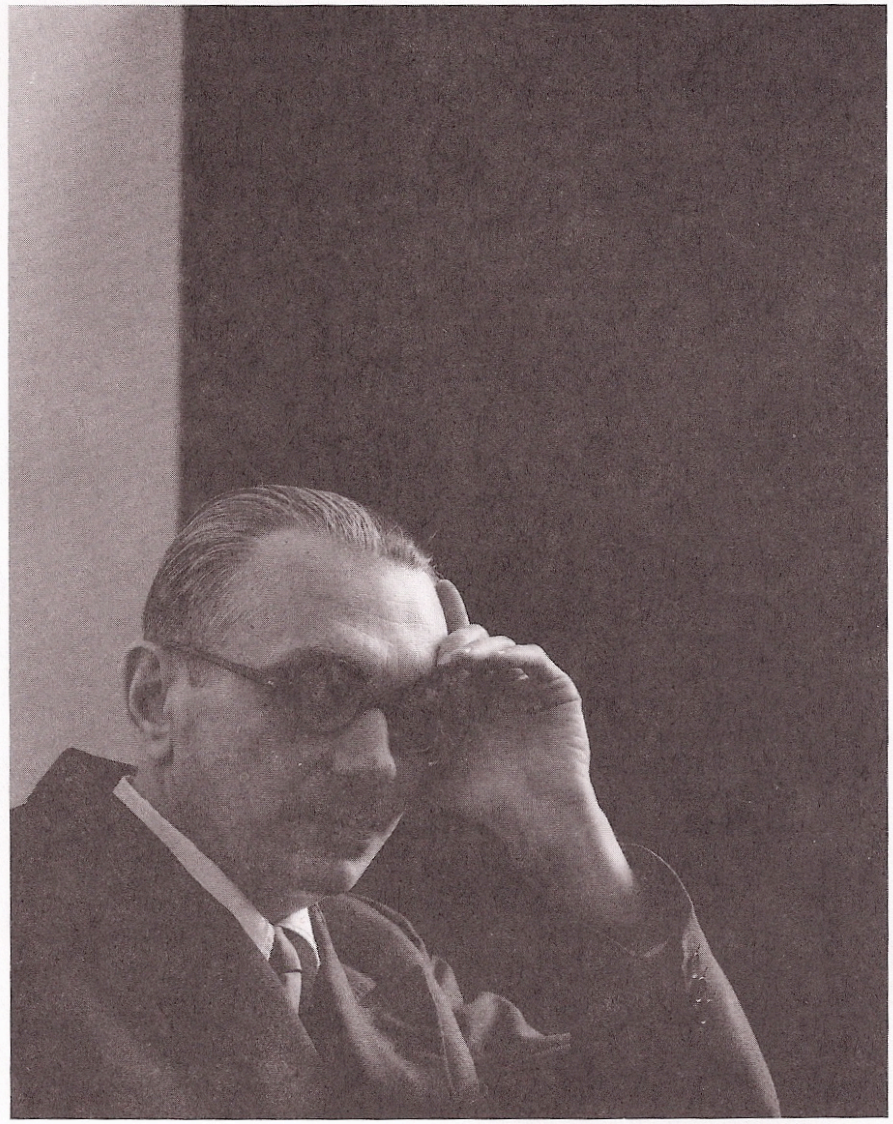
GÖDEL, NAGEL, MINDS AND MACHINES

Solomon Feferman

http://math.stanford.edu/~feferman/papers.html

Dr Kurt Gödel



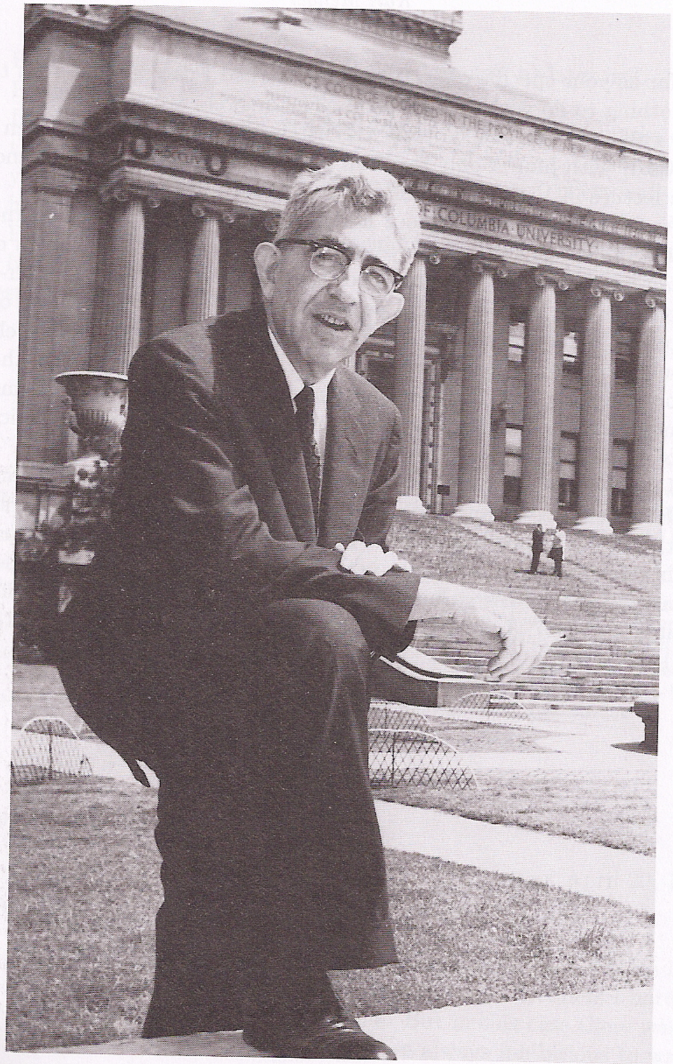
Kurt Gödel, 1956

The Gödel-Nagel imbroglio

Gödel's Proof, by Ernest Nagel and James R. Newman (NYU Press, 1957)

First popular exposition of Gödel's incompleteness theorems (1931)

Correspondence between Gödel and Nagel in Volume V of the Kurt Gödel *Collected Works*; introductory note by Charles Parsons and Wilfried Sieg



Ernest Nagel

Nagel and Newman

- Ernest Nagel, philosopher of science, Columbia
- James R. Newman, lawyer, mathematical enthusiast
- James R. Newman (ed.), *The World of Mathematics: A small library of the literature of mathematics from A'h-mosé the Scribe to Albert Einstein, presented with commentaries and notes*, vols. 1-4 (1957)

Nagel and Newman's proposal

An appendix that would include a translation of Gödel's 1931 paper on undecidable propositions together with the notes for lectures on that work that he had given during his first visit to the Princeton Institute for Advanced Study in 1934.

- Allan Angoff, editor of NYU Press

<u>Gödel to Angoff, April 1957</u>

Of course I shall have to see the manuscript of the book before I sign the contract, so that I can be sure that I am in agreement with its content, or that passages with which I don't agree can be eliminated, or that I may express my view about the questions concerned in the introduction.

Nagel to Angoff, May 1957

I could scarcely believe my eyes when I read his ultimatum that he is not only to see the manuscript of our essay before signing the contract with you, but that he is to have the right to eliminate anything in the essay of which he disapproves. In short, he stipulates as a condition of signing the contract the right of censorship.

This seems to me just insulting, and I decline to be a party to any such agreement with Gödel. … If [his] conditions were granted, Jim and I would be compelled to make any alterations Gödel might dictate, and we would be at the mercy of his tastes and procrastination for a period without foreseeable end. … Gödel is of course a great man, but I decline to be his slave.

<u>Nagel to Gödel, August 1957</u>

… I must say, quite frankly, that your … stipulation was a shocking surprise to me, since you were ostensibly asking for the right to censor anything of which you disapproved in our essay. Neither Mr. Newman nor I felt we could concur in such a demand without a complete loss of self-respect. I made all this plain to Mr. Angoff when I wrote him last spring, though it seems he never stated our case to you. I regret now that I did not write you myself, for I believe you would have immediately recognized the justice of our demurrer…

What were Gödel's concerns?

<u>Gödel to Nagel, unsent (?), March 1957</u>

Considerable advances have been made … since 1934. … it was only by Turing's work that it became completely clear, that my proof is applicable to every formal system containing arithmetic. I think the reader has a right to be informed about the present state of affairs.

Gödel PS to 1934 Princeton lectures

in Martin Davis (ed.) *The Undecidable* (1965)

In consequence of later advances, in particular of the fact that, due to A. M. Turing's work, a precise and unquestionably adequate definition of the general concept of formal system can now be given, the existence of undecidable arithmetical propositions and the non-demonstrability of the consistency of a system in the same system can now be proved rigorously for every consistent formal system containing a certain amount of finitary number theory.

Gödel PS to 1934 Princeton lectures, continued:

- Turing's work gives an analysis of the concept of "mechanical procedure" (alias "algorithm" or "computation procedure" or "finite combinatorial procedure"). This concept is shown to be equivalent with that of a "Turing machine".
- A formal system can simply be defined to be any mechanical procedure for producing formulas, called provable formulas. For any formal system in this sense there exists one in the [usual] sense … that has the same provable formulas (and likewise vice versa) ...

Gödel's possible concerns with the "Concluding Reflections" in Nagel and Newman

§1: The import of Gödel's conclusions is far-reaching, though it has not yet been fully fathomed. They seem to show that the hope of finding an absolute proof of consistency for any deductive system in which the whole of arithmetic is expressible cannot be realized, if such a proof must satisfy the finitistic requirements of Hilbert's original program.

[The incompleteness theorems] also show that there is an endless number of true arithmetical statements which cannot be formally deduced from any specified set of axioms … It follows, therefore, that an axiomatic approach to number theory … cannot exhaust the domain of arithmetical truth, and that mathematical proof does not coincide with the exploitation of a formalized axiomatic method.

Nagel &Newman, Concluding Reflections, §2:

Gödel's conclusions also have a bearing on the question whether calculating machines can be constructed which would be substitutes for a living mathematical intelligence. Such machines, as currently constructed and planned, operate in obedience to a fixed set of directives built in, and they involve mechanisms which proceed in a step-by-step manner.

But in the light of Gödel's incompleteness theorem, there is an endless set of problems in elementary number theory for which such machines are inherently incapable of supplying answers, however complex their built-in mechanisms may be and however rapid their operations.

It may very well be the case that the human brain is itself a "machine" with built-in limitations of its own, and that there are mathematical problems which it is incapable of solving. Even so, the human brain appears to embody a structure of rules of operation which is far more powerful than the structure of currently conceived artificial machines.

Nagel & Newman, *Concluding Reflections*,§ 3:

None of this is to be construed, however, as an invitation to despair, or as an excuse for mystery mongering. The discovery that there are formally indemonstrable arithmetic truths does not mean that there are truths which are forever incapable of becoming known, or that a mystic intuition must replace cogent proof. It does mean that the resources of the human intellect have not been, and cannot be, fully formalized, and that new principles of demonstration forever await invention and discovery. …

Nor do the inherent limitations of calculating machines constitute a basis for valid inferences concerning the impossibility of physico-chemical explanations of living matter and human reason. The possibility of such explanations is neither precluded nor affirmed by Gödel's incompleteness theorem. The theorem does indicate that in structure and power the human brain is far more complex and subtle than any nonliving machine yet envisaged.

The Gibbs Lecture

- "Some basic theorems on the foundations of mathematics and their implications" (1951, Kurt Gödel *Collected Works* Vol. III)
- SF, "Are there absolutely unsolvable problems? Gödel's dichotomy" (*Philosophia Mathematica*, 2006).

Gödel's Dichotomy

- *Either … 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 ...*

  [Italics Gödel's]

- *Diophantine problem*: a proposition of elementary number theory (aka first order arithmetic) of a relatively simple arithmetical form. The consistency of a formal system is equivalent to a diophantine problem.

Gödel's argument for his dichotomy

- If 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. [Italics Gödel's]

(Translations)

- *Finite machine*: a Turing machine
- *Well-defined axiomatic system*: an effectively specified formal system (according to Gödel, equivalent to a Turing machine)
- *Objective mathematics*: the totality of true statements of mathematics, which includes the totality of true statements of first-order arithmetic.

The assertion that objective mathematics is incompletable is a consequence of the second incompleteness theorem

- For any consistent effectively specified formal system S containing a certain basic system S_0 of (true) arithmetic, the number-theoretical statement Con(S) that expresses the consistency of S is true but not provable in S.
- NB 1: *the tacit assumption that the human mind is consistent* (as well as that S_0 is humanly accepted).
- NB 2: the dichotomy is not strict.

Nagel and Newman (N&N), Gödel (G), compared

- N&N: calculating machines "as currently constructed and planned, [that] operate in obedience to a fixed set of directives built in, and that involve mechanisms which proceed in a step-by-step manner."
- G: Turing machines
- N&N: incompleteness of formal systems (1st par.), incompleteness of calculating machines (2nd par);
- G: these are equivalent

N&N and G on the dichotomy

- N&N: the first incompleteness theorem implies the first disjunct (but they countenance the second disjunct)
- G: outside of the Gibbs lecture he was also convinced of the anti-mechanist position as expressed in the first disjunct of his dichotomy.

  (See: *From Mathematics to Philosophy* (Wang 1974), and *A Logical Journey. From Gödel to Philosophy* (Wang 1996, esp. Ch. 6).

Gödel countenances mind as machine in the Gibbs Lecture

- "[It is possible that] the human mind (in the realm of pure mathematics) is equivalent to a finite machine that, however, is unable to understand completely its own functioning."
- In other words, there is no *unassailable* human proof that mind is not equivalent to a machine.

Gödel countenances the *mechanist's empirical defense* or *escape hatch* in the Gibbs footnotes

(a) It 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.

(Gibbs footnotes, continued)

(b) The physical working of the thinking mechanism could very well be completely understandable; the insight, however, that this particular mechanism must always lead to correct (or only consistent) results would surpass the powers of human reason.

- But Gödel elsewhere speaks of the "impossibility of physico-chemical explanations of human reason."

Georg Kreisel on the escape hatch (1972)

It has been clear since Gödel's discovery of the incompleteness of formal systems that we could not have mathematical evidence for the adequacy of any formal system; but this does not refute the possibility that some quite specific system … encompasses all possibilities of (correct) mathematical reasoning … *In fact the possibility is to be considered that we have some kind of nonmathematical evidence for the adequacy of such [a system].* [italics mine]

Claims that the incompleteness theorems *prove* that mind is not mechanical

- J. R. Lucas, "Minds, machines and Gödel", 1961: "proof" that mind is not mechanical.
- Assumes that it is known that one's mind is consistent
- Lucas response to critics, 1996: The consistency of the machine that is supposed to model mind is established not by the mathematical ability of the mind, but on the word of the mechanist.
- Roger Penrose: Human mathematicians are not using a knowably sound algorithm in order to ascertain mathematical truth. (*Shadows of the Mind,* 1994)

Critiques of the "proofs"

See Stuart Shapiro, "Incompleteness, mechanism and optimism," *J. Symbolic Logic* 1998.

- None of the supposed proofs are resistant to the empiricists' escape hatch
- Infinitude of mathematics vs. finitude of mathematicians
- Idealization of what the individual mathematician (or community of logicians) can do, even *in principle*
- Analogous to the competence/performance distinction in linguistics

More critiques

- What about the assumption that the notions and statements of mathematics are fully and faithfully represented in a formal language (in order to compare with what a machine can process)?
- Even if restrict to the language of elementary number theory, what about accessibility to human comprehension of statements (like Con(S) for S an arbitrary formal system) of unlimited size?

Finite machine = Turing machine?

- Even if we accept Turing's idealization, none of the arguments against the mechanist are resistant to the empiricist defense.
- If the mechanist is right—that in some reasonable sense mind *is* equivalent to a finite machine—is it appropriate to formulate *that* in terms of the identification of what is humanly provable with what can be enumerated by a Turing machine?

Mechanists aims?

- Isn’t the mechanist instead after the *mechanisms* that govern the production of human proofs?
- The anti-mechanists’ <u>*equivocation*</u>: identifies *how the mathematical mind works* with *the totality of what it can prove.*
- Again analogous to the study of natural language, where the concern is with the *way* in which linguistically correct utterances are generated and *not* with the potential totality of all such

Modified formulation of the mechanist's (idealized) thesis

- Mind as constrained by the axioms and rules of some effectively presented formal system S.
- At best that identifies the mathematizing mind with the program for some non-deterministic Turing machine, and not with the set of its enumerable statements.

Still a problem

- Empirically possible, but implausible. No familiar system from Peano Arithmetic (PA) up to Zermelo-Fraenkel Set Theory (ZF) and beyond, actually underlies mathematical thought as it is experienced. No plausible candidate for that strengthened mechanist position.
- The reason for the implausibility of the modified version of the mechanist's thesis lies in the concept of a formal system S that is currently taken for granted in logical work, namely that the language of S is fixed once and for all.

Formal systems as currently conceived

- Peano Arithmetic (PA): Its language is determined by the choice of =, 0, 1,+, × as basic symbols.
- The induction axiom scheme of PA consists of all instances of $F(0)\ \&\ \forall x[F(x) \Rightarrow F(x+1)] \Rightarrow \forall x\, F(x)$

  for F(x) a well-formed formula of the language of PA.
- Zermelo-Fraenkel (ZF): language =, ϵ, Separation Scheme $\forall a \exists b \forall x[x \in b \Leftrightarrow x \in a\ \&\ F(x)]$

One way to straddle the Gödelian Dichotomy

- Modify the conception of schematic axiom system S so that the language of S is not fixed in advance but is open-ended
- Replace infinite scheme bundles by formal schemes for axioms and rules of inference.

  Implicit in the acceptance of given schemata is the acceptance of any meaningful substitution instances that one may come to meet in any context.

Examples

- The idea is familiar from logic, e.g. axioms like $P\&Q \Rightarrow P$ and $\forall x[P(x) \Rightarrow Q(x)] \Rightarrow [\forall x P(x) \Rightarrow \forall x Q(x)]$

  and rules of inference like: infer Q from P and $P \Rightarrow Q$.

- Induction axiom scheme: $P(0) \& \forall x[P(x) \Rightarrow P(x+1)]$

  $\Rightarrow \forall x P(x)$

- Separation scheme: $\forall a \exists b \forall x[x \in b \Leftrightarrow x \in a \;\&\; P(x)]$

Conclusions

- A mechanistic/anti-mechanistic hypothesis: All of mathematical reasoning is constrained by a finite number of such schematic axioms and rules of inference. But the language of mathematics is essentially open-ended.
- What logic can't supply: the cognitive science of actual mathematical reasoning (use of analogies, metaphors, physical and geometric intuition, visualization, etc.)

Reading recommendations for Gödel's theorems

- Torkel Franzén, *Gödel's Theorem. An incomplete guide to its use and abuse*, A. K. Peters (2005)
- Torkel Franzén, *Inexhaustibility: A non-exhaustive treatment*, A. K. Peters (2004)
- Peter Smith, *An Introduction to Gödel's Theorems*, Cambridge (2007)