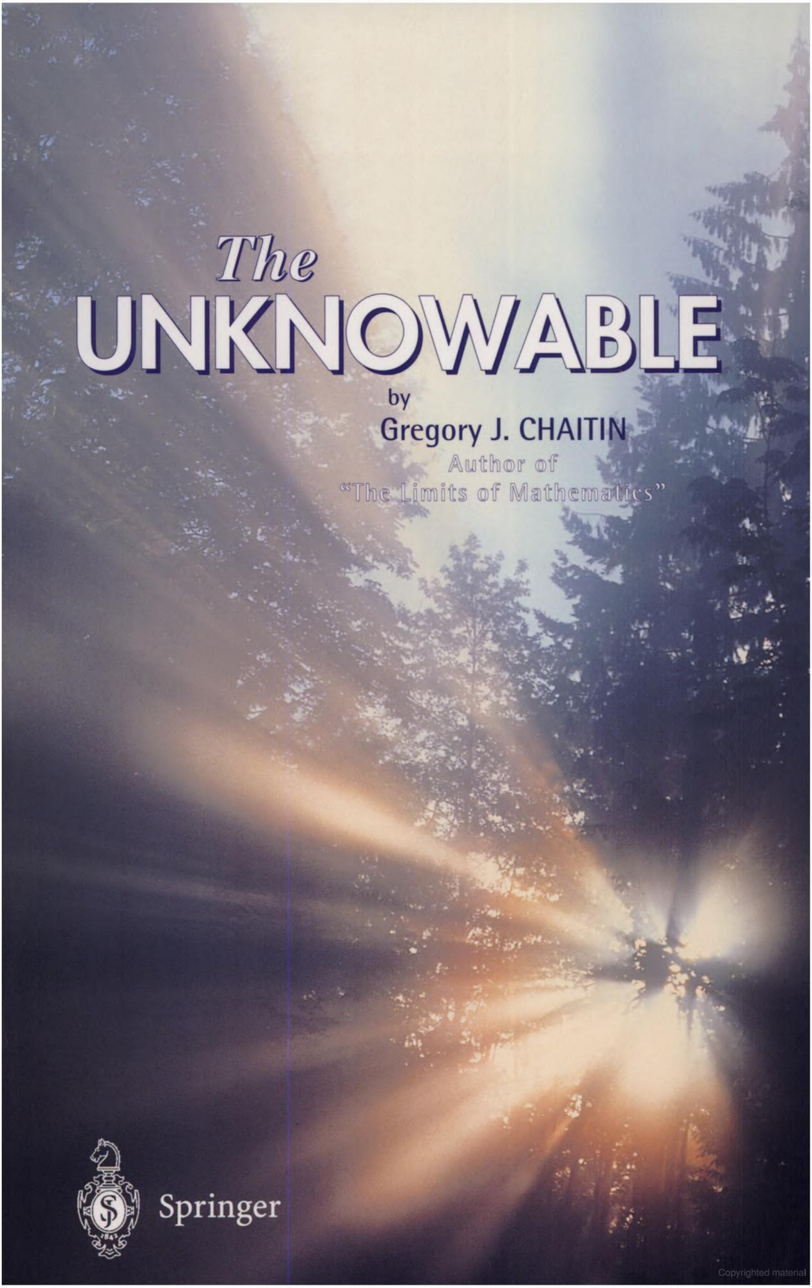

The UNKNOWABLE

by

Gregory J. CHAITIN

Author of
"The Limits of Mathematics"

Springer

The UNKNOWABLE

by

Gregory J. CHAITIN

Springer

Gregory J. CHAITIN
IBM Research Division
Thomas J. Watson Research Center
30 Saw Mill River Road
Hawthorne, NY 10532
USA

Library of Congress Cataloging-in-Publication Data

Chaitin, Gregory J.
The unknowable / Gregory J. Chaitin.
p. cm. -- (Springer series in discrete mathematics and theoretical computer science)
Includes bibliographical references.
ISBN 9814021725 (hardcover)
1. Machine theory. 2. Computational complexity. 3. Stochastic processes. I. Title. II. Series.
QA276.C437 1999
511.3--dc21 99-29345
CIP

ISBN 981-4021-72-5

Printed in Singapore
Typesetting: Camera-ready by the Author
SPIN 10790398 5 4 3 2 1

I

A Hundred Years of Controversy Regarding the Foundations of Mathematics[1]

Synopsis

What is metamathematics? Cantor's theory of infinite sets. Russell on the paradoxes. Hilbert on formal systems. Gödel's incompleteness theorem. Turing on uncomputability. My work on randomness and complexity. Is mathematics quasi-empirical?... The computer and programming languages were invented by logicians as the unexpected by-product of their unsuccessful effort to formalize reasoning completely. Formalism failed for reasoning, but it succeeded brilliantly for computation. In practice, programming requires more precision than proving theorems!... Each step Gödel ⇒ Turing ⇒ Chaitin makes incompleteness seem more natural, more pervasive, more ubiquitous—and much more dangerous!

What is Metamathematics?

In this century there have been many conceptual revolutions. In physics the two big revolutions this century were relativity theory and quantum mechanics: Einstein's theory of space, time and gravitation, and the

[1]Based on a lecture on "Cien años de controversia sobre los fundamentos de las matemáticas" given at several institutions during two visits to Buenos Aires in 1998.

theory of what goes on inside the atom. These were very revolutionary revolutions, dramatic and drastic changes of viewpoint, paradigm shifts, and very controversial. They provoked much anguish and heartache, and they marked a generational shift between so-called classical physics and modern physics.

Independently an earlier revolution, the statistical viewpoint, has continued, and now almost all physics, classical or modern, is statistical. And we are at the beginning of yet another conceptual shift in physics, the emphasis on chaos and complexity, where we realize that everyday objects, a dripping faucet, a compound pendulum, the weather, can behave in a very complicated and unpredictable fashion.

What's not much known by outsiders is that the world of pure mathematics hasn't been spared, it's not immune. We've had our crises too. Outsiders may think that mathematics is static, eternal, perfect, but in fact this century has been marked by a great deal of anguish, hand-wringing, heartache and controversy regarding the foundations of mathematics, regarding its most basic tenets, regarding the nature of mathematics and what is a valid proof, regarding what kinds of mathematical objects exist and how mathematics should be done.

In fact, there is a new field of mathematics called *metamathematics,* in which you attempt to use mathematical methods to discuss what mathematics can and cannot achieve, and to determine what is the power and what are the limitations of mathematical reasoning. In metamathematics, mathematicians examine mathematics itself through a mathematical microscope. It's like the self-analysis that psychiatrists are supposed to perform on themselves. It's mathematics looking at itself in the mirror, asking what it can do and what it can't do.

In this book I'm going to tell you the story of this century's controversies regarding the foundations of mathematics. I'm going to tell you why the field of metamathematics was invented, and to summarize what it has achieved, and the light that it sheds—or doesn't—on the fundamental nature of the mathematical enterprise. I'm going to tell you the extent to which metamathematics clarifies how mathematics works, and how different it is or isn't from physics and other empirical sciences. I and a few others feel passionately about this.

It may seem tortured, it may seem defeatist for mathematicians to question the ability of mathematics, to question the worth of their craft.

In fact, it's been an extraordinary adventure for a few of us. It would be a disaster if most mathematicians were filled with self-doubt and questioned the basis for their own discipline. Fortunately they don't. But a few of us have been able to believe in and simultaneously question mathematics. We've been able to stand within and without at the same time, and to pull off the trick of using mathematical methods to clarify the power of mathematical methods. It's a little bit like standing on one leg and tying yourself in a knot!

And it has been a surprisingly dramatic story. Metamathematics was promoted, mostly by Hilbert, as a way of confirming the power of mathematics, as a way of perfecting the axiomatic method, as a way of eliminating all doubts. But this metamathematical endeavor exploded in mathematicians' faces, because, to everyone's surprise, this turned out to be impossible to do. Instead it led to the discovery by Gödel, Turing and myself of metamathematical results, incompleteness theorems, that place severe limits on the power of mathematical reasoning and on the power of the axiomatic method.

So in a sense, metamathematics was a fiasco, it only served to deepen the crisis that it was intended to resolve. But this self-examination **did** have wonderful and totally unexpected consequences in an area far removed from its original goals. It played a big role in the development of the most successful technology of our age, the computer, which after all is just a mathematical machine, a machine for doing mathematics. As E.T. Bell put it, the attempt to soar above mathematics ended in the bowels of a computer!

So metamathematics did not succeed in shoring up the foundations of mathematics. Instead it led to the discovery in the first half of this century of dramatic incompleteness theorems. And it also led to the discovery of fundamental new concepts, computability and uncomputability, complexity and randomness, which in the second half of this century have developed into rich new fields of mathematics.

That's the story I'm going to tell you about here, and it's one in which I'm personally involved, in which I'm a major participant. So this will not be a disinterested historian's objective account. This will be a very biased and personal account by someone who was there, fighting in the trenches, shedding blood over this, lying awake in bed at night without being able to sleep because of all of this!!

What provoked all this? Well, a number of things. But I think it's fair to say that more than anything else, the crisis in the foundations of mathematics in this century was set off by G. Cantor's theory of infinite sets. Actually this goes back to the end of the previous century, because Cantor developed his theory in the latter decades of the 19th century. So let me start by telling you about that.

Cantor's Theory of Infinite Sets

So how did Cantor create so much trouble!? Well, with the simplicity of genius, he considered the so-called **natural numbers** (non-negative integers):

$$0, 1, 2, 3, 4, 5, \ldots$$

And he asked himself, "Why don't we add another number after all of these? Let's call it ω!" That's the lowercase Greek letter omega, the last letter of the Greek alphabet. So now we've got this:

$$0, 1, 2, \ldots \omega$$

But of course we won't stop here. The next number will be $\omega + 1$, then comes $\omega + 2$, etc. So now we've got all this:

$$0, 1, 2, \ldots \omega, \omega + 1, \omega + 2, \ldots$$

And what comes after $\omega + 1, \omega + 2, \omega + 3, \ldots$? Well, says Cantor, obviously 2ω, two times ω!

$$0, 1, 2, \ldots \omega, \omega + 1, \omega + 2, \ldots 2\omega$$

Then we continue as before with $2\omega + 1, 2\omega + 2$, etc. Then what comes? Well, it's 3ω.

$$0, 1, 2, \ldots \omega, \omega + 1, \omega + 2, \ldots 2\omega \ldots 3\omega$$

So, skipping a little, we continue with $4\omega, 5\omega$, etc. Well, what comes after all of that? Says Cantor, it's ω squared!

$$0, 1, 2, \ldots \omega, \omega + 1, \omega + 2, \ldots 2\omega \ldots 3\omega \ldots \omega^2$$

Then we eventually have ω cubed, ω to the fourth power, etc.

$$0, 1, 2, \ldots \omega, \omega + 1, \omega + 2, \ldots 2\omega \ldots 3\omega \ldots \omega^2 \ldots \omega^3 \ldots \omega^4 \ldots$$

Then what? Well, says Cantor, it's ω raised to the power ω!

$$0, 1, 2, \ldots \omega, \omega + 1, \omega + 2, \ldots 2\omega \ldots \omega^2 \ldots \omega^3 \ldots \omega^\omega \ldots$$

Then a fair distance later, we have this:

$$0, 1, 2, \ldots \omega, \omega + 1, \omega + 2, \ldots 2\omega \ldots \omega^2 \ldots \omega^\omega \ldots \omega^{\omega^\omega} \ldots$$

Then much later we start having trouble naming things, because we have ω raised to the ω an infinite number of times. This is called ϵ (epsilon) nought.

$$\epsilon_0 = \omega^{\omega^{\omega^{\cdots}}}$$

It's the smallest solution of the equation

$$\omega^\epsilon = \epsilon.$$

Well, you can see that this is strong stuff! And as disturbing as it is, it's only half of what Cantor did. He also created another set of infinite numbers, the **cardinal** numbers, which are harder to understand.[2] I've shown you Cantor's **ordinal** numbers, which indicate positions in infinite lists. Cantor's cardinal numbers measure the size of infinite sets. A set is just a collection of things, and there's a rule for determining if something is in the set or not.

Cantor's first infinite cardinal is aleph nought

$$\aleph_0$$

which measures the size of the set of natural numbers (non-negative integers). Aleph is the first letter of the Hebrew alphabet. Then comes aleph one

$$\aleph_1$$

[2] *Note for experts:* To simplify matters, I'm assuming the generalized continuum hypothesis.

It's like the paradox of the village barber who shaves every man in the village who doesn't shave himself. But then who shaves the barber?! He shaves himself iff (if and only if) he doesn't shave himself. Of course, in the case of the barber, there is an easy way out. We either deny the existence of such a barber, for he can't apply the rule to himself, or else the barber must be female! But what can be wrong with Russell's set of all sets that are not members of themselves?

The Russell paradox is closely related to a much older paradox, the liar paradox, which is also called the Epimenides paradox and goes back to classical Greece. That's the paradox **"This statement is false!"** It's true iff it's false, and therefore it's neither true nor false!

Clearly, in both cases the paradox arises in some way from a self-reference, but outlawing all self-reference would be throwing out the baby with the bath water. In fact, self-reference will play a fundamental role in the work of Gödel, Turing, and my own that I'll describe later. More precisely, Gödel's work is related to the liar paradox, and Turing's work is related to the Russell paradox. My work is related to another paradox that Russell published, which has come to be called the Berry paradox.

What's the Berry paradox? It's the paradox of **the first natural number that can't be named in less than fourteen words**. The problem is that I've just named this number in thirteen words! (Note that the existence of this number follows from the fact that only finitely many natural numbers can be named in less than fourteen words.)

This paradox is named after G.G. Berry, who was a librarian at Oxford University's Bodleian library (Russell was at Cambridge University), because Russell stated in a footnote that this paradox had been suggested to him in a letter from Berry. Well, the Mexican mathematical historian Alejandro Garciadiego has taken the trouble to find that letter, and it's a rather different paradox. Berry's letter actually talks about **the first ordinal that can't be named in a finite number of words**. According to Cantor's theory such an ordinal must exist, but we've just named it in a finite number of words, which is a contradiction.

These details may not seem too interesting to you, but they're tremendously interesting to me, because I can see the germ of my work in Russell's version of the Berry paradox, but I can't see it at all in

Berry's original version.[3]

Hilbert on Formal Systems

So you see, Cantor's set theory was tremendously controversial and created a terrific uproar. Poor Cantor ended his life in a mental hospital.

What was to be done? One reaction, or over-reaction, was to advocate a retreat to older, safer methods of reasoning. The Dutch mathematician L.E.J. Brouwer advocated abandoning all non-constructive mathematics. He was in favor of more concrete, less "theological" mathematics.

For example, sometimes mathematicians prove that something exists by showing that the assumption that it doesn't exist leads to a contradiction. This is often referred to in Latin and is called an existence proof via *reductio ad absurdum,* by reduction to an absurdity.

"Nonsense!" exclaimed Brouwer. The only way to prove that something exists is to exhibit it or to provide a method for calculating it. One may not actually be able to calculate it, but in principle, if one is very patient, it should be possible.

And the paradoxes led some other mathematicians to distrust arguments in words and flee into formalism. The paradoxes led to increased interest in developing symbolic logic, in using artificial formal languages instead of natural languages to do mathematics. The Italian

---

[3] Why not? Well, to repeat, Russell's version is along the lines of "the first positive integer that can't be named in less than a billion words" and Berry's version is "the first transfinite Cantor ordinal that can't be named in a finite number of words". First of all, in the Russell version for the first time we look at **precisely how long a text it takes to specify something** (which is close to how large a program it takes to specify it via a computation, which is program-size complexity). The Berry version is just based on the fact that there are a countable ($\aleph_0$) infinity of English texts, but uncountably many transfinite ordinals. So Russell looks at the exact size of a text, while Berry just cares if it's finite or not. Second, Russell is looking at the descriptive complexity of integers, which are relatively down-to-earth objects that you can have on the computer, while Berry is looking at **extremely big** transfinite ordinals, which are much more theological objects, they're totally nonconstructive. In particular, Berry's ordinals are **much bigger** than all the ordinals that I showed you in the previous section, which we certainly named in a finite number of words... I hope this explanation is helpful!

logician G. Peano went particularly far in this direction. And Russell and A.N. Whitehead in their monumental 3-volume *Principia Mathematica,* in attempting to follow Peano's lead, took an entire volume to deduce that $1 + 1$ is equal to 2! They broke the argument into such tiny steps that a volume of symbols and words was necessary to show that $1 + 1 = 2$![4] A magnificent try, but considered by most people to be an unsuccessful one, for a number of reasons.

At this point Hilbert enters the scene, with a dramatic proposal for a "final solution." What was Hilbert's proposal? And how could it satisfy everyone?

Hilbert had a two-pronged proposal to save the day. First, he said, let's go all the way with the axiomatic method and with mathematical formalism. Let's eliminate from mathematics all the uncertainties and ambiguities of natural languages and of intuitive reasoning. Let's create an artificial language for doing mathematics in which the rules of the game are so precise, so complete, that there is absolutely no uncertainty whether a proof is correct. In fact, he said, it should be completely mechanical to check whether a proof obeys the rules, because these rules should be completely syntactic or structural, they should not depend on the semantics or the meaning of mathematical assertions! In other words—words that Hilbert didn't use, but that we can use now—there should be a **proof-checking algorithm**, a computer program for checking whether or not a proof is correct.

That was to be the first step, to agree on the axioms—principles accepted without proof—and on the rules of inference—methods for deducing consequences (theorems) from these axioms—for **all** of mathematics. And to spell out the rules of the game in excruciatingly clear and explicit detail, leaving nothing to the imagination.

By the way, why are the axioms accepted without proof? The traditional answer is, because they are self-evident. I believe that a better answer is, because you have to stop somewhere to avoid an infinite regress!

What was the second prong of Hilbert's proposal?

It was that he would include unsafe, non-constructive reasoning

[4]They defined numbers in terms of sets, and sets in terms of logic, so it took them a long time to get to numbers.

in his formal axiomatic system for all of mathematics, like existence proofs via *reductio ad absurdum.* But, then, using intuitive, informal, safe, constructive reasoning **outside** the formal system, he would prove to Brouwer that the unsafe traditional methods of reasoning Hilbert allowed in his formal axiomatic system could never lead to trouble!

In other words, Hilbert simultaneously envisioned a complete formalization of all of mathematics as a way of removing all uncertainties, and as a way of convincing his opponents using their own methods of reasoning that Hilbert's methods of reasoning could never lead to disaster!

So Hilbert's program or plan was extremely ambitious. It may seem mad to entomb all of mathematics in a formal system, to cast it in concrete. But Hilbert was just following the axiomatic formal tendency in mathematics and taking advantage of all the work on symbolic logic, on reducing reasoning to calculation. And the key point is that once a branch of mathematics has been formalized, then it becomes a fit subject for **metamathematical** investigation. For then it becomes a combinatorial object, a set of rules for playing with combinations of symbols, and we can use mathematical methods to study what it can and cannot achieve.

This, I think, was the main point of Hilbert's program. I'm sure he didn't think that **"mathematics is a meaningless game played with marks of ink on paper"**; this was a distortion of his views. I'm sure he didn't think that in their normal everyday work mathematicians should get involved in the minutiae of symbolic logic, in the tedium of spelling out **every** little step of a proof. But once a branch of mathematics is formalized, once it is desiccated and dissected, then you can put it under a mathematical microscope and begin to analyze it.

This was indeed a magnificent vision! Formalize all of mathematics. Convince his opponents with their own methods of reasoning to accept his! How grand!... The only problem with this fantastic scheme, which most mathematicians would probably have been happy to see succeed, is that it turned out to be impossible to do. In fact, in the 1930s K. Gödel and A.M. Turing showed that it was impossible to formalize all of mathematics. Why? Because essentially **any** formal axiomatic system is either inconsistent or incomplete.

Inconsistency and incompleteness sound bad, but what exactly do they mean? Well, here are the definitions that I use. "**Inconsistent**" means proves false theorems, and "**incomplete**" means doesn't prove all true theorems. (For reasons that seemed pressing at the time, Hilbert, Gödel and Turing used somewhat different definitions. Their definitions are syntactic, mine are semantical.)

What a catastrophe! If mathematics can't be formalized, if no finite set of axioms will suffice, where does that leave mathematical certainty? What becomes of mathematical truth? Everything is uncertain, everything is left up in the air!

Now I'm going to tell you how Gödel and Turing arrived at this astonishing conclusion. Their methods were very different.

Gödel's Incompleteness Theorem

How did Gödel do it? Well, the first step, which required a tremendous amount of imagination, was to guess that perhaps Hilbert was completely wrong, that the conventional view of mathematics might be fatally flawed. John von Neumann, a very brilliant colleague of Gödel's, admired him very much for that, for it had never occurred to von Neumann that Hilbert could be mistaken![5]

Gödel began with the liar paradox, "**This statement is false!**" If it's true, then it's false. If it's false, then it's true. So it can neither be true nor false, which is not allowed in mathematics. As long as we leave it like this, there's not much we can do with it.

But, Gödel said, let's change things a little. Let's consider "**This statement is unprovable!**" It's understood that this means in a particular formal axiomatic system, from a particular set of axioms, using a particular set of rules of inference. That's the context for this statement.

Well, there are two possibilities. Either this statement is a theorem, is provable, or it isn't provable, it's not a theorem. Let's consider the two cases.

What if Gödel's statement is provable? Well, since it affirms that it itself is unprovable, then it's false, it does not correspond with reality.

---

[5] My source for this information is Ulam's autobiography.

lecture on "Logic and the understanding of nature." As is touchingly described by Hilbert's biographer Constance Reid, this was the grand finale of Hilbert's career and his last major public appearance. Hilbert's lecture ended with his famous words: *"Wir müssen wissen. Wir werden wissen."* We must know! We shall know!

Hilbert had just retired, and was an extremely distinguished emeritus professor, and Gödel was a twenty-something unknown. They did not speak to each other then, or ever. (Later I was luckier than Gödel was with Hilbert, for I at least got to talk with Gödel on the phone! This time I was the twenty-something unknown and **he** was the famous one.[7])

But the general reaction to Gödel, once the message sank in, was shock! How was it possible!? Where did this leave mathematics? What happens to the absolute certainty that mathematics is supposed to provide? If we can never have all the axioms, then we can never be sure of things. And if we try adding new axioms, since there are no guarantees and the new axioms may be false, then math becomes like physics, which is tentative and subject to revision! If the fundamental axioms change, then mathematical truth is time dependent, not perfect, static and eternal the way we thought!

Here is the reaction of the well-known mathematician Hermann Weyl: "[W]e are less certain than ever about the ultimate foundations of (logic and) mathematics...we have our 'crisis'...it directed my interests to fields I considered relatively 'safe,' and has been a constant drain on the enthusiasm and determination with which I pursued my research work."

But with time a funny thing happened. People noticed that in their normal everyday work as mathematicians you don't really find results that state that they themselves are unprovable. And so mathematicians carried on their work as before, ignoring Gödel. The places where you get into trouble seemed too remote, too strange, too atypical to matter.

But only five years after Gödel, Turing found a deeper reason for incompleteness, a different source of incompleteness. Turing derived incompleteness from uncomputability. So now let me tell you about

---

[7]I tell this story in my lecture "The Berry paradox" published in the first issue of *Complexity* magazine in 1995.

that.

Turing's Halting Problem

Turing's remarkable paper of 1936 marks the official beginning of the computer era. Turing was the first computer scientist, and he was not just a theoretician. He worked on **everything**, computer hardware, artificial intelligence, numerical analysis...

The first thing that Turing did in his paper was to invent the general-purpose programmable digital computer. He did it by inventing a toy computer, a mathematical model of a computer called the Turing machine, not by building actual hardware (though he worked on that later). But it's fair to say that the computer was invented by the English mathematician/logician Alan Turing in the 1930s, years before they were actually built, in order to help clarify the foundations of mathematics. Of course there were many other sources of invention leading to the computer; history is always very complicated. That Turing deserves the credit is as true, or truer, than many other historical "truths."

(One of the complications is that Turing wasn't the only inventor of what is now called the Turing machine. Emil Post came up with similar ideas independently, a fact known only to specialists.)

How does Turing explain the idea of a digital computer? Well, according to Turing the computer is a very flexible machine, it's soft hardware, it's a machine that can simulate any other machine, if it's provided with a description of the other machine. Up to then computing machines had to be rewired in order to undertake different tasks, but Turing saw clearly that this was unnecessary.

Turing's key idea is his notion of a **universal** digital machine. I'll have much more to say about this key notion of "computational universality" later. Now let's move on to the next big contribution of Turing's 1936 paper, his discussion of the halting problem.

What is the halting problem? Well, now that Turing had invented the computer, he immediately asked if there was something that can't be done using a computer, something that no computer can do. And he found it right away. There is no algorithm, no mechanical procedure, no computer program that can determine in advance if **another** computer program will ever halt. The idea is that before running a program P, in order to be sure that P will eventually stop, it would be nice to be able to give P to a halting problem program H. H decides whether P will halt or not. If H says that P halts, then we run P. Otherwise, we don't.

Why, you may ask, is there a problem? Just run the program P and see if it halts. Well yes, it's easy to decide if a program halts in a fixed amount of time by running it for that amount of time. And if it does halt, eventually we can discover that. The problem is how to decide that it never halts. You can run P for a million years and give up and decide that it will never halt just five minutes before it was going to!

(Since there's no time limit the halting problem is a theoretical problem, not a practical problem. But it's also a very concrete, down-to-earth problem in a way, because we're just trying to predict if a machine will eventually do something, if something eventually happens. So it's almost like a problem in physics!)

Well, it would be nice to have a way to avoid running bad programs that get stuck in a loop. But here is Turing's proof that there's no way to do it, that it's uncomputable.

The proof, which I'll give in detail in Chapter IV in LISP, will be a *reductio ad absurdum.* Let's assume that we have a way to solve the halting problem. Let's assume that we have a subroutine H that can take any program P as input, and that H returns "will halt" or "never halts" and always gets it right.

Then here's how we get into trouble with this halting problem subroutine H. We put together a computer program P that's self-referential, that calculates itself. We'll do this by using the same self-reference trick that I use in Chapter III to prove Gödel's theorem. Once this program P has calculated itself, P uses the halting problem subroutine H to decide if P halts. Then, just for the heck of it, P does the **opposite** of what H predicted. If H said that P would halt, then P goes into an infinite loop, and if H said that P wouldn't halt, then P immediately halts. And we have a contradiction, which shows that the halting problem subroutine H cannot exist.

And that's Turing's proof that something very simple is uncomputable. The trick is just self-reference—it's like Russell's set of all sets

Springer Series in Discrete Mathematics and Theoretical Computer Science

The UNKNOWABLE

This essential companion volume to CHAITIN's highly successful "The Limits of Mathematics", also published by Springer, gives a brilliant historical survey of the work this century on the foundations of mathematics, in which the author was a major participant. "The Unknowable" is a very readable and concrete introduction to CHAITIN's ideas, and it includes a detailed explanation of the programming language used by CHAITIN in both volumes. It will enable computer users to interact with the author's proofs and discover for themselves how they work. The software for "The Unknowable" can be downloaded from the author's Web site.

"Greg CHAITIN's new book, 'The Unknowable', is a welcome addition to his oeuvre. In it he manages to bring his amazingly seminal insights to the attention of a much larger audience and to place them into historical context. His work has deserved such treatment for a long time."
John Allen PAULOS, author of "Once Upon a Number"

"A 'prequel' to 'The Limits of Mathematics'; introduces metamathematical concepts from Cantor, Hilbert, Gödel, Turing and the author's algorithmic information theory. Written in a very straight-forward manner, CHAITIN's goal with 'The Unknowable' is to describe the simple ideas at the core of undecidability. He does an admirable job."
Michael D. SOFKA, Rensselaer Polytechnic Institute

ISBN 981-4021-72-5

9 789814 021722 >

http://www.springer.com.sg