Heuristically, this is because, in $Pr_\phi(\cdot)$, ϕ behaves like $\{\neg 0 = 1\}$ which has the dual $\{0 = 1\}$.)

I end this paper with a seemingly serendipitous—but lovely—application of fixed point E to a problem of relative interpretability: While ZF and GB are so intimately related as to be considered almost identical, the theories relatively interpretable in ZF are not the same as those so interpretable in GB.

Theorem *There is a sentence $\phi \in \Sigma_1$ such that*

i. *$ZF + \phi$ is relatively interpretable in ZF*
ii. *$GB + \phi$ is* not *relatively interpretable in GB.*

The existence of such a $\phi \in \Pi_2$ was first proven by Hájek [13] under an assumption of sufficient soundness. Later, Hájková and Hájek [16] reduced the requirement to mere consistency and Solovay [46] further reduced the complexity to Σ_1. The present greatly simplified proof is due essentially to Guaspari. (If one assumes $ZF \nvdash \neg Con_{ZF}$, it follows quickly from Theorem 3.3 of [10]; cf. also the discussion following Application 3.3 of [43].)

It should be noted that there are sentences $\phi \in \Pi_1$ such that $GB + \phi$ is relatively interpretable in GB, but $ZF + \phi$ is not so interpretable in ZF. (Cf. [13] and [46] for more information.)

To prove the theorem, we have to recall two basic facts about relative interpretability.

Facts: Let ϕ be a sentence, then

i. $ZF + \phi$ is relatively interpretable in ZF iff ϕ is Π_1-*con* over ZF
ii. The set $\{\phi\colon GB + \phi$ is relatively interpretable in $GB\}$ is r.e.

For the definition of relative interpretability and proofs of these facts, I refer the reader to [14], [15], and [10]. Fact i is a quick corollary of the Orey Compactness Theorem (for which see [6] and [30]). Fact ii is a trivial consequence of the finite axiomatisability of GB.

Proof of the theorem: Let $\chi v \in \Sigma_1$ define the set of all sentences θ such that $GB + \theta$ is relatively interpretable in GB and then choose $\phi \in \Sigma_1$ such that

$$ZF \vdash \phi \leftrightarrow .Pr_{\Sigma_1}(\ulcorner \neg\phi \urcorner) \preccurlyeq (Pr(\ulcorner \phi \urcorner) \vee \chi(\ulcorner \phi \urcorner)),$$

where $Pr(\cdot)$ is $Pr_{ZF}(\cdot)$. It suffices to show that $\mathbf{N} \models \neg\chi(\ulcorner \phi \urcorner)$. For then: (i) by the basic calculation, ϕ is Π_1-con over ZF; and (ii) by the definition of χv, $GB + \phi$ is not relatively interpretable in GB.

Suppose, by way of contradiction, that $\mathbf{N} \models \chi(\ulcorner \phi \urcorner)$. By the calculation, $ZF \vdash \neg\phi$, whence $GB \vdash \neg\phi$. But then $GB + \phi$ cannot be relatively interpreted in GB and $\mathbf{N} \models \neg\chi(\ulcorner \phi \urcorner)$.

A tiny remark: the disjunct $Pr(\ulcorner \phi \urcorner)$ is redundant. For, if θ is provable in GB, then $GB + \theta$ is relatively interpretable therein.

NOTES

1. I concede that, with the Recursion Theorem and the Diagonalisation Theorem coming to pretty much the same thing, many traditional sorts of results are accessible to both recursion theoretic and proof theoretic tools. However, when one desires a more sophisticated fine structure analysis, the situation is different.

2. There are beginnings: Peter Päppinghaus once showed me a method of translating recursion theoretic constructs into fixed points; but his verification was by example and he had made no analysis of the domain of applicability of his translation. Similarly, in [40], I reversed the direction by using a nested pair of pairs of effectively inseparable r.e. sets to simulate the use of the Shepherdson fixed point in a few of its applications.

3. This is shown by formalising the proof of cut-elimination in the theories and relying on the subformula property and the existence of partial truth definitions to prove by induction on the length of a cut-free derivation the truth of all theorems of the predicate calculus. For a simpler proof in the set theoretic case, cf. [39].

4. Alternatively, cf. [39] wherein Kent's application, if not the definition of the fixed point, is correctly repeated.

REFERENCES

[1] Bell, J., and M. Machover, *A Course in Mathematical Logic*, North-Holland, Amsterdam, 1977.

[2] Boolos, G., *The Unprovability of Consistency; An Essay in Modal Logic*, Cambridge University Press, Cambridge, 1979.

[3] Boolos, G. and R. Jeffrey, *Computability and Logic*, Cambridge University Press, Cambridge, 1974.

[4] Enderton, H., *A Mathematical Introduction to Logic*, Academic Press, New York, 1972.

[5] Ehrenfeucht, A. and S. Feferman, "Representability of recursively enumerable sets in formal theories," *Archiv für Mathematische Logik und Grundlagenforschung*, vol. 5 (1960), pp. 37-41.

[6] Feferman, S., "Arithmetization of metamathematics in a general setting," *Fundamenta Mathematicae,* vol. 49 (1960), pp. 35-92.

[7] Friedman, H., "The disjunction property implies the numerical existence property," *National Academy of Sciences, Proceedings*, vol. 72 (1975), pp. 2877-2878.

[8] Friedman, H., "Proof-theoretic degrees," to appear.

[9] Gödel, K., "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I," *Monatschefte für Mathematik und Physik*, vol. 38 (1931), pp. 173-198.

[10] Guaspari, D., "Partially conservative extensions of arithmetic," *American Mathematical Society, Transactions*, vol. 254 (1979), pp. 47-68.

[11] Guaspari, D., "Sentences implying their own provability," to appear.

[12] Guaspari, D., and R. Solovay, "Rosser sentences," *Annals of Mathematical Logic*, vol. 16 (1979), pp. 81-99.

[13] Hájek, P., "On interpretability in set theories," *Commentationes Mathematicae Universitatis Carolinae* (Prague), vol. 12 (1971), pp. 73-79.

[14] Hájek, P., "On interpretability in set theories II," *Commentationes Mathematicae Universitatis Carolinae* (Prague), vol. 13 (1972), pp. 445-455.

[15] Hájek, P., "On partially conservative extensions of arithmetic," in *Logic Colloquium '78*, eds., M. Boffa, D. van Dalen, and K. McAloon, North-Holland, Amsterdam, 1979.

[16] Hájková, M., and P. Hájek, "On interpretability in theories containing arithmetic," *Fundamenta Mathematicae*, vol. 76 (1972), pp. 131-137.

[17] Jensen, D. and A. Ehrenfeucht, "Some problem in elementary arithmetics," *Fundamenta Mathematicae*, vol. 92 (1976), pp. 223-245.

[18] Kent, C. F., "The relation of A to Prov $\ulcorner A \urcorner$ in the Lindenbaum sentence algebra," *The Journal of Symbolic Logic*, vol. 38 (1973), pp. 295-298.

[19] Kleene, S. C., "On notation for ordinal numbers," *The Journal of Symbolic Logic*, vol. 3 (1938), pp. 150-155.

[20] Kleene, S. C., *Introduction to Metamathematics*, van Nostrand, Princeton, New Jersey, 1952.

[21] Kreisel, G. and A. Levy, "Reflection principles and their use for establishing the complexity of axiomatic systems," *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, vol. 14 (1968), pp. 97-142.

[22] Leivant, D., "Maximality of logical calculi," *Annals of Mathematical Logic*, to appear.

[23] Manevitz, L. and J. Stavi, "Δ_2^0 operators and alternating sentences in arithmetics," *The Journal of Symbolic Logic*, vol. 45, (1980), pp. 144-154.

[24] Manin, Y., *A Course in Mathematical Logic*, Springer-Verlag, Heidelberg, 1977.

[25] Mendelson, E., *Introduction to Mathematical Logic*, van Nostrand, Princeton, New Jersey, 1964.

[26] Monk, D., *Mathematical Logic*, Springer-Verlag, Heidelberg, 1976.

[27] Montague, R., "Theories incomparable with respect to relative interpretability," *The Journal of Symbolic Logic*, vol. 27 (1962), pp. 195-211.

[28] Mostowski, A., *Sentences Undecidable in Formalized Arithmetic; An Exposition of the Theory of Kurt Gödel*, North-Holland, Amsterdam, 1952.

[29] Mostowski, A., "A generalization of the incompleteness theorem," *Fundamenta Mathematicae*, vol. 49 (1961), pp. 205-232.

[30] Orey, S., "Relative interpretations," *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, vol. 7 (1961), pp. 146-153.

[31] di Paola, R. A., "On sets represented by the same formula in distinct consistent axiomatizable Rosser theories," *Pacific Journal of Mathematics*, vol. 18 (1966), pp. 455-456.

[32] Putnam, H. and R. M. Smullyan, "Exact separation of recursively enumerable sets within theories," *American Mathematical Society, Proceedings*, vol. 11 (1960), pp. 574-577.

[33] Quinsey, J., Dissertation, Oxford.

[34] Rosser, J. B., "Extensions of some theorems of Gödel and Church," *The Journal of Symbolic Logic*, vol. 1 (1936), pp. 87-91.

[35] Rosser, J. B., "An informal exposition of proofs of Gödel's theorem and Church's theorem," *The Journal of Symbolic Logic*, vol. 4 (1939), pp. 53-60.

[36] Scott, D., "Algebras of sets binumerable in complete extensions of arithmetic," in *Recursive Function Theory*, ed. J. C. E. Dekker, American Mathematical Society, Providence, Rhode Island, 1962.

[37] Shepherdson, J., "Representability of recursively enumerable sets in formal theories," *Archiv für Mathematische Logik und Grundlagenforschung*, vol. 5 (1960), pp. 119-127.

[38] Shoenfield, J., *Mathematical Logic*, Addison-Wesley, Reading, Massachusetts, 1967.

[39] Smoryński, C., "The incompleteness theorems," in *Handbook of Mathematical Logic*, ed., J. Barwise, North-Holland, Amsterdam, 1977.

[40] Smoryński, C., "Avoiding self-referential statements," *American Mathematical Society Proceedings*, vol. 70 (1978), pp. 181-184.

[41] Smoryński, C., "Calculating self-referential statements I: explicit calculations," *Studia Logica*, vol. 38 (1979), pp. 17-36.

[42] Smoryński, C., "Calculating self-referential statements: nonexplicit calculations," *Fundamenta Mathematicae,* vol. 109 (1980), pp. 189-210.

[43] Smoryński, C., "Calculating self-referential statements: Guaspari sentences of the first kind," *The Journal of Symbolic Logic*, vol. 46 (1981), pp. 329-344.

[44] Smoryński, C., "A ubiquitous fixed point calculation," to appear.

[45] Smoryński, C., "Self-reference and modal logic," in *Handbook of Philosophical Logic*, eds., D. Gabbay and F. Günthner, to appear.

[46] Solovay, R. M., "On interpretability in set theories," to appear.

[47] Stegmüller, W., *Unvollständigkeit und Unentscheidbarkeit*, Springer-Verlag, Wien, Austria, 1959.

[48] Švejdar, V., "Degrees of interpretability," *Commentationes Mathematicae Universitatis Carolinae* (Prague), vol. 19 (1978), pp. 789-813.

Westmont, Illinois

Addendum The above is an expanded version of a lecture that I have given with some success in Oxford, Ann Arbor, and Warszawa. It has had a circuitous path to publication. At first I intended it for the Logic Colloquium at Praha. Its length dictated, however, that I give a different lecture. Thus, I submitted the paper to a major logic journal. When the colloquium organisers suggested that I include the present paper in the proceedings, I quickly withdrew it, being unsatisfied with the inordinate length of time it was taking the referee to read the paper. Then, as most readers know, the colloquium was cancelled due to technical reasons and I resubmitted the paper to the present *Journal.* I wish to make a special acknowledgment to the editors and the referee for their expeditious handling of the manuscript, allowing it to appear on the fiftieth anniversary of the publication of Gödel's paper.

It is now (February 1981) over a year since I wrote the above paper and, as one might expect, I have a few additional remarks to make. Because of the success I have had with this paper, I have decided not to tamper with the main text, but merely to add a (very) few relevant remarks in an addendum.

First, I must report that my friend Göran Sundholm, of Magdelen College, Oxford, found a terrific error in my history. It seems that Rosser's form of the Diagonalisation Lemma is already stated quite clearly on p. 91 of the 1934 edition (Springer, Wien) of Rudolf Carnap's *Logische Syntax der Sprache.* While I deny that this devastates my interpretation of the slow development of diagonalisation as a tool, I must admit that this reference does weaken my case.

Even before Sundholm savaged my history, Albert Visser and Per Lindström made observations that effectively buried Fixed Point D. These two gentlemen found easy derivations of two of my three original applications of this fixed point and the third has a similar easy derivation:

Theorem *Let T_0, T_1 be consistent r.e. theories extending (say) PA.*

i. *If $T_0 \subsetneq T_1$ and $X_0 \subseteq X_1$ are r.e., there is a formula ϕv such that ϕv semirepresents X_i in T_i.*
ii. *If T_0, T_1 are incomparable and X_0, X_1 are r.e., there is a formula ϕv such that ϕv semirepresents X_i in T_i.*
iii. *If T_0, T_1 are incompatible and (X_0, Y_0), (X_1, Y_1) are pairs of disjoint r.e. sets, there is a formula ϕv such that ϕv $(\neg\phi v)$ semirepresents X_i (Y_i) in T_i.*

Originally, I proved all of these by applying Fixed Point D. I will now show how they all quickly reduce to the uniform result: Let $T_0, T_1, \ldots$ be a uniformly r.e. sequence of consistent extensions of PA, and let X, Y be disjoint r.e. sets. There is a formula ϕv such that ϕv $(\neg\phi v)$ semirepresents X (respectively, Y) in each T_n.

Proof of the Theorem: i. (Lindström). Let $T_1 \vdash \theta$, $T_0 \nvdash \theta$ and choose

$\psi_0 v$ to uniformly semirepresent X_0 in T_0 and $T_0 + \neg\theta$
$\psi_1 v$ to uniformly semirepresent X_1 in T_0 and T_1,

and choose

$$\phi v\colon (\neg\theta \to \psi_0 v) \wedge \psi_1 v.$$

ii. (Visser). Let $T_i \vdash \theta_i$, $T_i \nvdash \theta_{1-i}$ and let

$\psi_i v$ uniformly semirepresent X_i in T_0, T_1, $T_0 + \neg\theta_1$, $T_1 + \neg\theta_0$

and choose

$$\phi v\colon (\neg\theta_0 \to \psi_1 v) \wedge (\neg\theta_1 \to \psi_0 v).$$

iii. As in ii, but assume

$\neg\psi_i v$ to uniformly semirepresent Y_i in T_0, T_1

as well.

(Remark: A careful inspection of the proof will reveal that in i one only needs the (nonuniform) existence of a correct Σ_1 semirepresentation of any r.e. set in an r.e. extension of *PA*, and in (ii), one only needs the existence of a semi-representation.)

A third point concerns my unexplained use of Feferman's dot notation. For twenty years this has been the accepted notation and, contrary to the opinion of the referee, I see no need to explain it here. The reader who is interested in the subject of the present survey will not get far in his further reading without coming across an explanation.

Speaking of further reading, let me finish by citing three important additions to my bibliography. The following papers of Per Lindström should be placed alongside Švejdar's paper as sources of nice applications of self-reference of a type not considered in the body of my paper.

1. "Some results on interpretability," in *Proceedings of the 5th Scandinavian Logic Symposium*, ed., F. V. Jensen, B. H. Mayoh, and K. K. Møller, Aalborg University Press, Aalborg, Denmark, 1979.
2. "Notes on partially conservative sentences and interpretability."
3. "On faithful interpretability."