Notre Dame Journal of Formal Logic
Volume 29, Number 1, Winter 1988

# Book Review

Kurt Gödel. *Collected Works, Volume I, Publications 1929–1936.* Edited by Solomon Feferman, et al. Oxford University Press, New York; Clarendon Press, Oxford, 1986. Pp. xvi, 474, $35.00.

This volume contains Gödel's publications that appeared by 1936. They raise two principal questions, which could not have been answered at the time, and are not prominent in the generally uninspired editorial additions either. They arise as follows.

Especially Gödel's best known (early) work was presented by reference to a sensational thesis about all mathematics being like doing sums, in a sense explained in more detail below. As with other extravagant theses, programs, ideals, or what have you, the first question is: *Why not simply ignore such things?* Next, since work on extravagances rarely suggests convincing improvements, there is a second question: *What might be done with the* (here mathematical) *tools used?*

At this point one might agonize about ignoring these two questions in turn. Instead we'll ask them first about another thesis, a thesis usually attributed to Pythagoras more than 2500 years ago: (natural) *number is the measure of all things.*

It isn't, because no natural number or ratio between such numbers measures the diagonal of the unit square, which has length $\sqrt{2}$. The tools used in this refutation are the geometric theorem of Pythagoras about right-angled triangles, and an arithmetic theorem about the (so-called diophantine) equation $n^2 = 2m^2$, which has no solution in natural numbers $n$ and $m$.

This refutation has not suggested any improvements, say, in the form of another measure for all things. On the contrary, being mundane enough to remove any sense of awe inspired by the Pythagorean thesis, the refutation raises questions about *assumptions behind the thesis*, for example:

What would be so wonderful if the thesis were true?

After all, even where only rational numbers are "needed", as in limits for experimental errors, others are used. The interval for a measured length of some diagonal of a unit square is often given in the form $\sqrt{2} \pm \epsilon$, with rational $\epsilon$ and thus irrational endpoints. More generally,

> Is it sensible to demand one measure for all things?

rather than relatively few measures for relatively many things? So to speak in the opposite direction:

> Are there phenomena, possibly far-removed from everyday experience, of Pythagoras or even of ourselves, that do lend themselves to one, as it were, fundamental measure?

and so forth.

Similar questions come up about the thesis refuted by Gödel's work, but also analogues to the sociological fact that a few bands of the faithful continue to pursue the Pythagorean thesis, in numerology or in reductions of mathematics to arithmetic.

As to the tools mentioned earlier, it would be unrealistic to try to be precise about cause and effect in the last 2500 years. (Did the yodeler or the echo trigger the avalanche?) As somebody said, such matters tend to be difficult just because they have so few consequences. Be that as it may, the tools used remain *memorable samples.*

Readers of Gödel's papers should not expect similarly colloquial language. Partly this is a matter of temperament. But also, those of us who know the detailed analyses made in the meantime can now judge which familiar ideas are typical enough to illustrate a particular general issue reliably. Occasional uses below of erudite language will serve as reminders of those analyses; for example, instead of 'doing sums', there will be 'formal procedures', that is, computations according to the logical idea(lization) of the perfect computer.

## *1 Background on formal or, equivalently, mechanical aspects of mathematics*

What are they and what are they supposed to do? (These words, but not the specific answers below, come from Dedekind's *Was sind und was sollen die Zahlen?*) Computations with 0, 1, +, ×, as done in elementary school, are quite typical of formal procedures. Appendix 1(a) sketches both the drill involved and shortcuts resulting from—humanly inevitable—reflection on it.

Today, computations on an electronic computer are familiar, and are even better examples, with one proviso: *not* too much attention to details either of the hardware or of the wetware (computer jargon for 'intellect') since the logical idea of the computing machine corresponds only to Simple Simon's image of computing Man.

Both kinds of examples convey very well what formal procedures *are*, but they do not provide an effective background for Gödel's (best known) result. This states some odd things that even an ideal computer can*not* do, which are wholly overshadowed by the many things that even actual computers *can* do; realistically speaking, far beyond Leibniz's dream.[1]

The extravagant thesis about doing sums provides, of course, a much more glamorous background. The first of the details about that thesis, promised at the outset, is a *distinction*, between thought processes in mathematics and thoughts (as in "it's a thought"), or, more simply, mathematical results. The pioneers, especially Frege, concentrated on the logical relations between results; without any claims on the full range of the mathematical imagination, or even dismissing questions about the latter as irrelevant (to mathematics).[2] Brief indications of this distinction, applied to addition and multiplication, are in Appendix 1, particularly (a) (ii)-(iii), and (c). The remainder of this section is more relaxed.

What is at issue is an understanding or theory of reasoning, at least, in mathematics. And the claim of the thesis is that the formal elements reflect all that is significant. Thus a theory takes the form of a *formalization*, which consists of a formal language with a basic alphabet, a formal grammar, and formal rules of derivation. Appendix 1 provides a sample, but the formalization of elementary (predicate) logic, which appeared more than one hundred years ago, remains more impressive; above all because of the expressive power of its simple notation (vocabulary and grammar). This became relatively soon part of mathematical, and even general, culture; much more so than its rules of inference, inevitably reminiscent of processes.

Let there be no mistake: even the formalization for doing sums is amazingly simple compared to the phenomena that present themselves naturally; specifically, in the many nuances of natural—written, and especially spoken—mathematical language(s). The price to pay for this simplicity is a malaise: the formal elements constitute a *very pale picture* of even the two elementary parts of mathematics in the last paragraph, let alone of broad mathematical experience.

One consolation, albeit overlooked by Goethe himself, is implicit in 1.2037 of *Faust I*: *all* theory is grey. So paleness by itself is not a defect of any theory, logical or not. More substantial encouragement comes from the mechanical picture of the physical world around us in terms of point masses and their motion in space-time, which leaves out colors and shapes, not to speak of chemical composition. This picture, hardly less pale than the formal picture above, has not only been most successful in its domain, but remained for centuries a model for theoretical understanding. Readers of Appendix 2 may pursue the parallel with the formal picture(s) of the world of mathematics a step further: not only are the objects of mechanics here represented, but also spatio-temporal relations between them; cf. Appendix 2(a)(ii) on Gödel's twist representing properties of Cantor's numerical representations of words.

Memorable formalizations were proposed in the last decade of the last and the first decade of the present century; by Whitehead and Russell for *all of mathematics* in the three volumes of *Principia Mathematica*, by Hilbert for various *branches* of mathematics beginning with elementary geometry; with respect for both the venerable ideal of purity of method and for the mathematical tradition of concise exposition. But for the sequel, and probably *sub specie aeternitatis*, the following difference between the two styles is much more consequential.

In his so-called metamathematics Hilbert paid attention to global *mathematical properties* of the formal pictures such as completeness, taken up in the next paragraph. *Principia* did not; in the tradition of natural history, which is content with a compact description of data that happen to catch our attention; by mathematical formulas when it uses mathematics at all. Admittedly, this side of metamathematics got lost in Hilbert's later rhetoric, for example, about real and ideal elements in the tradition of Ockham's razor, already mentioned in Note 2.

A specific principal requirement on formal pictures is this. For propositions $P$ represented by, say, $p$ in the formalization considered, $p$ should be a formal theorem iff $P$ is true. When $P$ is about some particular notion or 'structure' then either $P$ is true or its negation (is true). The corresponding mathematical property of the formalization is that *either* $p$ is a formal theorem *or* the formal negation of $p$ (is a formal theorem). This is called *formal completeness*: cf. Appendix 1(b)(i).

In general, the restriction to particular notions is needed, as seen in the case of pure logic where, as Leibniz put it, truth in *all* possible worlds is meant. So $P$ is not logically true if it is false in some such world. But the negation of $P$ is logically true only if $P$ holds in *no* possible world. Thus neither $P$ nor its negation need be logically true, and so formal completeness is not required here. Nor is it required when $P$ is about, say, sets, before we have made up our minds on the particular kind of set to be considered; for some $P$, neither $P$ need be true for all kinds of sets contemplated, nor its negation. Perhaps it is worth adding that, even when $P$ is about some specific structure, formal completeness would not be generally even plausible, if thought *processes* were the main object of study. There is simply not a shred of evidence that every problem of, say, Higher Arithmetic is solvable in any even remotely realistic sense, let alone that we should want to look at every problem (in current formalizations).

***2 A refutation and a Pyrrhic victory*** Gödel's—most famous— incompleteness theorem was originally stated for *Principia* and related systems; in fact, for the parts that serve to represent arithmetic properties. For each such system $S$

some true proposition of arithmetic is not a formal theorem

(of $S$), where the proposition depends upon $S$. Thus, doing higher arithmetic is *not* like doing sums (cf. Appendix 1(b) and (c)). Readers who have any taste for doing mathematics at all can probably get a good idea of the proof from Appendix 2.

The particular true proposition obtained there has the following simple interpretation. It is (the so-called arithmetization of) an instance of the principle *each formal theorem is true* (cf. Appendix 2(c)(i)).

For our ordinary view of—the logical relations between results in—mathematics, the principle above is a minimal consequence of understanding the rules of $S$ at all (and, it may be added, rules actually used are understood, at least, enough for this consequence). Viewed this way, $S$ would simply be said not to prove the particular property of itself expressed in the principle, though $S$ proves many things, also about itself. Thus *the thesis is refuted according to the letter* since one of its explicitly formulated (cl)aims, completeness for arithmetic, is not realized.

But for grand theses, and especially for ideals—here of theoretical understanding—a broader sense of 'refutation' is philosophically appropriate:

*When the ideal is realized the realization is found to be unsatisfactory.* To spell it out: The realization is found to lack or have certain properties that, as can *now* be seen, had been tacitly assumed to go with, respectively against, the ideal. In this way inspection of a realization can identify *tacit assumptions behind the ideal,* which constitutes a refutation in the broader, second sense, by a Pyrrhic victory, as it were.

An example of such a refutation of the thesis is Gödel's other very well-known result in this volume: the completeness of Frege's formalization for elementary predicate logic, but now in the sense that for each formula *F, either F* is a formal theorem *or F* is not true in all possible worlds.

Viewed dispassionately this result does not at all give a privileged place to Frege's rules. It just shows that they are sufficient-in-principle, as Kant liked to say, or, more soberly, sufficient to generate the logically true formulas in the formalization. On the contrary, by using the concept of logical truth the result draws attention to quite different possibilities of proving logical theorems; specifically, the possibility of drawing on knowledge—if not of all possible worlds, at least—of many corners of our world; in other words, the possibility of proving logical theorems by so-called logically impure methods. Specialists will think here of theorems that are or can be formulated in Frege's logical language, for example, about ordered or real closed fields, but are proved by topological methods.

In point of fact such impure methods have been used increasingly since the formalization of pure logic more than one hundred years ago, and especially since Gödel's result more than fifty years ago; incidentally, this is often done by people totally ignorant of the formal rules (so that it is wide open in which way the conviction carried by their proofs can be realistically related to the formalization).

Before turning to more positive aspects of the two refutations above by the incompleteness and completeness theorems, a couple of comments seem in order. The first is general. The proofs of both theorems are pearls of logic. But not both results can be sensational! If the completeness of a formalization for mere elementary logic is a sensation, then incompleteness of Higher Arithmetic is not.[3] Historical counterfactuals aside, one hundred years ago it would have been fitting to give pride of place to the lesser known result, bolstering up the then-tentative project of formalization, with the incompleteness theorem ratifying, formally as it were, the then general distrust of that project.

The second comment concerns the open secret that, outside mathematical logic, neither of those two very well-known results turns up in the ordinary mathematical literature. There is no conspiracy against them. As stated, they just have not found a use; fittingly, in view of the fact that they are tailor-made for—refuting or supporting, no matter—a *refuted thesis.* This is the situation considered at the outset, with all its problems, in particular the problem of finding some sober use for the tools employed.

Especially when, as in the present case, the tools are of obvious "raw" interest, the principal obstacle to solving those problems is blindness to them; in particular, the illusion that pretty mathematics must "somehow" have already solved those problems. *Tractatus* 6.21, about mathematics not expressing any thought, is surely literally false. But equally surely mathematics very often leaves questions open that require more (demanding) thought than the mathematical solution.

*3 Shifts of emphasis* What more do we know from formalization? Obviously this question does not even arise if the thesis is elevated, as indeed has been done, to the doctrine that there is nothing (precise) besides formalization.

The question is less innocent than it may look. For effective contributions to some particular area the additional knowledge—in answer to: What more . . . ?—must be expressed in terms used in that area, and is thus liable to require familiarity with it. The following two general points have a larger market.

First, there is the matter of choosing a formal system rationally. By Gödel's incompleteness theorem the only general idea for reducing the arbitrariness of such a choice, completeness, is not even 'in principle' available for systems of arithmetic. One of the current favorites for answering the question above provides a better, and sometimes even practical, idea for a choice, based on the *rate of growth* of those functions that can be shown in $S$ to solve suitable problems, say, $P$ provided one wants to know about such things. Here, *$P$ comes first*, and *$S$ is a tool*, not a 'foundation'. For specialists: 'Suitable' means in practice that $P$ has the form $\forall\exists$, and that incompleteness applies even if all true, purely universal propositions are added to $S$ as axioms. The bounds are established by so-called consistency proofs for those—necessarily incomplete—$S$.

The second point concerns complete formalizations such as (Frege's) logical rules. Unless—to repeat what cannot be repeated too often—formalization is required as a matter of doctrine, the question What more . . . ? is just as hot here as for incomplete systems. The answer is simply: *a new description of the object involved*; in the case above, of the logically true propositions (in Frege's language). For effective knowledge this description competes then with others such as the "impure" kinds in the last section. Logicians think here of the competition between descriptions in model-theoretic and diverse proof-theoretic terms. There is an obvious parallel here with the Pythagorean thesis, specifically with the use of irrationals even when they could be avoided-in-principle.

But a more significant element of the parallel is this: the proofs of completeness and incompleteness are "mundane enough to remove any sense of awe inspired by" the refuted formal(ist) thesis, and leave us free to examine some *assumptions behind* the latter.

*4 What is so wonderful about formalization?* People have been thrashing about for an answer. Only one will be considered here. It is the tacit assumption of some ethereal need—here satisfied by formalization—for an ultimate norm of precision; a tacit assumption popular not only in the foundations of mathematics but almost throughout all Western culture.

Presumably according to the books they read in their teens, particular authors writing on such norms refer to the finiteness of formal objects, to their spatio-temporal or, more generally, their public character, or, going the whole hog, to the idea that the thought processes themselves are formal (i.e., mechanical in the sense of the perfect computer), in which case only formal rules can be unambiguous. All this would carry little weight without the master assumption that there are *genuine doubts about the precision or reliability of principles currently in use*. (Viewed this way, all those little paradoxes are a godsend for this assumption.)

Now if the principles in question are in fact 100% reliable, doubts about them are dubious, which doubts can be just as much as can assertions. The privileged place given to doubts, for example, by Descartes, looks very much like other pious conventions that are only too familiar. But it pays to be more specific.

To put first things first, there can be perfectly proper doubts, even about principles. Thus, not so long ago, – antecedents of – those now current about sets were problematic. The problems were solved by saying out loud which sets were meant (and not by putting principles into formal dress which, after all, Frege did in his *Grundgesetze*); cf. the end of Section 1 on completeness. Though, occasionally, obviously problematic principles are investigated, most often those proper doubts are about – the probability of – incorrect applications of correct principles. In this case preoccupation with reliability-in-principle, that is, reliability of principles, simply *distracts from the dominant factor*, here, dominant source of error.

Is this factor a foundational concern? According to a principal tradition of the subject, it is not. In this case the topic of reliability has been *discovered* not to be primarily foundational; by the way, not necessarily a comedown; see the end of Appendix 2. A discovery of the nonfoundational character of a topic – here of reliability – is to be compared to discoveries in the natural sciences; for example, of the gravitational or magnetic character of some phenomenon, say, near the surface of the earth; in other words, whether the dominant force, if there is *one*, is gravity or the magnetic field of the earth. Similarly, in the remark for specialists in Section 2 concerning "impure" methods, problems stated in logical language but solved by topological methods are thereby discovered to have topological character. As in other scientific experience, research has produced ways both to cope with such delicate points as mixtures and, above all, ways to recognize when *enough is enough*: that is, enough for using such characterizations without looking for new evidence or even referring to the old. (Our existing knowledge of those characterizations is thus treated as a priori.) Given the level of generality of all this, it applies of course to norms as well, here of precision but surely also in practical life. For the literal sense of "philosophy", recognizing when enough is enough has always been a central concern. Thus in the *Metaphysics* Γ 4, 1006a, 6-9, Aristotle relied on good breeding. The following couple of points help too.

First, there is extended experience, which may but need not confirm ordinary practice. For example, as already explained in Section 2, experience has rehabilitated our ordinary 'norms' or, more simply, checks in the case of logical reasoning, as opposed to the demands of so-called formal reasoning. In contrast, after the discovery in the last century of so-called abstract reasoning we have never looked back. Specifically, we express in axiomatic terms what we feel to be essential about an argument. For example, in elementary number theory analysis in terms of abstract finite groups is used; we do not isolate only, say, the 'numerical content'.

Secondly—and this is particularly significant for the present section—a good deal of the foundational literature obscures the consequences of such extended experience, including logical experience. Thus, it wails about "far-reaching reductions" that are (allegedly) lost by the refutation of the thesis here considered, apparently without remembering the fact that those would-be magic reductions *are already available* in substantial areas—such as elementary geometry or logic—but *not used*: least of all, as a norm of precision.

Readers familiar with foundational (bad) habits surely know much more along these lines. But the snippets above are enough to show that the topic of reliability is not a rewarding market for formalization, and certainly less so than the answers in Section 3 to the question: What more . . . ? This suggests also a genuinely philosophical conclusion.

It is simply not plausible that this particular topic will get very far without closer attention to thought processes, which, by Section 1, are neglected in the pale formal picture. (See also the Remark at the end of Appendix 1.) *Warning* for readers of pages 394-399 in the volume under review: The heading "length of proofs" suggests consequences for understanding thought processes; but the formal picture is just too pale to support such colorful interpretations! A short description of a long formal proof, especially with underlining of its memorable parts, is in fact easier to process than a relatively short formal proof. In a similar vein, the process of *discovering a new axiom* to prove some given result is often not only less demanding than *discovering a proof* of it *from given axioms*, but has a similar flavor. Though such a proof by given axioms can in principle be found mechanically by trial and error, in practice it is not so found.[4]

*5* ***Short answers to the initial questions*** The refutations, including the Pyrrhic victories, of the thesis were involved in locating and examining assumptions behind it; such as dubious doubts or doctrinaire norms of precision. Admittedly, those assumptions can be and were questioned before settling or even formulating the thesis. But the refutation and, particularly, its elementary character (stressed in Appendix 2(c)) can help establish *a sense of proportion for the examination*: above all, by eliminating undue worry about not having grasped the full inwardness of the thesis.

As to the tools, sketched in Appendix 2, the situation is not too different from that of the Pythagorean thesis except that sixty, not 2500, years have passed. This affects both the choice of problems and the methods of solution. Thus there is a greater difference between the proofs of Mordell's conjecture, a recent highlight, and the irrationality of $\sqrt{2}$ in the theory of diophantine equations than between, say, "double" diagonalization in so-called priority arguments of Higher Recursion Theory and simple diagonalization in Appendix 2(b). But Gödel's proofs remain *memorable samples* in the second subject.

For the record, I still find his direct and self-assured style in the volume under review appealing, compared not only to the extremes of pedantry and sloppiness rampant at the time but also to some constipated parts of the editorial material. The so-called substance, of proofs and results, has been superseded, in accordance with Buffon's *ces choses sont hors de l'homme* (those things—

meaning results—are impersonal), which preceded his much better known pronouncement (on August 25, 1753): *le style est l'homme.*

So much for logic, the business of this journal, at least according to its name. But logicians do not live by logic alone—not even intellectually. To me, a principal reward for refuting rather than ignoring extravagant theses is in that broader area, as follows.

*6* ***Refutations: For a better quality of life*** Evidently, a parallel with material pollution is meant, which presents a similar choice. In the austere 50's those complaining of noise or other pollution were told to ignore it. Progressives, always characterized more by temperament than by specific views, had only contempt for the (to them, obviously absurd) Air Force general played by George C. Scott in the film *Dr. Strangelove,* who worried about chemicals in drinking water. And for all I know they may have a point if selection by resistance to pollution helps the species flourish in our cold, unfriendly universe. But for some (of us) it was not so easy to ignore the pollution.

Similarly, (we) logically sensitive souls do not so easily ignore the logical atrocities in (Hilbert's) presentations of the thesis, especially when research stagnated and the claims inflated. The pollution was all around, spread by a band of the faithful who found the presentation so congenial that it matters little whether it helped form or "only" consolidate their views. Here are a couple of samples.

(a) The laws of thought are mechanical and *Non ignorabimus.*

Actually, the idea was that those laws were already formulated in Hilbert's systems, and that we shall—want to—know the answers to all the problems formulated there.

As a kind of fallback Hilbert had a weaker, would-be cute meaning for '*non ignorabimus*':

(b) It is consistent to assume that every problem can be solved.

In other words, to assume for every proposition $P$ that either $P$ is a formal theorem (of the "foundational" system) or its negation is.

Gödel's results refute (a) and (b) conclusively and most elegantly. For (a) this is clear without further analysis. For (b), a corollary, derived in Appendix 2(c)(ii), is needed that is valid only for a more special class of systems, but certainly for all current at the time:

(b′) It is even consistent to assume that every proposition can be proved,

from which (b) follows trivially. We breathe more freely.

This relief is hardly necessary in practice; not, for example, for the robust among us who ignored the thesis in their work even if they gushed about it in private. Philosophically, the relief is *only* a palliative. It just distracts from the source of the pollution, the assumptions behind the thesis. But, as seen by Section 4, cleaning this up is a costly business, requiring more capital (i.e., scientific experience) and labor (sustained reflection). Palliatives have a wider market.

To pursue the parallel a little further, but also to end on an irreverent note, it may be remembered that similar technologies can create and clean up pollution, and that quite often the same firm manufactures both types. In the 50's, Gödel himself (as so often in this review, in effect, not in these words) modified (a) above by adding *nil nisi externum* (to *non ignorabimus*: what he actually asserted was that we know everything about our own constructions). Then his incompleteness result implies the hot news:

> Either mind is not mechanical or the natural numbers are not our own construction

nor (*pace* Kronecker) the reals. For the record, I don't choke at that emission, but almost cherish it, together with memories of many conversations with him, spiced with spontaneous twists of a similar flavor.

But all in all it's fair to say that elementary metamathematics and, particularly, Gödel's contributions are good value; at least for those who for one reason or another have learned about them. What seems to me wide open is how effective those contributions are for conveying *that which is of general interest in them*, in particular, that which is genuinely generally accessible; effective compared to metaphors from more widely available knowledge, as in the Pythagorean thesis. The matter is wide open because my skepticism seems to be shared by others in the trade, and so existing (unsuccessful) attempts at popular exposition have been perpetrated by the uninformed in line with the proverb: . . . where angels fear to tread.[5]

## *Appendixes*[6]

### *1 Doing sums formally*

**(a) Numerical equations:** The *alphabet* (i.e., symbols): 0, 1, +, $\cdot$, (, ), =.
*Terms:* 0 and 1; $(t + t')$ and $(t \cdot t')$ if $t$ and $t'$ are terms.
*Equations* are of the form $t = t'$ where, here and below, $t$ and $t'$ stand for arbitrary terms.
*Axioms:* $1 = (0 + 1)$ and substitutions of terms for $a$ and $b$ in the schemata (i.e., recursion equations)

$$(a + 0) = a, \ (a + (b + 1)) = ((a + b) + 1);$$
$$(a \cdot 0) = 0, \ (a \cdot (b + 1)) = ((a \cdot b) + a)$$

*Derivations* are finite partially—or, for ordinary writing, linearly—ordered sequences of equations $E$ such that $E$ is *either* an axiom *or* obtained by substituting $t$ for one or more occurrences of $t'$ in $t_1 = t_1'$ if both $t = t'$ and $t_1 = t_1'$ precede $E$ in the sequence. This rule of "inference" is called substitution of equals for equals.
*Exercises:* Derive (all instances of) $a = a$. *Hint:* $(t + 0) = t$ is an axiom. Substitute $t$ for $(t + 0)$ in $(t + 0) = t$.—From (any derivation of) $t = t'$ infer $t' = t$. *Hint:* Substitute $t$ for the right $t'$ in $t' = t'$.

(i) *Numerals and numerical values:* By definition, the numerals are 0, (0 + 1), ((0 + 1) + 1), . . . ; that is, if $\bar{n}$ is a numeral the next numeral is $(\bar{n} + 1)$.
*Exercise:* For each term $t$ there is a numeral, $|t|$, such that $t = |t|$ is a

formal theorem. *Hint:* 0 is itself a numeral. $1 = (0 + 1)$ is an axiom. Note that for numerals $|t|$ and $|t'|$, $(|t| + |t'|$ and $(|t| \cdot |t'|)$ can be proved to be equal to numerals, and follow the buildup of terms.

(ii) *Reflection on numerical values*, when the symbols of the alphabet are given their usual arithmetic meaning (parentheses being part of the notation for addition and multiplication). The words *completeness* and *soundness* of formal rules, used in the text, mean here that exactly the true equations $t = t'$ are formal theorems.
*Exercise:* Prove this. *Hint:* The axioms are true and the rules preserve truth. So all formal theorems are true. But $t = t'$ is true only if $|t|$ and $|t'|$ are identical, in which case $|t| = |t'|$ is an instance of $a = a$ and thus a formal theorem.

(iii) *Reflection on possible orders of applying the rules of (a)* for computing the numerical value, say, of $(0 \cdot t)$. Like other formal rules, those of (a) leave a choice in their order of application. Knowledge of arithmetic properties—here, that $|(0 \cdot t)| = 0$—helps in an efficient choice.
*Exercise:* $(0 \cdot (b + 1)) = (0 \cdot b)$. *Hint:* For $a = 0$, $(a \cdot (b + 1)) = ((a \cdot b) + a)$ becomes $(0 \cdot (b + 1)) = ((0 \cdot b) + 0)$, which becomes $(0 \cdot (b + 1)) = (0 \cdot b)$ when $(0 \cdot b)$ is substituted for $a$ in $(a + 0) = a$.
*Conclusion:* When $t$ has the form $(t' + 1)$, do not compute $t'$ first, as in the exercise of (i), but use $(0 \cdot (t' + 1)) = (0 \cdot t')$ as the first step in computing $(0 \cdot t) = 0$.

**(b) Numerical inequalities: $\mathbf{t \neq t'}$.** One *additional axiom* schema: $(a + 1) \neq 0$. *Rules:* substitution of equals for equals is extended: from $t = t'$ and $t_1 \neq t_1'$ infer any inequality obtained by substituting $t$ for $t'$ in $t_1 \neq t_1'$. Also, a new rule: Infer $(t + 1) \neq (t' + 1)$ from $t \neq t'$. (This corresponds to the cancellation rules for equations: Infer $t = t'$ from $(t + 1) = (t' + 1)$, which is superfluous in the sense that the rules of (a) are complete without it.)

(i) *Completeness* of the rules for inequalities. *Exercise:* Prove it. *Hint:* If $t \neq t'$ is true, one term, say $|t'|$, of the pair of numerals $|t|$ and $|t'|$ is a proper part of the other. Let $|t_1|$ be (the numeral of) the difference. Then $|t_1| \neq 0$ is an instance of $(a + 1) \neq 0$, and $|t| \neq |t'|$ is inferred from $|t_1| \neq 0$ by ($|t'|$ applications of) the new rule.

**(c) Diophantine questions,** mentioned already in connection with $\sqrt{2}$, concern equations between polynomials with numerical coefficients; in the notation of **(a)**, such polynomials are the terms generated by adding the "variables" $x_1, x_2, \ldots$ to the "constants" 0 and 1. Diophantine questions ask whether or not an equation has a solution by natural numbers. For *positive* answers, the rules in **(a)** are enough since, if $x_1, x_2, \ldots$ are—the numerals of—solutions, this fact is verified by computation. But there are simply *no formal rules* at all that *are*—correct and—*enough for all negative answers*, so-called diophantine inequalities; recall $n^2 \neq 2m^2$ in connection with $\sqrt{2}$. For specialists: negative response to Hilbert's demand for such rules (in his tenth problem).

Thus, even if only logical relations are considered, *solving diophantine equations is not like doing sums.*

(i) There is a plausible parallel between diophantine questions and metamathematical questions about derivability by formal rules. If a formula *is* derivable, this fact is verified by (nonnumerical) computation with formal objects. Underivability is usually established by use of specific properties of the formula considered, as illustrated by non-Euclidean models in the case of Euclid's fifth postulate. Specialists know a precise sense of this parallel from work on Hilbert's tenth problem above.

**(d) Thinking about sums:** facts of experience. Building up derivations, for example, by following the rules in **(a)** mechanically, is more exhausting than, say, reflecting on shortcuts, and thus, realistically speaking, more liable to error. This fact is dismissed in foundations as human weakness and thus as irrelevant to logic. Perhaps; but if so, the broader philosophical topic of human data processing just does not have a (primarily) logical character; in particular, those "weaknesses" may be of the essence in determining the extent to which human data processing is *not discrete.*

*Remark* for readers familiar with—the literature on—Wittgenstein's queries about *following* (mechanical) *rules correctly:* His wording is purely logical since it concerns the idea of a correct application of a given rule (like Kant, *Kritik der reinen Vernunft*, A 132-133), and possibly irreconcilable conflicts over different interpretations of that idea in certain imagined situations. Actually, computational errors do occur, with the difference that their presence is recognized (even if no erroneous step is located!) and so those (imagined) conflicts are rare. However, it is to be noted that the errors are of a kind that would not at all be expected to occur frequently in wholly discrete data processing.

## 2 *Two twists by Gödel on Cantor's enumerability results*

**(a) Words of a numbered alphabet and relations between them:**

(i) Cantor's enumerations of pairs and of finite sequences of objects in an enumerated set are familiar. Reminders: Rationals (pairs of integers), algebraic numbers determined by the sequence of coefficients of their primitive equations; cf. Appendix 1(c). Formal objects like those of Appendix 1 are sequences (of letters in the alphabet of the system used).

Cantor himself did not pay attention to the numerical properties and relations that correspond to those for numbered sequences. Below, two operators (i.e., functional relations), $\lambda_*$ and $\sigma$ will be used where

$\lambda_* w = *w$, that is, the sequence resulting from putting the element $*$ in front of $w$, and

$\sigma(w, v)$ is the result of substituting (the sequence) $v$ for some chosen element in $w$.

But given a numbering it is, in practice, a matter of routine to write down the corresponding operations, which pass from the numbers of $w$ and $v$ to the numbers of $\lambda_* w$ and of $\sigma(w, v)$. Here "in practice" means that any of the familiar arithmetic operations, like exponentiation, are used; in any case, polynomials alone, the subject of Appendix 1, would not be enough.[a]†

The following twist by Gödel involves something new, even if it is little more than remembering—the possibility of—incompleteness, in other words, a difference between truth and formal provability.[b] It concerns:

(ii) *Two kinds of formal representation*, specifically, of properties of numbers (of formal objects); here, the property $T$ of being the number of a *formal theorem* of the system considered. As in Appendix 1(a), $\bar{n}$ is the numeral of $n$. Let $F_T$ be a formula with one free variable, here chosen to be $a$. Then, here and below, $F_T$ is said to represent $T$ numerically iff, for each number $n$,

$T(n)$ is true iff $F_T(\bar{n})$ is a formal theorem.

*Exercise:* Verify that in an inconsistent system, where every formula is a formal theorem, only one property is representable (which holds for all $n$). End of exercise. Let $\neg$ be formal negation.

For a complete (formal) system, where either $F_T(\bar{n})$ or $\neg F_T(\bar{n})$ is a formal theorem, it follows that

(*) $T(n)$ is false iff $\neg F_T(\bar{n})$ is a formal theorem.

To underline the point the literature uses a new word for representations that also satisfy (*): originally, *entscheidungsdefinit*: today, more often *invariant* definitions.

The idea is extended to *sequences*, $\vec{P}$, of properties $P_1, P_2, \ldots, P_m$. A formula $F_{\vec{P}}$ with two variables represents $\vec{P}$ if, for all $m$ and $n$, $P_m(n)$ is true iff $F_{\vec{P}}(\bar{m}, \bar{n})$ is a formal theorem. Evidently, even if each member of a sequence has a representation the whole sequence need not be representable by any formula of the system; for example, in Appendix 1(c), each polynomial is represented, but no sequence that includes all the polynomials. This warning serves as a foil to

*representing a sequence of all representable properties*

by use of $F_T$ and a variant $\bar{\sigma}$ of the substitution operation $\sigma$ in Appendix 2(a)(i), where now not the word with number $n$ but the numeral $\bar{n}$ replaces the privileged variable $a$. (In the notation of Appendix 2(a)(i), $v$ is $\bar{n}$, and $a$ is the "chosen element".)

If $P_m$ is the property represented by the formula with number $m$ then

$P_m(n)$ is true iff $F_T[\bar{\sigma}(\bar{m}, \bar{n})]$ is a formal theorem.

*Reminder:* If the system considered is complete, the representation is a definition, and so $F_T[\bar{\sigma}(m, n)]$ would *define* an enumeration of all representable properties, a notion familiar from Cantor's cardinal arithmetic.

†Superscript letters refer to the Addenda, p. 175.

**(b) A formal counterpart to Cantor's diagonal construction:** As usual, two properties of numbers are called *different* iff *some number has one of the properties, but not the other.*

(i) For every sequence $\vec{P}$ of properties there is a property, say $D_{\vec{P}}$, depending of course on $\vec{P}$, that is different from each property of the given sequence. Let

(**) $D_{\vec{P}}(n)$ be true iff $P_n(n)$ is false.

Consider any $P_m$. It differs from $D_{\vec{P}}$; specifically, at the argument $m$, which, by (**), has the property $D_{\vec{P}}$ iff it does not have the property $P_m$.

*Remark:* The literature sometimes speaks of "self-reference" here. This is literally true since an argument where $D_{\vec{P}}$ differs from $P_m$ is the subscript of $P_m$ itself; after all, for most function terms their evaluation at the argument $n$ refers to $n$. Psychoanalysts may speculate on the fact of experience that the word has clouded the critical judgment of many, but the practice itself is harmless. We return to Cantor.

For a long time the principal use of the passage (i.e., the operation) $\vec{P} \vdash D_{\vec{P}}$ was made in cardinal arithmetic, Cantor's pet. Specifically, in contrast to (say) the set of all algebraic numbers, and like the set of all real numbers, *the set of all sets of natural numbers is not enumerable.* An enumeration is nothing else but a sequence, and it would not include the corresponding diagonal set (**).

Long before the representation of all representable sets (by Gödel) at the end of Appendix 2(a)(ii), Cantor's argument caused malaise, and people thrashed about for ways of expressing this malaise. There were those dubious doubts about the existence of (the) uncountable sets mentioned above, but also talk about the language in which they are defined, although cardinal arithmetic is not restricted to sets that happen to be specified in any particular language.

Similar words—but with quite a different meaning!—get a point in the twist from cardinal arithmetic to formal representations.

(ii) Gödel's twist on Cantor's diagonal construction is applied to the sequence represented by $F_T[\bar{\sigma}(m,n)]$.[c]

*If* the system considered were complete, the diagonal set of that sequence would be defined by $\neg F_T[\bar{\sigma}(a,a)]$ or, equivalently, by $F_T[\bar{\sigma}'(a,a)]$, where $\bar{\sigma}'(a,a)$ is short for $\lambda_\neg[\bar{\sigma}(a,a)]$ introduced in Appendix 2(a)(i).

Now, the formula $F_T[\bar{\sigma}'(a,a)]$ has a number, say, $g$. By the diagonal construction this purported representation of the diagonal set is certainly not a definition for $a = g$, whence incompleteness with respect to $F_T[\bar{\sigma}(\bar{g},\bar{g})]$.[d]

This is a little slick, and it pays to *interpret* the formulas used. Write $G$ for $F_T[\bar{\sigma}'(\bar{g},\bar{g})]$, which has the number $\bar{\sigma}(\bar{g},\bar{g})$. Suppose $G$ is a formal theorem. Since, by Appendix 2(a)(ii), $F_T$ represents the property of being a formal theorem, the formula with number $\bar{\sigma}(\bar{g},\bar{g})$, namely, $\neg G$, is also a formal theorem—a contradiction.

Conversely, suppose $\neg G$ is a formal theorem. Then, by the other direction of Appendix 2(a)(ii), $F_T[\bar{\sigma}'(\bar{g},\bar{g})]$, that is, $G$ is also a formal theorem.

*Remarks:* The twist above is only implicit in Gödel's famous paper. It is explicit in Turing's work in 1936, but also in Gödel's letter to Zermelo (1931), which can be found in [1], but not in the volume under review. Secondly, readers familiar with the literature are warned that, instead of $G$, usually the formula $\neg F_T[\bar{\sigma}(\bar{g}_0,\bar{g}_0)$ is used where $g_0$ is the number of $\neg F_T[\bar{\sigma}(a,a)]$ above.

(c) **Further interpretations** of (b) above in terms of two related metamathematical properties of formal systems:

(i) *Each formal theorem is true:* This is applied in Section 2 of this review to the formula $\neg G$ with number $\sigma'(g,g)$. Thus $G$, short for $F_T[\bar{\sigma}'(\bar{g},\bar{g})]$, represents "$\neg G$ is a formal theorem" and, as usual, $\neg G$ means "$\neg G$ is true". So the instance of the principle used is expressed by $G \to \neg G$, which reduces to $\neg G$.

As a corollary, (even) the particular case of the principle above applied only to $\neg G$, is not a formal theorem of the system considered.

(ii) *Consistency* is applied in Section 6 to the pair of formulas $G$ and $\neg G$. (As usual a system is called consistent if such pairs are not both formal theorems.)

First, the principle (i) applied to both $G$ and $\neg G$ implies that not both $G$ and $\neg G$ are formal theorems (since not both are true). Evidently, here no specific restriction on $G$ is used.

Secondly, for systems that are *complete for G*, that is, for which $G \to F_T[\bar{\sigma}(\bar{g},\bar{g})]$ is a formal theorem, consistency implies the principle (i) for $\neg G$. For, by consistency, even assumed only for the pair $G$ and $\neg G$,

$$F_T[\bar{\sigma}(\bar{g},\bar{g})] \to \neg F_T[\bar{\sigma}'(\bar{g},\bar{g})], \text{ for short, } F_T[\bar{\sigma}(\bar{g},\bar{g})] \to \neg G,$$

which, together with the completeness for $G$, leads to $G \to \neg G$; cf. Appendix 2(c)(i). Here two special properties are used: that of $G$ being, by Appendix 1(c)(i), *like* solvability of a diophantine equation; and of course of the system considered, which must prove of itself that verification by computation *loco citato* is possible.

As a corollary, for *such* systems *S, their consistency is not a formal theorem*, and so it is consistent to assume (that is, add to $S$ the axiom) that $S$ is inconsistent.[e]

For use in (b′) of Section 6 it should be recalled that many current systems satisfy the additional condition that, demonstrably, each inconsistency implies an arbitrary formula. With this *additional* condition *it is consistent to assume that every proposition is a formal theorem.* For specialists: some current systems, for example, cut-free ones, do not satisfy this additional condition.

Finally, it is worth noting that the (second) relation above, with the Principle 2(c)(i), makes consistency—for the special systems considered—more than a purely necessary, so to speak negative virtue (i.e., the absence of particularly crass errors like contradictions).

(d) **Discussion:** We have conflicting requirements for a scientifically successful *pursuit* of mathematical logic and for a philosophically adequate *examination* of foundational claims like Hilbert's thesis. For the former it is most rewarding to pursue the details of **(a)** and **(b)**, including the details of $F_T$, thus establishing the potential of formalization. But for a proper assessment of the intended thesis it is essential to realize how few details are needed (for its refutation)!

*Either* the various representations are not available in the system considered when it fails because it lacks expressive power, *or* the system is incomplete and so fails because it lacks deductive power.

This refutation is, perhaps, little more than a wisecrack. If so, the punishment fits the crime, as it were: the thesis is so badly wrong that a refutation is so *un*demanding. This kind of thing is familiar in memorable foundations, and not only in the case of refutations. Thus Tarski gave an impeccable translation *of* " 'snow is white' is true" *by* "snow is white" (used in Appendix 2(c)(i) above). Whatever problems there may be here, they concern the optical properties of snow, not the general notion of truth. This shows, for example, that Pilate had no reason to stay for an answer. In any case, it does not seem to be recorded whether he asked his question with bated breath or a shrug-and-a-wink.

*Reminder:* This use of mathematics, in particular, mathematical logic, is a refrain of the whole review. Being mundane enough to remove any sense of awe inspired by those Big Words, it *corrects* our view of them.

## *Addenda*

(a) Numberings—of words of a numbered alphabet and of other syntactic objects—are used traditionally, and appropriately if some proposition about numbers is to be shown formally independent. But if the rhetoric about set-theoretic foundations were taken literally, one would consider systems for sets, and code (i.e., represent) their syntactic objects by means of hereditarily finite sets, as is done in some elementary texts. For the rhetoric mentioned the "identification" of symbols with sets is a matter of course, and the representation of sequences of sets by sets is familiar. Viewed this way it is quite lopsided to present *arithmetization* as a most central component, let alone novelty, in Gödel's proofs of incompleteness. *NB*. Arithmetization or, more precisely, ingenious variants have *become* central for such delicate later developments as reducing the number of variables in a "universal" diophantine equation; cf. Addendum (c) below.

(b) Actually, Gödel's proofs apply also to suitable sets (of axioms) that are not formal (or, equivalently, recursively enumerable). *Reminder:* the set of consequences should be representable. It is if, for example, all true $\forall$-sentences are added to formal arithmetic.

(c) This is a *representation of a sequence of all representable properties* (of the system considered) in the sense of (a)(ii) in Appendix 2 or, equivalently, an *enumeration* of them for $m = 0, 1, \ldots$. It is perhaps *satisfaisant pour l'esprit* that Kleene called such enumerations, which are indeed central to *in*completeness, *complete* (for recursively enumerable sets); "universal" is more usual now. In the language of functions they correspond to partial recursive enumerations of all partial recursive functions. Here the difference between *total* and *partial* functions corresponds to the difference between truth and provability, mentioned at the end of Appendix 2(a)(i).

(d) Up to this point the notion of numerical representation, explained in Appendix 2(a)(ii) for *properties* or, equivalently, sets, has been perfectly adequate. Specifically, the results stated so far hold for all representations (of the particular set in question), even though the latter need not be formally equivalent. For example, if for each $n$, $A(\bar{n})$ is a formal theorem, $F$ and $F \wedge A$ represent the same set even if $\forall x A(x)$ is not a formal theorem. Similarly, if $\bar{\sigma}'$ represents the substitution operation, *loco citato*, so does $\bar{\sigma}_1'$ where $\bar{\sigma}_1' = \bar{\sigma}' + 0$, but the corresponding $g_1 \neq g$. Both $F_T[\bar{\sigma}'(\bar{g},\bar{g})]$ and $F_T[\bar{\sigma}'(\bar{g}_1,\bar{g}_1)]$ satisfy the condition for so-called *Gödel sentences* $S$: $S \leftrightarrow F_T(\bar{s})$ is a formal theorem, where $s$ is the number of the formula $\neg S$. In (literate) English the popular use of the definite article—"the" Gödel sentence—requires some equivalence relation connecting all formulas that satisfy the condition above. For broad classes of common systems all Gödel sentences are in fact formally equivalent, for others they are not or not known to be equivalent. (Specialists will think here of so-called Rosser systems, and readers of Note 5 will raise similar questions about "this" in "This sentence is false".)

More generally, attention is required not only by the representation of sets, as in Appendix 2(a)(ii), but also of *propositions*. Here, as so often, the best guide for progress comes from broad mathematical experience rather than from the recent logical literature; specifically, from the introduction of (algebraic) coordinates in (synthetic) geometry.[7]

Codes—arithmetic or set-theoretic—correspond to the coordinates; syntactic objects—either thought of as finite words of an enumerated alphabet or as finite trees, say, in print, respectively on a blackboard—correspond to geometric points; finally, properties of the objects—such as terms, formulas, derivations—and relations, say, between derivations and their last formula correspond to geometric relations, for example, collinearity.

At least one (historical) difference is to be noted, especially with respect to the delicate cases of projective or desarguean geometry, which would be called *weak* systems in contemporary logical jargon (compared to the geometry of the full Euclidean plane). The geometric axioms came first, and the corresponding algebraic ideas—of skew fields, fields etc.—afterwards. In contrast, axioms for arithmetic including (weak) subsystems of familiar formal arithmetic came first, while formal systems for syntax—concatenation theory or theories of finite trees—are still not very familiar; especially, the choice of axioms satisfied by the syntactic properties and relations above is usually left implicit. Put differently, it is left open which *data* determine the formal rules considered for generating terms etc. (If Post production rules are meant, the corresponding additional axioms of the concatenation theory have the form of so-called elementary inductive definitions including the principle of proof by induction, which expresses that the least solution of the inductive definition is meant.)

Readers will recall two elements in the introduction of coordinates. First, there is the matter of determining (algebraic) coordinates for points *uniquely up to transformations* of a suitable kind, usually, some transformation group. Secondly, for a given choice of coordinates it is shown that algebraic relations that satisfy the geometric axioms must be of a certain form; for example, a relation satisfying the axioms for collinearity in two dimensions will be linear in the algebraic sense.

*Warning:* There is a difference in jargon. The mathematical tradition does not have the refrain, introduced at the end of Appendix 2(a)(i), about the difference between truth and *provability*, but about truth in, say, the Euclidean plane, and *validity* for all projective or desarguean planes. The definitions involved in the introduction of coordinates are *uniform* for all projective or desarguean planes considered.

The notion, called in the logical literature *canonical representation* (of formal rules), follows the model just recalled; with modifications for *canonical* invariant *definitions*, especially of operations, corresponding to Appendix 2(a)(ii). In particular, given a (canonically defined) coding of the syntactic objects, such canonical representations of syntactic properties are arithmetic predicates satisfying, (demonstrably), the inductive definitions induced by the Post production rules for those properties. Any two canonical representations of a given property are then demonstrably equivalent, provided of course the system of arithmetic considered contains enough induction.

The corresponding *uniqueness condition* satisfied by the (canonical) codings themselves is adequately illustrated by the humble matter of (surjective) *pairing* $\pi$ with its left and right inverses, say $\lambda$ and $\rho$, satisfying $\pi[\lambda(z),\rho(z)] = z$ for all $z$. If $(\pi',\lambda',\rho')$ is another such pairing then the transformation: $z \vdash z'$ is given by: $z' = \pi'[\lambda(z),\rho(z)]$. (The weaker condition on—not necessarily surjective—pairings that used to be quoted, namely, that $\pi(x,y) = \pi(x',y') \rightarrow (x = x' \wedge y = y')$, would of course not be enough for that transformation.) This uniqueness-up-to-definable-isomorphism here corresponds to uniqueness within the geometric transformation group in the case of coordinates. Before going further into the function(s) of canonical representations it is as well to recognize a:

*Wide-spread malaise* about details of coding, often felt to be boring. This is a 'subjective correlative' of the fact that, generally, the results actually proved—about a particular coding or by use of it—are also valid for any other coding that may come to mind. In other words, the details are introduced for some ritual of 'precision', which draws attention away from the more demanding questions of what one is being precise about, and why. At an *elementary* stage there are two principal strategies for progress. One, exemplified by the canonical representations above (and categorical axioms in another sphere), is to formulate some kind of maximal requirement, enough for any developments in sight. At an opposite extreme one looks at minimal requirements for each of the more prominent results. At a *later* stage, brute power may be used as in the 'ingenious variants' alluded to at the end of Addendum (a), and, as always, there is the delicate job of discovering relatively few requirements adequate in relatively many situations. (All this may be hard, but it is not boring.) A good example of the latter is provided by the popular and successful:

*Modal language of provability logic.* It realizes the idea of 'relatively many situations' by its established expressive power. A moment's thought shows that the full panoply of canonical requirements is liable to be excessive here. They—are enough to—cover also propositions about *proofs* (i.e., derivations) while the language is restricted to prov*ability* statements. The latter require only conditions on representing logical operations and the set of consequences, say $F_T$ and $F_{T'}$, which ensure that statements in the modal language are equivalent when $F_T$, respectively $F_{T'}$ are substituted for $\Box$. Because of the iteration of $\Box$, it is not enough that $F_T$ be a representation in the sense of Appendix 2(a)(ii). As experts surely know, but do not publicize, *if modus ponens* or, pedantically, $\Box(p \to q) \to (\Box p \to \Box q)$ is derivable for the representation $F_T$ used, the required condition simplifies. Instead they talk of *modus ponens* being "natural", ignoring both the *discovery* that cut-free rules are useful, and the *philosophical* observation, at the end of Section 4, about the general weakness of the formal picture for understanding the phenomena of proof.

*Bibliographical comment* for readers familiar with the (sloppy) literature on *natural* representations, especially common in connection with consistency statements (in weak systems). The classical literature on the foundations of geometry was far more sophisticated! It did *not assume* that what happened to be familiar from ordinary analytic geometry since Descartes would also be natural for, say, all projective planes. On the contrary it investigated *whether* familiar techniques were adequate in the case of unfamiliar, so to speak nonstandard (projective) planes; 'adequate' for its elegant, though half-forgotten representation theorems with their explicitly stated adequacy conditions. *Remark:* For the record I was not conscious of the close relation between canonical representations and the introduction of coordinates when I introduced the former some twenty-five years ago (though I had learnt the latter from Hilbert's *Foundations of Geometry* twenty years earlier; cf. the footnote on p. 261 of [2]).

(e) This oblique reference to systems that prove their own consistency is worth amplifying, specifically, in terms of canonical representations in the sense of (d). *Reminder:* The "deviant" (self-referential) sentence introduced by Rosser—to eliminate Gödel's condition of $\omega$-consistency for a system $S$, say—is simply a Gödel sentence, in the sense of (d), for the canonical representation of the following system, say, $S_R$. Here

> $d$, with end formula $A$, is a derivation in $S_R$ iff $d$ is a derivation in $S$, and there is no derivation $d'$, before $d$, of $\neg A$.

I do not know whether, in general, $S_R$ proves its own consistency, of course formulated canonically, but $S_R^+$ certainly does, where it is required in addition that

> if $A'$ is the end formula of $d'$, then $A$ is not $\neg A'$ either.

Note that the passage $S \mapsto S_R^+$ is not only a simple but a recursive operation (here meant in the usual, strong sense, not merely the sense of having recursive values at recursive arguments). Incidentally—and without forgetting limitations of *all* formal systems; cf. the end of Section 4 and of (d)—$S_R^+$ *possesses a feature of actual experience with proofs that is not present in* *more usual systems*: the results are cross-checked against background knowledge.

For consistent $S$, $S$ and $S_R^+$ have not only the same set of theorems but the same proofs; only the procedures for checking the latter are different. Thus the—still common—formulation of Gödel's second theorem "for all sufficiently strong systems" is not at all sloppy but just sadly ignorant, in particular of anything like $S_R^+$ above. A civilized formulation (modulo minimal conditions) is in the

**Theorem** *Either consistency or completeness for $\Sigma_0^1$ statements is not (internally) derivable.*

In fact this can be sharpened and expressed in modal language by restricting completeness to provability statements, where not arbitrary Gödel sentences $G$: $\Box(G \leftrightarrow \Box\neg G)$ are used, but only $G$ of the form $\Box\neg G_0$; see the opening paragraph of (d). This is a modal counterpart of so-called literal Gödel sentences, actually constructed in Appendix 2(b).

*Proof* (following Jeroslow): Completeness with respect to provability statements, $(\Box A) \to (\Box\Box A)$, is enough to derive a literal Gödel sentence from consistency, $\neg(\Box B \wedge \Box\neg B)$. Specifically, for $A = \neg G_0$ and $B = \Box A$. For, by the defining property of such Gödel sentences,

$$(\Box\neg G_0) \leftrightarrow \Box\neg\Box\neg G_0;$$

hence,

$$\Box\neg G_0 \to \Box\neg\Box\neg G_0,$$

while completeness for $\Box\neg G_0$ means

$$\Box\neg G_0 \to \Box\Box\neg G_0.$$

**Corollary** *Since the proof does not use closure under modus ponens, and ordinary cut-free systems are complete with respect to $\Sigma_1^0$ statements, such systems do not prove their own consistency either.*

Finally, to repeat what cannot be repeated too often: Without completeness with respect to $\Sigma_1^0$ statements, *consistency is very pale indeed*: it does *not* even ensure the truth of (proved) $\Pi_1^0$ statements; cf. Appendix 2(c)(ii), where $G \in \Sigma_1^0$ and so $(\neg G) \in \Pi_1^0$. *Remark:* As readers will have noticed, above consistency was taken in a sensible form, not merely in Hilbert's coy version, $\neg\Box\bot$, which implies consistency in the presence of *modus ponens*: cf. end of Appendix 2(c). In cut-free systems, $\neg\Box\bot$ *is* provable in the system itself.

## NOTES

1. Even if not the oft-quoted part about settling moral or legal squabbles mechanically. It seems to me it should not be too hard to program expert systems that generate, statistically, more or less the usual judgments. But the market is likely to be limited. Lawyers would not be enthusiastic, for obvious reasons; nor those, among the accused and litigants, who are not satisfied with statistical justice (even when they

accept the two cherished principles of the uniqueness of each individual and of equality before the law, which would make it wise to be satisfied).

2. In contrast to the pioneers a century ago, around the middle of our century Turing proposed a test for 'identifying' thinking behavior by its not being distinguishable from human performance. Naturally, distinctions by reference to certain aspects, so-called results, were meant; comparable in the case of AL (artificial locomotion) to 'identifying' walking with roller skating as long as one starts and finishes together. For AL viewed as a branch of engineering, which lives on achieving given results (i.e., tasks) by novel processes, Turing's proposal is a matter of course. It becomes remarkable if it is viewed as contributing somehow to elucidating the processes in data processing, by putting them into black boxes, as it were, or ignoring them in other ways. Needless to say that 'test' has become popular, among the vulgar, like the most vulgar uses of Ockham's razor (where the fact that something is not needed to explain a particular bunch of phenomena is interpreted to show that it does not exist). For the record, that 'test' is the only indiscretion of Turing that I both know of and have found at all disturbing.

3. For example, by 1929, a couple of years before Gödel's discovery of incompleteness, the number theorists Siegel and Weil discounted, in effect though in different terms, the possibility of a complete formal theory for diophantine equations. In fact, their equations had just two variables, for integral, respectively rational solutions. (Weil's equations were even only of degree four.) The question of whether or not there is a complete formal theory for these special cases is still open. Incidentally, the two papers referred to were immediately famous.

4. The ethereal business of possibilities-in-principle has been most prominent in foundation; not only in the writings of Kant, mentioned already, but also of logicians like Russell, who described *Principia* as "a parenthesis in the refutation of Kant". Here, as at the end of Appendix 2, the punishment fits the crime. Many of Kant's observations on reasoning apply impeccably to actual phenomena but are false if interpreted as needs-in-principle. Thus, contrary to Kant, appeal to geometric experience (especially, visualization, also called *Anschauung*) is not needed in principle for mathematical deductions. But it continues to be used widely, and to good effect. More specifically, Euclidean geometry does not have the privileged place that Kant, taken literally, gave to it (nor, of course, for physical space near massive bodies; and it is not the geometry of visual space either). But, to this day, *mathematicians continue to think in Euclidean terms*, also *when defining non-Euclidean spaces*.

5. At least in my view, Smullyan's *What is the name of this book?* does not belong here at all. (For one thing, the author is well informed.) It contains a remarkable collection of puzzles, puns, and other *jeux d'esprit* in which their logical aspects may fairly be said to be dominant. They are understood by use of propositional or at most (propositional) provability logic. But realistically speaking this recreational corner of experience seems to me to be of very specialized interest; for example, more so than the broader matters in the last chapter of his book, with more jokes, but without much (relation to any earlier) point. True, Smullyan's fancies are no further from ordinary linguistic experience than, say, Galileo's bags of feathers falling behind leaden spheres are from ordinary mechanical phenomena. But they bring at least to my mind a jolly wake for a defunct two-thousand-year-old tradition, that of the Liar, rather than a first step to higher things like celestial mechanics.

6. It will not have escaped the reader's notice that the preceding pages of this review are meant for a broad audience. The same applies to the following two appendices.

However, the latter may be more useful, at least, for specialists if some relations to the technical literature are pointed out *explicitly*. This is done at the end, in the Addenda (a) to (e).

7. Readers familiar with *interpretations* in the sense of Tarski, that is, uniform definitions of a model for one theory in another, may wish to compare the introduction of coordinates to the interpretation of the particular systems of geometry considered in the corresponding algebraic systems; to be quite precise, systems for vectors over the coordinate space. *Reminders:* Projective geometry corresponds to skew fields, Pappus to commutative fields, etc. But over and above an interpretation, coordinates provide an *embedding*, which induces the (definable algebraic) relations that interpret the so-called nonlogical constants of the geometry considered. All this is well known.

   Hilbert went further in his *Foundations*, and interpreted the algebraic systems in the corresponding geometry too. Specifically, he defined ternary relations, say, $A$ and $M$ in the language of projective geometry, and proved for them geometrically the laws of *addition* and *multiplication* that hold in the corresponding algebra, with the same (formulas) $A$ and $M$ for all the extensions of the projective axioms, respectively those for skew fields (and the same embedding, the identity, for the planes, respectively algebras in question). This *tour de force* is a high spot in the tradition to which so-called reverse mathematics belongs.

   In the Addendum (d) coding of syntax is meant as interpreting a suitable formal theory of syntax, that is, of finite words or finite trees with certain inductive definitions – of such syntactic properties as: being terms, proofs, provable – in generally weak systems of arithmetic. The reverse direction, corresponding to Hilbert's *tour de force*, does not seem to have been investigated. Roughly speaking, it looks for an arithmetic 'structure' in – the language of – syntax (even though so far this has no more been needed in coding than the *tour de force* was needed in geometry). Reminder: Strictly speaking, a *relative* interpretation is involved since the embedding is not *onto*, that is, not all natural numbers are codes of some syntactic object.

## REFERENCES

[1] Gödel, K., "Letter to Zermelo (1931)," *Historia Mathematica*, vol. 6 (1979), pp. 294-304.

[2] Hodge, W. V. D. and D. Pedoe, *Methods of Algebraic Geometry*, Vol. I, Cambridge University Press, Cambridge, 1947.

GEORG KREISEL

*Salzburg, Austria*