Gödel's Theorem, or an Evening with Mr Homais

Jean-Yves Girard

English translation

The other evening at the Club, I meet Homais beside the pedestal table; he tells me point-blank that he will not stand again for mayor of Yonville.

Homais: I no longer believe in politics; besides, as Régis Debray has said, Gödel's theorem destroys the very idea of a system.

Me: How dreadful! What exactly is this Claudel theorem?

H *(condescending)*: Gödel, my dear fellow, not Claudel... Have you never read Gödel, Escher, Bach?

M: Escher, yes... Planète, around 1963. So what?

H: Well, the theorem is categorical: it says that a system cannot justify itself.

M: What kind of system?

H *(pedantically)*: A ma-the-ma-ti-cal system, my dear fellow, a for-mal theory.

M: Whew, that's a relief: at least our system of body hair is safe... Forgive the cheap joke; I mean our philosophical system!

H: Sor-ry, my dear fellow: all systems - philosophical, artistic, political... In fact, Gödel proved that thought cannot think itself, that our intelligence is limited...

M *(under my breath)*: Speak for yourself, old man...

M *(aloud)*: Surely a system can still be improved. Even Windows 95 has a service that takes care of bugs... Why not do the same with philosophy, art...?

H *(triumphantly)*: But the improved system will be imperfect too, since Gödel's theorem applies to every system; besides, patched versions of Windows are just as buggy. You see: there is no hope.

M: But if you are giving up politics, what are you going to do?

H: Numerology. That surprises you, but think for a moment: if no theory is possible, everything becomes possible - hence magic, my dear fellow. And the first of all magicians is Gödel: he proved his theorem with magic numbers. At bottom, logician = magician. I await The Morning of the Logicians.

M: Magic, numerology... it sounds a little like Paco Rabanne.

H: I call it AI: ar-ti-fi-cial in-tel-li-gence.

M: So you need some?

H *(piqued)*: My friend, you should know that every property has a Gödel number, and that all one need do is calculate - as on Des chiffres et des lettres, 'Numbers and Letters' - to obtain an answer to every question. Naturally, we use a computer to go faster.

(Knocking from inside the pedestal table.)

M: But you have just said that no system is perfect. Isn't your numerological computer a system too?

H *(pauses, then)*: Do you realize that mathematicians still do not know how to divide by zero? For me, the answer is 7, even if these gentlemen laugh... Better a lame answer than no answer at all! And besides, 7 signifies...

M: So much for logic!

H: It is non-monotonic logic, that is, non-deductive logic: 4 / 0 = 7, but one does not infer from this that 7 x 0 = 4.

M: In other words, your logic is a kind of car without an engine.

H *(sneering)*: What does that matter, if the dashboard is made of burr walnut...?

M: So, to sum you up: theoretically nothing works, but in practice one can do whatever one likes... What a strange idea to abandon politics when you had such a promising career ahead of you in Education!

The pedestal table begins to move; Homais places on it a computer connected to a stethoscope; the pedestal table joins the conversation.

Pedestal table: All this ♠♠♠ you have been talking - a German word - has woken me up.

H *(terrified)*: My... Sir...

G: Gödel. Kurt Gödel, 1906-1978. I am the one who proved the two incompleteness theorems in 1931, and I have rarely heard such a tissue of positivist ♠♠♠. You have not missed a single one.

H: You, the destroyer of theories... you who revealed the limitations of human thought...

G: Let us separate the issues. I showed certain limitations of formal processes, and those limitations apply to every deductive approach. They therefore prevent you from carrying your AI project through to completion.

H: I merely provide answers, outside any formal system; and besides, this is a Republic!

M: One could always answer 'yes': it is simple and offends no one.

G *(embarrassed)*: I had not thought of so radical a solution, because my theorem applies only when one answers a certain number of basic questions correctly, as ordinary formal systems do.

H *(triumphantly)*: Very well: we shall answer the basic questions correctly, and then complete the system mechanically. We take a formal system T, and whenever A is not provable in T, we add its negation, ¬A. Our logic is based on the idea that not knowing means knowing that not.

M: At a pedestrian crossing, the red light no longer comes on. Do you cross, then?

G: Gentlemen... That is precisely the principle refuted by my first incompleteness theorem. When a formal search goes on and on, how can one know that it will not produce an answer later - in an hour, in a year, in a century? It is simple and irrefutable: there is no way to know.

M: Doubting Thomas, you go too far! Are we to know only what we have verified with our own eyes, without being able to predict anything?

G: Prediction is sometimes possible, but only case by case, and it requires intelligence. Any systematic recipe, however, will necessarily make mistakes.

H *(sulkily)*: I see that you oppose technological progress. Let us return to systems and human intelligence. You must admit that the ggrrreat ideas take quite a blow from your second theorem.

G: Only insofar as those theories present themselves in a precise formalism - which is never the case outside mathematics...

M: The mathematicians I know are not all that precise; besides, they write in English, which is not a formal language.

G: True, but the total formalization of mathematics is possible in principle... and the computer has made it a reality.

H *(startled)*: I no longer understand anything at all. You silenced me on AI, and now you are taking back what you said.

G: Not at all! We have software capable of writing out a complete formal proof from 'human' language. But it does not invent anything: the only intelligence involved is that of the prover, and also that of the software designer.

H: AI has nevertheless advanced further than that: H. Simon built a program that rediscovered one of Kepler's laws entirely by itself.

G: Kepler told me that by far the hardest part was not deriving the law from the numerical data, but thinking to analyze those data rather than others. He would also like the computer to rediscover the abandoned fourth law, the one connecting the planets with the regular polyhedra.

H: So Kepler does not appreciate Simon's work?

G: No. He even believes that he found the law before the machine did.

H: Very well. So incompleteness is connected with the precision of language?

G: And that is why one can hardly see how to apply it to politics, painting, or poetry. Let me remind you that the great Spinoza wanted to produce a purely formal Ethics: the first page looks like mathematics riddled with errors; as for what follows, the less said the better.

H: But you associated numbers with statements when you conceived that project for a formal philosophy designed as a kind of arithmetic?

G: Wrong. Leibniz was the one who had that idea.

H: Which you took up.

G *(sighing)*: Not really. Leibniz was heading more in the direction of the Kabbalah; and, like many commentators, you believe that my numbers have a meaning. Sorry: they have none. They are merely an enumeration with no remarkable property. My first incompleteness theorem is only a property of enumerations already observed by Cantor in the nineteenth century. It is an elaborated form of the Liar paradox, or the Cretan paradox.

H: Your second incompleteness theorem says that a theory cannot speak about itself, and that one must jump to a 'meta' level.

G: The second theorem says that the formula asserting the consistency of a theory T is not provable within T itself... which presupposes that this is a formula of T - in other words, that T can speak about itself. It was Hilbert, around 1900, who discovered that mathematics could speak about itself. The Gödel numbers that impress you so much are merely a way of making this internalization, this reflection, explicit.

H: All right, all right... one can speak about oneself, but one cannot know oneself; one cannot be one's own 'meta'. Admit that the idea reaches far beyond your narrow formal framework!

G *(mischievously)*: All the more so because the consistency formula is of little interest, and we had to wait until the 1970s to find more interesting examples of incompleteness. Outside mathematics it is easier: the second theorem tells us that 'one cannot tighten the screws on one's glasses while still wearing them,' or that 'the frame of the painting is not the painting'... But we should not forget that glasses are made for seeing, not for having their screws tightened.

M: As for those who shift the problem from the canvas to the frame, that is probably more a matter of incompetence than incompleteness!

H: Do not mock modern art, my friend. Each school announces the next, which itself prefigures another, in a whirl...

G: System, then meta-system, then meta-meta-system, and so on. Spare glasses, then, are meta-glasses... This Greek word, meaning 'beside, after,' has acquired a perverse meaning. It conceals more than it explains.

H: But the reality that exists prior to formalism - that too is the 'meta'!

G: That is Tarski's realist viewpoint, dominant today: a stable external world of which we are both part and observer. A formula is therefore true or false - an absolute notion, as realism requires - and provable, refutable, or neither - a notion relative to a theory T. My first theorem gives a true statement that is not provable in T; the second gives a more particular statement: the consistency of T.

H: Even so, you missed Tarski's theorem, the one saying that truth is not definable in T.

G *(with a small laugh)*: If you like. In fact, when I began my work, it was taken for granted that truth and provability were identical. So I formally defined truth by means of provability and reused Cantor's old argument to obtain a contradiction in T. Only then did I understand that what I had actually proved was the contradiction of the hypothesis 'true = provable' - hence my first theorem. A few years later Tarski 'rediscovered' this primitive version of my theorem; it cannot have exhausted him.

H: You are very hard on Tarski.

G: The principal strength of Tarskism is its flatness - a flatness corresponding to Western rationalism. Yet however much one speaks of mathematical reality, one never manipulates anything but systems: the only reference should therefore be internal. At bottom, completeness and incompleteness are internal properties, referring to no pre-existing external world. But all this requires profound reflection, which will be the work of the coming century.

M: And might thereby correct the positivism of twentieth-century logic.

G: Let us hope so. Besides, not everything in twentieth-century logic should be discarded. Take Hilbert: his program, which my theorem refutes, was a program for the internal justification of mathematics by consistency proofs. It can be seen as a reductionist program inspired by a very Germanic scientism; but it can also be seen as an attempt to escape realism, to escape truth. Formal consistency is a form of immanence - unfortunately, a poor man's immanence.

H: All right, all right... and the limitation of human intelligence?

M: Mr Gödel is telling you that the theorem limits formal systems. So you see your intelligence as a formal system, as a machine?

H *(pompously)*: Indeed. I have the impression that everything I do ought to be done by a machine, which could moreover speak just as well in my place.

G: From where I am, it is difficult to tell whether you exist or whether you are the virtual creation of a GAT - a Generator of Automatic Truisms. Intelligence does not exist without error, perhaps even without obstinacy in error; but who would take the risk of giving a computer that kind of psychology? As for the incompleteness theorem, it certainly did not foresee bad-tempered theories...

H: Then this theorem says nothing?

G: It states one primary impossibility among many others.

M: So, if I have understood correctly, there is no way to read the 'Great Book of Truth' kept by Mr Meta-Tarski, any more than there is a 'Great Book of Initial Positions' from which to predict the roulette wheel.

G: Quantum mechanics goes so far as to deny the existence of that Book of Positions: beyond a certain degree of precision, the very notion of position loses its meaning. It is possible that the Book of Truth has no more reality than that.

H: So in the end your thing is useful only for wrecking everyone else's constructions?

G: True. It does not show the way; it shows only where not to go. Is that not enough for you?

Exasperated, Homais abruptly switches off his computer (which is not recommended).

H: You know what? I do not think that was Gödel. If pedestal tables start lying now...!

LOGICAL PARADOXES

The oldest known paradox is that of the Cretan: 'Cretans are liars.' This is not really a paradox, because the Cretan, although a liar, may for once have told the truth; or there may be another Cretan who is not a liar. A genuine paradox arises only with a single Cretan and a single statement: 'I am lying' - meaning that what I am saying is a lie. The Liar paradox is not a mathematical paradox, but it underlies diagonalization, which is at work in Cantor's and Russell's paradoxes and in the first incompleteness theorem.

Cantor proved that one cannot enumerate all sets of integers. Indeed, suppose we have such an enumeration X_n (so $m \in X_n$ means that the integer m belongs to the nth set), and consider the set $Y = \{m \mid m \notin X_m\}$ of integers that do not belong to the set bearing the same number. This set Y must be equal to one of the X_n, say X_N; but $N \in X_N \leftrightarrow N \in Y \leftrightarrow N \notin X_N$. Diagonalization consists in setting $m = n$ in $m \in X_n$.

Diagonalization appears again in Russell's paradox (1902). Let A be the set of all sets that do not belong to themselves, $A = \{x \mid x \notin x\}$. Then A belongs to itself if and only if it is not an element of itself: $A \in A \leftrightarrow A \notin A$. This paradox marks the starting point of the foundational crisis (see p. 6).

Russell's result has many paraphrases. For example, one cannot compile a catalogue of all catalogues that do not mention themselves. In the same spirit, Richard's paradox (1905) considers 'the smallest integer that cannot be defined in fewer than one hundred letters' - a phrase that defines an integer in fewer than one hundred letters. This is not a mathematical paradox, since we would be hard pressed to define the word 'define'; but something of Richard's paradox passes into the incompleteness theorem. In the same vein, Queneau's paradox considers the first uninteresting integer, only to observe that this makes it interesting... precisely for that reason.

The first incompleteness theorem is Cantor's paradox applied to the enumeration of the theorems - the provable statements - of T (see p. 7). Many attempts have been made to get around it: people thought that one merely had to torture logic a little... In fact, the first theorem is more robust than the second, because it does not really assume that T is a logical system. For example, the theorem tells us that there is no algorithm for detecting 'loops,' that is, the non-termination of computations. The existence of such an algorithm would be equivalent to the existence of a loop-free computational method: if loops can be detected, they can be eliminated. Let f_n be an enumeration of algorithms defining functions from the integers to the integers. Then $g(n) = f_n(n) + 1$ defines a total algorithm whose consideration leads to a contradiction: if $g = f_N$, then $f_N(N) = g(N) = f_N(N) + 1$. This paradox shows the absurdity of the 'no loops' hypothesis, because in reality $f_N(N)$ is computed as $f_N(N) + 1$, which is computed as $f_N(N) + 2$... Thus there is no loop-testing algorithm.

In other words, the absence of a result cannot in any way be treated as a result. Knowledge does not commute with negation: 'not knowing' and 'knowing that not' can under no circumstances be identified. There is no way around incompleteness, which is not even a logical limitation, since it is a general property of enumerations. The result displeases people, like a stain on logic. We should set the record straight: who would want a complete, Confucian world in which everything had been found and parents, hierarchical superiors, priests, and so on were always right?

THE FOUNDATIONAL CRISIS

Developing an idea of Raymond Lull's, Leibniz conceived a calculus ratiocinator that would represent properties by numbers and make it possible to resolve every question by an arithmetical calculation.

At the end of the nineteenth century, Cantor introduced new entities, sets. They were initially used for delicate questions in analysis, such as Fourier series, and were then systematized by means of a single principle, the comprehension schema, according to which every property P defines a set: the set of all x such that P[x], written $\{x \mid P[x]\}$. This naive set theory has one immense virtue. By reducing all mathematical constructions - integers, real numbers, and so on - to sets, it gives the various branches of mathematics, such as analysis and algebra, a unity of principle, since everything can, in principle, be translated into a single language.

Unfortunately, Burali-Forti found a contradiction in it in 1898, simplified by Russell in 1902. For mathematics, the danger was limited, since it was confined to a new and very marginal domain; moreover, around 1910 Zermelo discovered the 'good' version of set theory, which is still in use today. This is why the overwhelming majority of mathematicians completely ignored the crisis, which Hilbert and Brouwer, by contrast, dramatized.

Hilbert, the white knight of mathematics, wanted to eliminate the paradoxes by means of an internal justification of mathematics: mathematical methods would be used to prove that mathematics is not contradictory. This seems doubtful at first sight, but Hilbert was anything but an idiot. The proof would use very elementary arithmetic, which itself needed no justification. This formulation of Hilbert's program is refuted by the second incompleteness theorem. If T_0 is the system of 'elementary' arithmetic, then the formula asserting the formal consistency of T_0 is not provable in T_0. If 'elementary' arithmetic cannot justify itself, it is difficult to see how it could justify anything stronger than itself.

A less well-known version of Hilbert's program is the elimination of the infinite. If a statement A of the 'refutable' kind is provable in mathematics - at the time this morally meant: if A is true - then it can already be proved by the limited methods of T_0. The first incompleteness theorem refutes this form of the program, and more generally Leibniz's project, of which Hilbert's program is only a variant. If the elementary mathematics of T_0 suffices, then one can 'ratiocinate' by attempting a proof within T_0.

Brouwer proposed a highly original path: radically changing the rules of logic. Intuitionism therefore rejects proof by contradiction - or, equivalently, the law of excluded middle, $A \vee \neg A$. Its account of the world is non-realist, with a tendency toward subjectivism. Brouwer and Hilbert meet in their rejection of realism: Brouwer saw mathematics only as an activity of the mind, whereas Hilbert saw in it only formal properties of language without any external reference, or at least with only a minimal external reference. Apart from that, their positions could hardly have been more opposed: Brouwer was something of a mystic, while Hilbert was a hard-line scientistic thinker. With time, and despite the failure of their respective programs, there is still far more to glean around these two questionable figures than along the avenues of reality paved over with concrete by Tarski.

THE INCOMPLETENESS THEOREMS

Gödel numbers are simply an enumeration of the formal expressions in a given language. One need only adopt, say, a lexicographic ordering of those expressions in order to assign a Gödel number G(A) to every expression. This plainly belongs to the 'positive' part of Hilbert's program: carrying out the program requires that logical artefacts - formulas and proofs - be encoded by integers, and Gödel was indeed the first to do so. We therefore fix a formal system T in which a little arithmetic can be performed - very little suffices. We assume the consistency, or non-contradiction, of T. If T is inconsistent, then everything becomes provable in it, so it is easy to see that this hypothesis is strictly necessary for what follows.

Next consider the one-parameter statement B[n]: 'The formula numbered n, with n substituted for its parameter - namely $A_n[n]$ - is not provable.' B is itself a formula with one parameter, and therefore has a Gödel number N. If we assign the value N to the parameter n, then B[N] literally means 'B[N] is not provable,' or 'I am not provable.' If provable = true, we obtain the genuine Liar paradox, 'I am not true'; provability and truth must therefore diverge. In fact, Gödel's formula is true without being provable, provided of course that T is consistent. It destroys the Leibnizian form of Hilbert's program, the calculus ratiocinator. This is the first form of incompleteness.

A new hall of mirrors: we have in fact just proved something - Gödel's formula - since we know that it is true; but we have proved it only under the hypothesis that T is consistent. All this can be perfectly formalized and yields, within T, a proof of the implication $Con(T) \Rightarrow B_N[N]$. Since $B_N[N]$ is not provable in T, neither is its formal consistency, or logical non-contradiction, still under the hypothesis that T is consistent, of course. This is the second form of incompleteness, by far the best known.

But what does completeness mean? Traditionally, it is the adequacy between a language and an interpretation, and that adequacy depends on the logical statements under consideration. Thus, if a statement of the form 'there exists an n such that $f(n) = 0$' - also called a verifiable statement, since one need only try the values $n = 0, 1, 2, 3, \ldots$ successively - is true, then it is provable. This is the spirit, if not the letter, of the completeness theorem of... Gödel, from 1930. If we call refutable the negation of a verifiable statement - that is, a statement of the form 'for every integer n, $f(n) \neq 0$,' or one reducible to that form, as in the case of Fermat's theorem, consistency, or Gödel's formula B[N] - then no false refutable statement can be proved in a consistent theory. For if a refutable statement A is false, its negation $\neg A$ is verifiable and true, and therefore provable by completeness.

This is also why Gödel's formula B[N] is true: if it were false, it would be provable, and hence true. Incompleteness therefore applies as soon as formulas are refutable. That said, this incompleteness in principle rarely manifests itself in interesting regions of mathematics, to the point that most mathematicians have never encountered any manifestation of it.

A more internal meaning of completeness would be: 'nothing is missing.' Why is nothing missing from pure logic, whereas something is necessarily missing from arithmetic? Can one speak of lack without presupposing totality? These are important questions for the century to come.

Source: Jean-Yves Girard, "Le théorème de Gödel ou une soirée avec M. Homais."