Philosophia Mathematica Advance Access published January 9, 2006

Philosophia Mathematica (III) Page 1 of 19. doi:10.1093/philmat/nkj003

Are There Absolutely Unsolvable Problems? Gödel's Dichotomy[†]

SOLOMON FEFERMAN*

This is a critical analysis of the first part of Gödel's 1951 Gibbs lecture on certain philosophical consequences of the incompleteness theorems. Gödel's discussion is framed in terms of a distinction between objective mathematics and subjective mathematics, according to which the former consists of the truths of mathematics in an absolute sense, and the latter consists of all humanly demonstrable truths. The question is whether these coincide; if they do, no formal axiomatic system (or Turing machine) can comprehend the mathematizing potentialities of human thought, and, if not, there are absolutely unsolvable mathematical problems of diophantine form.

Either ... the human mind ... infinitely surpasses the powers of any finite machine, or else there exist absolutely unsolvable diophantine problems.

1. Does Subjective Mathematics Coincide with Objective Mathematics?

The above striking dichotomy was enunciated by Kurt Gödel in a lecture entitled 'Some basic theorems on the foundations of mathematics and their implications', on the day after Christmas, 1951, at a meeting of the American Mathematical Society at Brown University in Providence, Rhode Island. The lecture itself was the twenty-fifth in a distinguished series set up by the Society to honor the nineteenth-century American mathematician, Josiah Willard Gibbs, famous for his contributions to both pure and applied mathematics. Soon after Gödel delivered the Gibbs lecture he wrote of his intention to publish it, but he never did so; after he died, the text[1] languished with a number of other important essays and

[†] This is a revised version of a lecture given at the Royal Swedish Academy of Sciences on 22 October 2003 as part of a symposium on my work cited for the 2003 Rolf Schock Prize in Logic and Philosophy. I have benefitted from comments on a draft of this article by Jeremy Avigad, David Israel, Geoffrey Hellman, Wilfrid Hodges, Hannes Leitgeb, Wilfried Sieg, Stewart Shapiro, and Richard Tieszen.

* Departments of Mathematics and Philosophy, Stanford University, Stanford, California 94305, U. S. A. sf@csli.stanford.edu

[1] [Gödel, 1951]. In my discussion of Gödel's lecture I shall in part be enlarging on some of the points made by George Boolos in his introductory note to it on pp. 290–304 of [Gödel, 1995], especially §§1 and 2. Italics in quotes from Gödel are his throughout.

lectures in his *Nachlass* until it was retrieved by our editorial group for publication in Volume III of the Gödel *Collected Works*.

At the risk of going over familiar ground, in this section and the next I will explain what Gödel meant by the various terms in the above statement, and then fill in the ellipses to help us understand better what he was after. The basic results on the foundations of mathematics referred to in the title of his lecture are the two stunning *incompleteness theorems* that he had discovered twenty years earlier. But it was mainly the consequences of the *second incompleteness theorem* that were stressed by Gödel in his lecture, and which he there expressed informally as follows:

> *For any well-defined system of axioms and rules ... the proposition stating their consistency (or rather the equivalent number-theoretical proposition) is undemonstrable from these axioms and rules, provided these axioms and rules are consistent and suffice to derive a certain portion of the finitistic arithmetic of integers.* [Gödel, 1951, pp. 308–309]

As indicated by Gödel in a footnote (p. 308, fn. 10), by axioms and rules sufficient for the 'finitistic arithmetic of integers' he meant the usual system PA of Peano Axioms for the arithmetic of the natural numbers.

The implications that Gödel drew from this incompleteness theorem concerned both the potentialities and possible limitations of human thought as expressed in the above dichotomy in terms of a distinction between *objective mathematics* and *subjective mathematics*. According to him, objective mathematics consists of 'the body of those mathematical propositions which hold in an absolute sense, without any further hypothesis', as contrasted with theorems that are only conditionally true, such as those of axiomatic geometry. A mathematical statement constitutes an *objective problem* if it is a candidate for objective mathematics, *i.e.*, if its truth or falsity is not conditional on any hypotheses and is independent of whether or how it may be demonstrated. At a minimum, among the objective problems are those concerning the arithmetic of integers that are in diophantine form, as explained below.

By subjective mathematics[2] Gödel means the body of all *humanly demonstrable* (or *knowable*) mathematical truths, *i.e.*, all the propositions which the human mind is capable of demonstrating. More precisely, subjective mathematics consists of all those theorems whose truth is demonstrable in *some* well-defined system of axioms all of whose axioms

[2] Subjective mathematics in this sense is no less objective than objective mathematics; so Gödel's terminology may be confusing. I have decided to retain it nevertheless, so as to avoid conflicts with his text.

are recognized to be objective truths and whose rules preserve objective truth.

Of the relation between objective and subjective mathematics, Gödel says that his second incompleteness theorem

> *makes it impossible that someone should set up a certain well-defined system of axioms and rules and consistently make the following assertion about it: All of these axioms and rules I perceive (with mathematical certitude) to be correct, and moreover I believe that they contain all of mathematics.* If someone makes such a statement he contradicts himself. For if he perceives the axioms under consideration to be correct, he also perceives (with the same certainty) that they are consistent. Hence he has a mathematical insight not derivable from his axioms. [Gödel, 1951, p. 309]

Gödel goes on to say that one has to be careful in order to understand clearly the meaning of this state of affairs.

> Does it mean that no well-defined system of correct axioms can contain all of mathematics proper? It does, if by mathematics proper is understood the system of all true mathematical propositions; it does not, however if one understands by it the system of all demonstrable mathematical propositions.

By the above argument, no well-defined system of correct axioms can contain all of objective mathematics. As to subjective mathematics,

> it is not precluded that there should exist a finite rule producing all its evident axioms. However, if such a rule exists, we with our human understanding could certainly never know it to be such, that is, we could never know with mathematical certainty that all the propositions it produces are correct. . . . [W]e could perceive to be true only one proposition after the other, for any finite number of them. The assertion . . . that they are all true could at most be known with empirical certainty.

In these terms, Gödel's dichotomy comes down to whether objective and subjective mathematics coincide. If they do, then demonstrations in subjective mathematics are not confined to any one system of axioms and rules, though each piece of mathematics is justified by *some* such system. If they do not, then there are objective truths that can never be humanly demonstrated, and those constitute absolutely unsolvable problems.

2. Formal Systems, Finite Machines and the Second Incompleteness Theorem

To explain more precisely the assertions in Gödel's dichotomy and the incompleteness theorem behind it, I shall elaborate his terminology according to his own explanations. To begin with, a system of axioms and rules of inference is said to be *well-defined* if it is a formal system in the usual sense, *i.e.*, if, first of all, its propositions are stated in a formal language all of whose expressions are finite sequences of a specified finite stock of basic symbols, and if, secondly, one can check by a *finite procedure* (on a *finite machine*) the following three things: (i) whether or not a given sequence of basic symbols constitutes a statement of the language, (ii) whether or not a given statement is an axiom, and, finally, (iii) whether or not a given statement follows directly from other statements by one of the rules of inference.[3]

As Gödel explains at the beginning of his lecture, by a finite machine he means a Turing machine, and by a finite procedure he means one that can be carried out by such a machine. Furthermore, he is in agreement with the Church-Turing thesis that identifies effective computability with computability on a Turing machine. In accordance with more current terminology, I shall occasionally use *effective procedure* in place of 'finite procedure', and *effectively given (or presented) formal system* in place of 'well-defined system of axioms and rules'. I shall also use the letter 'S' to denote any such system. Standard examples are Peano Arithmetic (PA) and Zermelo-Fraenkel set theory (ZF). In Gödel's view, both PA and ZF are part of objective mathematics, but he is at pains to formulate the incompleteness theorem and its consequences in a way that holds for other positions—such as those of the finitists and intuitionists—as to what counts as mathematics proper.

The proof of Gödel's incompleteness theorems for sufficiently strong formal systems S makes use of the arithmetization of syntax of S, via the attachment of (Gödel) numbers to the symbols of S, then the formulas of S, and finally the sequences of formulas of S. We can effectively determine whether or not a given number is the Gödel number of a formula of S. Proofs from the axioms of S are finite sequences of formulas, each of which is either an axiom of S or is obtained from earlier formulas in the

[3] In one of Hao Wang's conversations with Gödel in 1972, reported in [Wang, 1996, p. 204], Gödel says that 'a formal system is nothing but a many-valued [non-deterministic] Turing machine which permits a predetermined range of choices at each stage'.

sequence by application of one of the rules of inference. Again, we can effectively determine whether or not a given number is the Gödel number of a proof from S. Furthermore, when it is, we can effectively extract the number of the final statement in the proof, *i.e.*, the theorem established by the proof. It follows that we can mechanically run through all the integers in turn, testing each to see if it is the number of a proof from S, and when it is, give as output the number of the theorem established by that proof. In other words, with each effectively given formal system is associated a Turing machine M which enumerates the set of theorems of S, or—more picturesquely—prints out the theorems of S one after another. Conversely, given any formal language L, any Turing machine M can be made to correspond to a formal system S in L by extracting from the numbers it enumerates those that are Gödel numbers of formulas of L, and taking their deductive closure to be the theorems of S. In this way, talk of well-defined or effectively given formal systems can be converted into talk of Turing machines and vice versa.

Given any effectively presented formal system S we may construct in a canonical way from its presentation an effectively computable relation $\mathrm{Prf}_S(x, y)$ which expresses that x is the Gödel number of a proof in S of the formula with Gödel number y. Fixing y to be the number n_0 of some standard false statement such as $\neg(0 = 0)$, the statement Con_S defined as $\forall x\ \neg\mathrm{Prf}_S(x, n_0)$ then expresses that S is consistent. In his original proof of the incompleteness theorems, Gödel showed that for the case that the relation Prf_S is primitive recursive, it is definable in the language of Peano Arithmetic. Later it was recognized through the work of Kleene that statements of the form $\forall x\ \mathrm{P}(x)$ with P Turing computable can be effectively re-expressed in the form $\forall x\ \mathrm{R}(x)$ with R primitive recursive, hence definable in the language of arithmetic. Thus for any effectively presented formal system S, the statement Con_S may be construed to be a statement of that language. It is with this understanding that Gödel's second incompleteness theorem is expressed as follows.

> *If* S *is an effectively given formal system which contains PA and* S *is consistent, then* Con_S *is not provable in* S.

Much stronger formulations of this theorem have since been established, obtained by replacing PA by a relatively weak subsystem proof-theoretically equivalent to the system PRA of Primitive Recursive Arithmetic.[4] However, this and even stronger improvements do not affect Gödel's general argument for his dichotomy.

[4] See, *e.g.*, [Hájek and Pudlak, 1991, p. 164].

Note well that whether or not the axioms of S are objectively true and whether or not the rules of inference of S preserve objective truth, the consistency statement for S is an objective number-theoretical problem: either the system is consistent or it is not, so either Con_S is true or it is not. Gödel refers to it as a *diophantine problem.* The appellation comes from the work *Arithmetica* of the third-century AD Greek mathematician, Diophantus of Alexandria, on the solutions in integers of polynomial equations with integer coefficients. In an unpublished and undated manuscript from the 1930s found in Gödel's *Nachlass* and reproduced in Vol. III of the *Collected Works*,[5] he showed that every statement of the form $\forall x\, \mathrm{R}(x)$ with R primitive recursive is equivalent to one in the form

$$\forall x_1 \ldots x_n \exists y_1 \ldots y_m [p(x_1, \ldots, x_n, y_1, \ldots, y_m) = 0]$$

in which the variables range over natural numbers and p is a polynomial with integer coefficients; it is such problems that Gödel referred to as diophantine in the Gibbs lecture.[6] It follows from the later work on Hilbert's 10^{th} problem by Martin Davis, Hilary Putnam, Julia Robinson and—in the end—Yuri Matiyasevich that, even better, one can take $m = 0$ in such a representation when the $=$ relation is replaced by $\neq$.[7] Thus, in the following, one may take diophantine problems to be those in the latter stronger form rather than in Gödel's form.

3. Gödel's Dichotomy and Its Proof

We are now in a position to return to Gödel's dichotomy and state it in full as done in his Gibbs lecture:

> *Either mathematics is incompletable in this sense, that its evident axioms can never be comprised in a finite rule, that is to say, the human mind (even within the realm of pure mathematics) infinitely surpasses the powers of any finite machine, or else there exist absolutely unsolvable diophantine problems of the type specified.* [Gödel, 1951, p. 310]

According to Gödel, this is a 'mathematically established fact' which is a consequence of his incompleteness theorem. However, all that he says by way of an argument for it, in consequence of the second incompleteness theorem, is the following:

[5] Appearing there as *Gödel *193?*.
[6] [Gödel, 1951, p. 307].
[7] See, *e.g.*, [Davis, Matiyasevich, and Robinson, 1976].

> [I]f the human mind were equivalent to a finite machine, then objective mathematics not only would be incompletable in the sense of not being contained in any well-defined axiomatic system, but moreover there would exist *absolutely* unsolvable problems . . . , where the epithet 'absolutely' means that they would be undecidable, not just within some particular axiomatic system, but by *any* mathematical proof the mind can conceive. [Gödel, 1951, p. 310]

Looking more closely at this, there are some unstated underlying assumptions that I make explicit as follows:

(I) The human mind, in demonstrating mathematical truths, only makes use of evidently true axioms and evidently truth-preserving rules of inference at each stage.
(II) The axioms and rules of inference accepted as evident for the human mind include those encapsulated in the system PA of Peano Arithmetic.
(III) By a finite machine is meant a Turing machine which enumerates only theorems that are among those provable by the human mind.

The argument then runs as follows. To say that the human mind—in its capacity as a producer of mathematical truths—is equivalent to a finite machine amounts, as we have seen, to the same thing as saying that the set of humanly demonstrable theorems can be axiomatized by an effectively given formal system S. Suppose this is the case; then since by (I) the human mind proves only truths, S must be consistent. Moreover, by (II) S contains PA. Hence by the second incompleteness theorem, the consistency Con_S of S is true but not provable in S; so it is not humanly provable. Since the negation of Con_S is not true it is also not provable in S by (I). But Con_S can be brought to the form of a diophantine statement, and so if the human mind is equivalent to a finite machine, it provides an example of an absolutely undecidable diophantine problem. On the other hand, if the human mind is not equivalent to a finite machine, then for any such machine the mind can prove a statement which cannot be produced by a machine, or, as Gödel puts it, 'the human mind . . . infinitely surpasses the powers of any finite machine'.

4. Skeptical Considerations

What are we to make of this argument? Gödel's claim that its conclusion is a mathematically established fact can only be accepted if we grant mathematical meaning to all the notions involved and mathematical certitude to the assumptions. But we have problems with that straight off: hardly any mathematicians would ascribe mathematical clarity to the concept of 'the human mind' or even of what is humanly demonstrable within mathematics, or even more specifically of what is humanly demonstrable within the realm of arithmetic. Moreover, since what is at issue is the producibility of an infinite set of propositions, no one mathematician, whose life is finitely limited, can produce such a list; so either what one is talking about is what the individual mathematician *could do in principle*, or one is talking in some sense about the potentialities of the pooled efforts of the community of mathematicians now or ever to exist. Again, we ought to regard that as a matter of what can be done *in principle*, since it is most likely that the human race will eventually be wiped out either by natural causes or through its own self-destructive tendencies by the time the sun ceases to support life on earth. One might hope that somehow or other, despite these dire probabilities, humanity, or whatever it evolves into, will persist into the indefinite future somewhere in the universe and continue to produce new mathematical theorems. But it is clear that whichever way this is taken and whatever speculations about the future are accepted, there is a *highly idealized* concept of the human mind in its mathematizing behavior that is at play in Gödel's dichotomy.[8]

This is not to say that the perennial question: *Is mind mechanical?*, is not taken seriously. One has only to look at the contents of such works as: *Minds and Machines*; *How the Mind Works; Minds, Brains, and Computers; Mind Design; etc., etc.*, to see that this has engaged the attention of a number of researchers in cognitive science, especially from the fields of philosophy, psychology, and artificial intelligence, and has even attracted the interest of some mathematicians willing to forego mathematical precision. The question addressed by these thinkers is whether human mental activity can in general be described in terms of the workings of a machine. But in pursuing that question, even the concept of machine is up for grabs. Some cognitive scientists have proposed mechanical models quite different from those provided by Turing machines. For example, there are the so-called *connectionist machines* (or *neural networks*, or *parallel distributed processors*, as they are also called); unlike the idea of Turing machines, there is no established mathematical explanation of what constitutes a connectionist machine,

[8] The problematic character of these idealizations has been stressed by Stewart Shapiro [1998], among others. An indication of Gödel's own view is provided by a conversation with Hao Wang reported in [Wang, 1996, p. 189 (6.1.23)]: 'By *mind* I mean an individual mind of unlimited life span. This is still different from the collective mind of the species.' *Cf.* also [Wang, 1996, p. 205 (6.5.5)].

but there is a general idea of such, given specificity by a number of interesting examples.[9]

Setting aside the problem of minds and mental capacities, another concept that is presumed in Gödel's dichotomy and that may be viewed askance is that of objective mathematical truth, which is supposed to be considered in an absolute sense, independent of how one may arrive at such. Even in the most basic realm of arithmetic, there is some dispute among mathematicians of various foundational tendencies about this idea. The formalists identify truth—if they are willing to speak about it at all—with what can be proved in a formal axiomatic system. The intuitionists identify truth with what can be demonstrated by constructive means, though what constitutes a constructive proof is to be understood without reference to a formal system. And the finitists limit truth to statements of a very restricted kind that can be verified without appeal to the completed infinite. Unlike any of these, I think it is fair to say that the great majority of working mathematicians view what they are after is determining truths, and that the questions of truth or falsity of statements of arithmetic, and in particular those in diophantine form, are definite mathematical problems. Gödel himself counts the statements of set theory among the objective problems, but, as I have said, he formulates his dichotomy in such a way that there can be as divergent understanding of what those come to as that given by the finitists and the intuitionists. In doing so, he makes the statement of the dichotomy even vaguer.

Moving on, what about Gödel's talk of 'evident axioms'? This reminds me of the apocryphal story told about Norbert Wiener who at a certain point in his lecture on some recondite points in the theory of Fourier series asserted that something is obvious, stopped, reconsidered, went out of the room for a half hour, and then returned saying, 'Yes, it's obvious.' What is left out of the story is that what is obvious to such a one as Wiener need not be obvious to the students, no matter how hard they try to grasp what is asserted, even if it is something that is supposed to be 'really obvious'. And that already applies in the case of so-called evident axioms that take a certain amount of mathematical sophistication to appreciate. Even Gödel

[9] See, for example, [Rumelhart and McLelland, 1986] and [Churchland and Sejnowski, 1992]. Any connectionist machine with rational weights and computable activation functions at each unit can be treated in terms of Turing machines. But the essential feature of experiments with connectionist machines to model assorted cognitive phenomena is the use of procedures for 'training' them so as to decrease the error between correct outputs and actual outputs over a sample of inputs, by systematically modifying the weights. There are various procedures in practice for minimization of errors but no general theory of such.

suggested, in his famous article on Cantor's continuum problem, that what is evident can be cultured:

> [T]here may exist, besides the usual axioms ... other (hitherto unknown) axioms of set theory which a more profound understanding of the concepts underlying logic and mathematics would enable us to recognize as implied by these concepts. [Gödel, 1990, p. 261]

As suggested in this quote, it is not at all evident what makes a proposed axiom evident.

Let us take this one step farther: recall assumption (I) that the human mind, in doing mathematics, makes use only of evident axioms and evident rules of inference. But mathematicians hardly mention axioms at all in support of their proofs in their daily practice, and some go through their entire careers without appealing once to an axiom of *any* kind. So assumption (I) requires an argument that underneath it all, mathematicians proceed step by step from what is already known, and so on, back to more and more basic knowledge. According to this view, there can be no infinite regress in this; so an account of the logical structure of mathematical practice must eventually reach statements which are regarded as so evident that they take on the status of axioms. An alternative view of mathematics is promoted by such thinkers as Imre Lakatos who—inspired by Karl Popper—assimilated mathematics to empirical science and other areas of fallible knowledge. Lakatos pointed both to the history of the subject, which is full of controversy, confusion, and even error, marked by periodic reassessments and occasional upheavals, and to the mathematician at work, who relies on surprisingly vague intuitions and proceeds by fumbling fits and starts with all too frequent reversals. I have argued against the Lakatosian view of mathematics and in favor of an account in terms of its logical structure,[10] but want to emphasize that the latter cannot be taken for granted. Even if it is accepted that mathematics proceeds at bottom in a logical way, one may ask whether the formalization of mathematics in effectively given axiomatic systems S of the sort presumed by Gödel provides the appropriate model of the logical structure of mathematics. In this respect I have developed views which diverge from the metamathematical approach that is currently in the saddle, but it would take me too far afield and into more technical territory to explain that here.[11]

[10] *Cf.* [Feferman, 1998, Ch. 3].
[11] *Cf.* [Feferman, 1996].

5. Taking the Issues at Face Value

What I have been concerned with in the preceding section is to put into question Gödel's assertion that his dichotomy is an established mathematical fact. But at an informal, non-mathematical, more every-day level, there is nevertheless something to the ideas involved and something to the argument that we can and should take seriously. If we then take these at face value, we will want to say more, namely, *which* disjunct ought to be accepted and *why*? There was no uncertainty about the choice in the mind of David Hilbert. In his famous lecture entitled 'Mathematical problems' for the meeting of the International Congress of Mathematicians held in Paris in 1900, Hilbert emphasized the importance of taking on challenging problems for maintaining the progress and vitality of mathematics. But he went beyond that to express a remarkable conviction in the solvability of all mathematical problems, which he even called an 'axiom'. To quote from his lecture:

> Is the axiom of solvability of every problem a peculiar characteristic of mathematical thought alone, or is it possibly a general law inherent in the nature of the mind, that all questions which it asks must be answerable? ... This conviction of the solvability of every mathematical problem is a powerful incentive to the worker. We hear within us the perpetual call: There is the problem. Seek its solution. You can find it by pure reason, for in mathematics there is no *ignorabimus*.[12]

To be sure, one problem after another had been vanquished in the past by mathematicians, though sometimes only after considerable effort and only over a period of many years. And Hilbert's personal experience was that he could eventually solve any problem he set his mind to. But it was rather daring to assert that there are *no* limits to the power of human thought, at least in mathematics.

At any rate, Hilbert clearly said that there are no absolutely unsolvable problems, but in answer to the question as to why he believed that, he told

[12] [Hilbert, 1900], as translated in [Browder, 1976, p. 7]. For the relevance of Hilbert's problems to logic see the articles there by A. D. Martin and G. Kreisel, as well as Ch. 1 of [Feferman, 1998]. Hilbert qualified his idea of the solvability of every problem by saying that a solution might consist of demonstrating that the problem cannot be solved by prescribed methods, *e.g.*, the problem of the duplication of the cube by straightedge and compass and the problem of the solution of the general fifth-degree equation by radicals. It should also be noted that some of the problems on Hilbert's list are not at all specific but, rather, are programmatic. For example, Problem 6 calls for one to 'treat ... by means of axioms, those physical sciences in which mathematics plays an important part.'

us only that it is an axiom or a matter of conviction.[13] What was Gödel's own view? Typically cautious, in the Gibbs lecture he stated his conclusion from the second incompleteness theorem only as a disjunction, despite his personal conviction that mind is not equivalent to a finite machine. Apparently the reason he did that is because he did not feel he had a knock-down proof of the falsity of the mechanist position. Rather, he put forward various arguments against that position, including several communicated to Hao Wang and first recounted in the latter's book, *From Mathematics to Philosophy*[14] and then at greater length in *A Logical Journey. From Gödel to Philosophy*.[15] Gödel thought that Hilbert was right to reject the possibility of absolutely unsolvable problems. Otherwise, 'it would mean that human reason is utterly irrational by asking questions it cannot answer, while asserting emphatically that only reason can answer them.'[16] Further, in a note entitled 'A philosophical error in Turing's work' prepared for publication but never actually published, Gödel wrote:

> Turing gives an argument which is supposed to show that mental procedures cannot go beyond mechanical procedures. However, this argument is inconclusive. What Turing disregards completely is the fact that *mind, in its use, is not static, but constantly developing, i.e.*, we understand abstract terms more and more precisely as we go on using them ... though at each stage the number and precision of the abstract terms at our disposal may be *finite*, both ... may *converge toward infinity* ....[17]

Note that, in contrast to the ascription of an error in the title, Gödel does not say that Turing is *mistaken*, only that his argument is *inconclusive*. Moreover, in the Gibbs lecture he countenances the possibility that 'the

[13] For an interesting stab at a theoretical formulation of Hilbertian 'rationalistic optimism' see [Shapiro, 1997, pp. 207–211].

[14] [Wang, 1974, pp. 324–326].

[15] [Wang, 1996], especially Ch. 6. Cf. also [Wang, 1993].

[16] [Wang, 1974, p. 324]. Geoffrey Hellman (personal communication) has pointed out that this is hyperbole (uncharacteristic of Gödel), for 'first of all, in any particular case, we may not know or have reason to think that we're dealing with an unsolvable problem, so why is it irrational to pursue a proof? And second, to assert that *only* reason can answer is by no means that reason *can* answer—a problem may simply be beyond "reason's" capacity to answer (which is not to say it has no objectively correct answer, a separate issue).'

[17] [Gödel, 1990, *1972a* Remark 3, p. 306]. A close version of same is to be found in [Wang, 1974, p. 325]. *Cf.* also the relevant conversations reported in [Wang, 1996, pp. 195–200].

human mind (in the realm of pure mathematics) *is* equivalent to a finite machine that, however, is unable to understand completely its own functioning'. And in a footnote he says that

> [I]t is conceivable ... that brain physiology would advance so far that it would be known with empirical certainty
>
> 1. that the brain suffices for the explanation of all mental phenomena and is a machine in the sense of Turing;
>
> 2. that such and such is the precise anatomical structure and physiological functioning of the part of the brain which performs mathematical thinking. [Gödel, 1951, p. 309, fn. 13]

Gödel's cautious statement concerning minds and machines is also curious in view of his assertion near the outset of the Gibbs lecture that the 'phenomenon of the inexhaustibility of mathematics' follows from the fact that

> [T]he very formulation of the axioms [of set theory over the natural numbers] up to a certain stage gives rise to the next axiom. It is true that in the mathematics of today the higher levels of this hierarchy are practically never used. It is safe to say that 99.9% of present-day mathematics is contained in the first three levels of this hierarchy. So for all practical purposes, all of mathematics *can* be reduced to a finite number of axioms. However, this is a mere historical accident, which is of no importance for questions of principle. [Gödel, 1951, p. 307]

More positive claims to prove that mind is not equivalent to a finite machine in Turing's sense have come from several sources, most prominently the philosopher J. R. Lucas and the mathematician Roger Penrose. In Lucas's article 'Minds, machines and Gödel' [Lucas, 1961], he argued that whatever Turing machine M is proposed to describe mind, knowing the program for M one can produce a sentence which is true but not provable by the machine, namely the consistency statement for M. This presumes that if mind is equivalent to a Turing machine then its exact program can be produced; it may well be that empirical investigations support the existence of such a machine without one being able to tie it down in complete detail. The second thing that Lucas's argument presumes is that mind is consistent, and moreover that we *know* it is consistent. In his book *Shadows of the Mind*, Penrose developed a refined version of Lucas's argument designed to be immune from such criticisms.

But his conclusion is concomitantly weak: ‘Human mathematicians are not using a knowably sound algorithm in order to ascertain mathematical truth’.[18] In a trenchant survey of these and other such ‘proofs’, Stewart Shapiro has concluded that ‘*there is no plausible mechanist thesis on offer that is sufficiently precise to be undermined by the incompleteness theorems*’.[19]

While I agree completely with Shapiro, this leaves open the possibility that there are grounds, other than those coming from the incompleteness theorems, for coming to the conclusion that there are no absolutely unsolvable problems. Indeed, Per Martin-Löf has proved exactly that, in the form: *There are no propositions which can neither be known to be true nor be known to be false* [Martin-Löf, 1995, p. 195]. However, this is established on the basis of the constructive explanation of the notions of ‘proposition’, ‘true’, ‘false’, and ‘can be known’. The argument goes roughly as follows. A proposition A can be known to be true just in case it can be demonstrated and it can be known to be false just in case its negation ¬A can be demonstrated, *i.e.*, A can be shown to lead to a contradiction. Then to assert that A cannot be known to be true implies ¬A. Similarly, to assert that A cannot be known to be false implies ¬¬A. Thus the assumption that one has a proposition A which both cannot be known to be true and cannot be known to be false is to have one for which ¬A∧¬¬A holds, and that is demonstrably false.[20]

For the non-constructive mathematician, Martin-Löf’s result would be translated roughly as: *No propositions can be produced of which it can be shown that they can neither be proved constructively nor disproved constructively*. For the non-constructivist this would seem to leave open the possibility that there are absolutely unsolvable problems A ‘out there’, but we cannot *produce* ones of which we can *show* that they are unsolvable.

[18] [Penrose, 1994, p. 76]. I am deliberately ignoring here Penrose’s ‘new argument’ in the 1996 *Psyche* symposium on his book; *cf.* [Shapiro, 1998, p. 284] and [Lindström, 2001].

[19] [Shapiro, 1998, p. 275]. For useful supplements to Shapiro’s bibliography see also [Hellman, 1981], [Lindström, 2001], [Franzén, 2005], and http://cons.net/online2.html#godel in David Chalmers’s excellent collection of online sources on the philosophy of mind.

[20] Gödel makes a similar observation in [1951, p. 310, fn. 15], as follows: ‘It is to be noted that intuitionists have always asserted the first term of the disjunction [in the dichotomy] (and negated the second term, in the sense that no demonstrably undecidable propositions can exist).’

6. Minds Are Not Finite Machines and Yet There Are Absolutely Unsolvable Problems

Where does this leave us? First of all, though I have used 'Gödel's dichotomy' to refer to the disjunctive formulation in section 3 above, the disjuncts are not on the face of them mutually contradictory. Gödel himself asserts in a parenthetical remark directly following the statement that 'the case that both terms of the disjunction are true is not excluded, so that there are, strictly speaking, three alternatives'. Speaking for myself, I have to say that I find it plausible that *both* disjuncts of Gödel's dichotomy are true! But here I depart from problematic matters of principle to matters of current and foreseeable practice.

Let me begin with the first disjunct. It seems to me that there is an equivocation in Gödel's 'proof' of the dichotomy and in the Lucas-Penrose arguments, in which one countenances the assumption that mind is a finite machine with respect to the production of mathematical truths. There are two ways of taking this, if one grants that such truths are established only by proofs. The first is that in carrying out these proofs human minds are (explicitly or implicitly) following some one formal system S of evident axioms and rules of inference. The second is that human minds are employing an algorithm or program for a Turing machine to carry out the production of mathematical truths.[21] The equivocation lies in identifying the *processes* for producing proofs of theorems of S with their *results*, that is, with the set of all theorems of S. If one considers *only* the results then for mind to be constrained to follow a single S is the same as to be a finite machine since, under the given assumptions on formal systems, the set of theorems of S is effectively enumerable. But if we look instead at the *processes* by means of which they are obtained, it is obvious that the *way* mathematicians prove theorems is not at all the way that machines (at least as currently conceived) churn out theorems. It is a travesty of the former to picture the mathematician trying to prove a particular statement A by enumerating the theorems of S one after another to see if A is among them. By contrast, the actual human search process is guided by a combination of experience, intuition, and trial-and-error; frequently it requires intense concentrated work, intermixed with periods of gestation and sudden leaps of realization ('Ah, hah!').[22] The elements of a solution to a non-routine problem are usually novel and can even be suprising; they may require the introduction of new connections and new concepts that are nowhere on the surface.

[21] The two ways are brought closer together by allowing consideration of non-deterministic Turing machines; *cf.* fn. 3 above.

[22] *Cf.* [Hadamard, 1949], [Pólya, 1968], [Fischbein, 1987], and [Feferman, 2000].

Moreover, finding proofs of theorems that have previously been stated is only part of the story. A good part of the mathematician's work in practice is devoted to arriving at new fruitful conjectures to be tackled, and that is a matter of informed judgment as well as an exercise of creative intelligence. There is nothing at all in the machine picture that accounts for *this* aspect of the mathematical mind at work.

Note that my conclusion that—insofar as the process of producing theorems is concerned—mind is not a finite machine in the sense of Turing, does not exclude the possibility that the actual creative mathematical process can be modeled by some other kind of machine; however, there is no proposal for such remotely on the horizon.

Note also that it is not excluded that the totality of humanly acceptable principles for proving theorems is bounded by a single formal system S, for example the system ZFC with the addition of all large-cardinal axioms that have been considered to date.[23] But even though such axioms border on inconsistency, as they do at the very highest reaches,[24] it is of course conceivable that one will think of new axioms beyond those considered to date for which consistency can still be entertained. On the other hand, I don't know of anyone who says that we can be assured that all the large-cardinal axioms that have been considered to date lead only to mathematical truths, let alone that they are 'evident' as required by Gödel in his disjunctive formulation.[25]

Let us turn, finally, to the second disjunct. We do not have any precise criterion for the solvability of individual problems which would allow us to prove the existence of problems that are absolutely unsolvable *in principle*; so it is idle to ask for examples of such. Instead I wish to point to two problems at the extremes of current mathematics that I will argue are absolutely unsolvable from the standpoint of *practice*.

(P1) Is the value of the digit in the $10^{10^{10^{10}}}$ th place of the decimal expansion of $\pi - 3$ equal to 0?

As Robert Solovay and I have written about such problems,[26] this is an example of a mathematical 'yes/no' question, whose answer can be determined in principle by a mechanical check, but which, in all probability, cannot be settled by the human mind because it is beyond

[23] *Cf.* [Kanamori, 1994, p. 471].

[24] *Ibid.*, p. xxii.

[25] If one is only concerned with the completely minimal requirement that every diophantine statement proved be true, one need only seek assurance as to the consistency of such axioms. But what could provide such assurance, given Gödel's second incompleteness theorem?

[26] Feferman and Solovay, Introductory Note to *Gödel 1972a*, Remark 2, in [Gödel, 1990, p. 292].

all remotely conceivable computational power on the one hand and there is no conceptual foothold to settle it by a proof on the other.

The second problem that is a candidate to be absolutely unsolvable is Cantor's continuum problem, which Hilbert placed first on his list of 23 open mathematical problems in his 1900 address.

(P2) Is there an uncountable subset of the set **R** of real numbers which is not in one-to-one correspondence with **R**?

Gödel took this problem as belonging to the realm of objective mathematics and thought that we would eventually arrive at evident axioms to settle it. In terms of Cantor's notation, we have that the cardinal number of **R**, Card(**R**), is greater than or equal to the first uncountable cardinal number $\aleph_1$; Cantor conjectured that they are equal. But all efforts to determine the value of the cardinal number of **R** ('the continuum') on the basis of currently accepted axioms or any plausible extension S of those axioms proposed thus far have failed. Using the forcing method introduced by Paul Cohen, Azriel Levy and Robert Solovay have shown that both $\mathrm{Card}(\mathbf{R}) = \aleph_1$ and $\mathrm{Card}(\mathbf{R}) > \aleph_1$ are consistent with any such S, provided S is consistent.[27] [28] Of course, Cantor's problem is not a diophantine problem, and if it is absolutely unsolvable that does not mean that there exist absolutely unsolvable diophantine problems. But there are closely related problems which are of that form. Namely, in an effort to settle the continuum problem, some set theorists have proposed adding to the axioms of set theory statements of the form that some very large cardinals exist. Well, then one can ask whether the resulting system is consistent, and *that* question *is* a diophantine problem. How could we hope to settle it? As remarked in fn. 25, Gödel's incompleteness theorem makes that a non-starter.

My conclusion from all this is that—stimulating as Gödel's formulations in the Gibbs lecture of the consequences for mathematics of his incompleteness theorems are—they are a long way from establishing anything definitive about minds, in their general mathematizing capacity, and machines. Gödel himself was more concerned with the philosophical consequences, namely that 'under either alternative [of the disjunction] they are very decidedly opposed to materialistic philosophy', or, to put it more positively, they support 'some form or other of Platonism or "realism" as to the mathematical objects' [1951, p. 311], and he devoted

[27] *Cf.* [Martin, 1976, p. 86]. More precisely, the 'plausible extensions' S of ZFC are those by some large-cardinal axiom asserting the existence of a cardinal κ which is preserved under Cohen extensions relative to forcing by P with card(P) smaller than κ. This result was proved in [Levy and Solovay, 1967].

[28] Conceivably, Hilbert would have regarded the proof of independence of CH from ZFC set theory as a 'solution' to P2. *Cf.* fn. 12.

the main part of his lecture to making the case for that. A discussion of these claims could be largely independent of the preceding considerations.

Note Added in Proof

Alasdair Urquhart brought to my attention interesting notes that Emil Post made in November, 1940, of two conversations he had had a year or two before with Kurt Gödel on the subject of absolutely unsolvable problems; these notes are reproduced in part in [Grattan-Guinness, 1990, pp. 82–83]. At that time, Gödel offered the Continuum Hypothesis as a possible candidate for such a problem if, as he conjectured, its independence from the axioms of set theory could be established. Seemingly, the idea was that independence of CH, like its consistency, would hold for any extension of the axioms of set theory that might reasonably be contemplated. As is well known, Gödel later rejected this possibility in his 1947 article on Cantor's continuum problem [Gödel, 1990, pp. 176–187]. For Post's own speculations about absolutely unsolvable problems, see [Urquhart, forthcoming].

References

BROWDER, FELIX, ed. [1976]: *Mathematical Developments Arising From Hilbert Problems*. Proceedings of Symposia in Pure Mathematics, Vol. 28. Providence, R. I.: American Mathematical Society.

CHURCHLAND, PATRICIA S., and TERRENCE J. SEJNOWSKI [1992]: *The Computational Brain*. Cambridge, Mass.: Bradford/MIT Press.

DAVIS, MARTIN, YURI MATIYASEVICH, and JULIA ROBINSON [1976]: 'Hilbert's tenth problem. Diophantine equations: Positive aspects of a negative solution', in [Browder, 1976], pp. 323–378.

FEFERMAN, SOLOMON [1996]: 'Gödel's program for new axioms: Why, where, how and what?', in P. Hájek, ed., *Gödel '96: Logical Foundations of mathematics, Computer Science, and Physics—Kurt Gödel's Legacy*. Lecture Notes in Logic 6. Berlin: Springer-Verlag, pp. 3–22.

—— [1998]: *In the Light of Logic*. New York: Oxford University Press.

—— [2000]: 'Mathematical intuition vs. mathematical monsters', *Synthese* **125**, 317–332.

FISCHBEIN, EFRAIM [1987]: *Intuition in Science and Mathematics*. Dordrecht: Reidel.

FRANZÉN, TORKEL [2005]: *The Incompleteness Theorem*. Wellesley, Mass.: AK Peters.

GÖDEL, KURT [1951]: 'Some basic theorems on the foundations of mathematics and their implications', in [Gödel, 1995], pp. 304–323.

—— [1990]: *Collected Works, Vol. II. Publications 1938-1974*. S. Feferman, *et al*., eds. New York: Oxford University Press.

—— [1995]: *Collected Works, Vol. III. Unpublished Essays and Lectures*. S. Feferman, *et al*., eds. New York: Oxford University Press.

GRATTAN-GUINNESS, IVOR [1990]: 'The manuscripts of Emil L. Post', *History and Philosophy of Logic* **11**, 77–83.

HADAMARD, JACQUES [1949]: *The Psychology of Invention in the Mathematical Field*. Princeton, N. J.: Princeton University Press.

HÁJEK, PETR, and PAVEL PUDLAK [1991]: *Metamathematics of First-order Arithmetic*. Berlin: Springer-Verlag.

HELLMAN, GEOFFREY [1981]: 'How to Gödel a Frege-Russell: Gödel's incompleteness theorem and logicism', *Noûs* **15**, 451–468.

HILBERT, DAVID [1900]: 'Mathematische Probleme. Vortrag, gehalten auf dem internationalen Mathematiker-Kongress zu Paris, 1900', *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen*, pp. 253–297; translated into English in [Browder, 1976], pp. 1–34.

KANAMORI, AKI [1994]: *The Higher Infinite*. Berlin: Springer-Verlag.

LEVY, AZRIEL, and ROBERT SOLOVAY [1967]: 'Measurable cardinals and the Continuum Hypothesis', *Israel Journal of Mathematics* **5**, 234–248.

LINDSTRÖM, PER [2001]: 'Penrose's new argument', *J. Philosophical Logic* **30**, 241–250.

LUCAS, J. R. [1961]: 'Minds, machines, and Gödel', *Philosophy* **36**, 112–137.

MARTIN, DONALD A. [1976]: 'Hilbert's first problem: The Continuum Hypothesis', in [Browder, 1976], pp. 81–92.

MARTIN-LÖF, PER [1995]: 'Verificationism then and now', in W. De Pauli-Schimanovich, *et al.*, eds. *The Foundational Debate*. Dordrecht: Kluwer, pp. 187–196.

PENROSE, ROGER [1994]: *Shadows of the Mind*. Oxford: Oxford University Press.

PÓLYA, GEORGE [1968]: *Mathematics and Plausible Reasoning*, Vols. 1 and 2. 2nd ed. Princeton, N. J.: Princeton University Press.

RUMELHART, DAVID E., and JAMES L. MCLELLAND, eds. [1986]: *Parallel Distributed Processing*. Vols. 1 and 2. Cambridge, Mass.: Bradford/MIT Press.

SHAPIRO, STEWART [1997]: *Philosophy of Mathematics. Structure and Ontology*. Oxford: Oxford University Press.

—— [1998]: 'Incompleteness, mechanism, and optimism', *Bulletin of Symbolic Logic* **4**, 273–302.

URQUHART, ALASDAIR [forthcoming]: 'Emil Post', in the *Handbook of the History of Logic*.

WANG, HAO [1974]: *From Mathematics to Philosophy*. London: Routledge and Kegan Paul.

—— [1993]: 'On physicalism and algorithmism: Can machines think?', *Philosophia Mathematica* (3) **1**, 97–138.

—— [1996]: *A Logical Journey. From Gödel to Philosophy*. Cambridge, Mass.: MIT Press.