

# **THE LOGIC IN COMPUTER SCIENCE COLUMN**

**BY**

**YURI GUREVICH**

Computer Science and Engineering  
University of Michigan, Ann Arbor, MI 48109, USA  
gurevich@umich.edu

# INCOMPLETENESS EX MACHINA

Sebastian Oberhoff

oberhoff.sebastian@gmail.com

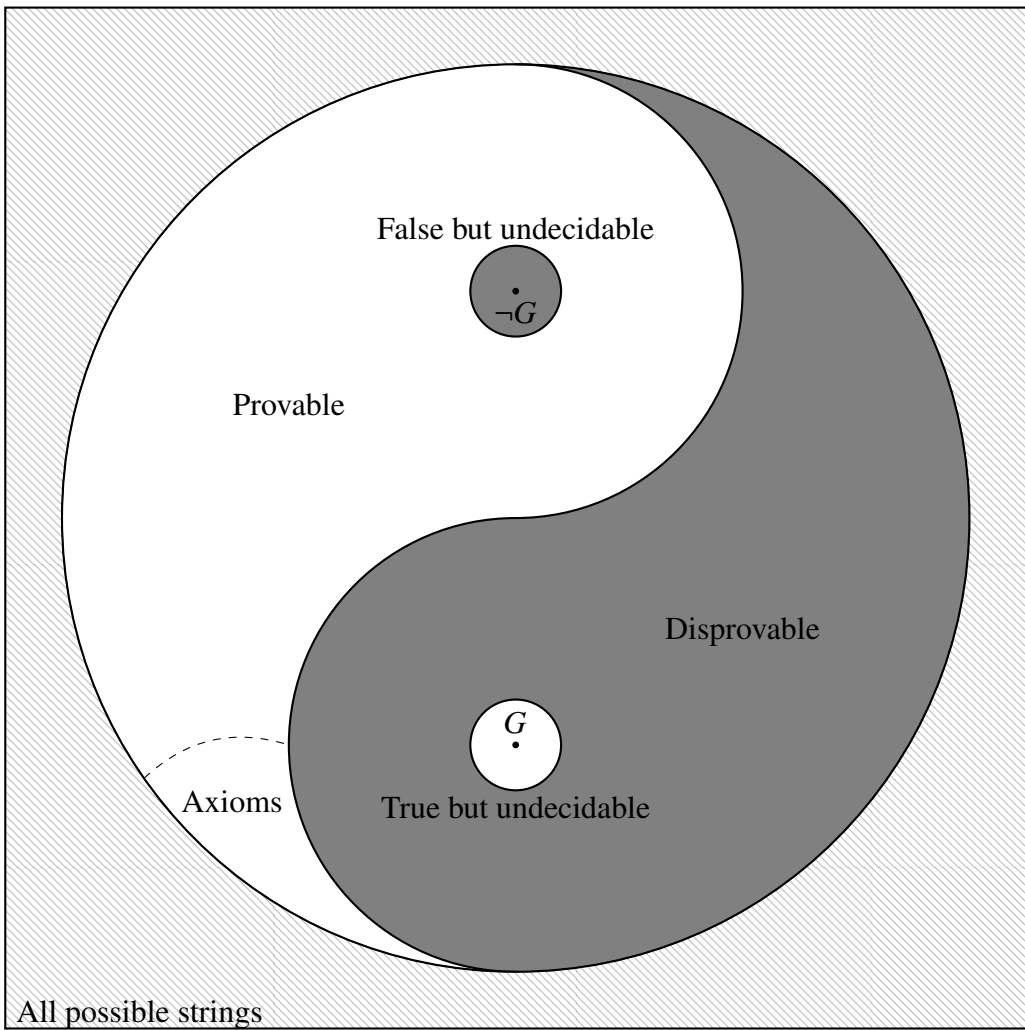
## Abstract

In this essay we'll prove Gödel's incompleteness theorems twice. First, we'll prove them the good old-fashioned way. Then we'll repeat the feat in the setting of computation. In the process we'll discover that Gödel's work, rightly viewed, needs to be split into two parts: the transport of computation into the arena of arithmetic on the one hand and the actual incompleteness theorems on the other. After we're done there will be cake.

It is a profoundly erroneous truism, repeated by all copy-books and by eminent people when they are making speeches, that we should cultivate the habit of thinking of what we are doing. The precise opposite is the case. Civilization advances by extending the number of important operations which we can perform without thinking about them. Operations of thought are like cavalry charges in a battle—they are strictly limited in number, they require fresh horses, and must only be made at decisive moments.

---

*Alfred North Whitehead*



# 1 The Incompleteness Theorems

## On Fast-Forward

Kurt Gödel's incompleteness theorems are clearly the most significant results in the history of mathematics (fight me). The first establishes that no single "proper" formal system can fully settle all mathematical questions; that truth and provability are distinct concepts.<sup>1</sup> The second shows that such a formal system also can't prove itself free of contradiction. If it tries, its wings will melt and it will crash to the ground.

Gödel accomplished these triumphs as follows. Suppose we're given a formal system<sup>2</sup>  $\mathcal{F}$  that is capable of reasoning about elementary arithmetic. Then, as Gödel showed through a lot of toil [1], it is possible to construct a sentence  $G$  which essentially says "I am not provable in  $\mathcal{F}$ ". Once he had built this sentence he then simply asked: can  $G$  be proven or disproven—that is *decided*—in  $\mathcal{F}$ ?

**Suppose  $G$  is provable:**

**Either  $\neg G$  is also provable:**

$\Rightarrow$  Both  $G$  and  $\neg G$  are provable in  $\mathcal{F}$ . Any system in which such a situation arises is also called *inconsistent*.

**Or  $\neg G$  isn't provable:**

$\Rightarrow$  Since  $G$  is provable there exists some concrete proof of  $G$  which we can write out *inside*  $\mathcal{F}$ . This is done using a device called Gödel numbering which allows numbers to talk about proofs. It's complicated.

$\Rightarrow$  Because this proof leads to  $G$  this allows us to construct a new proof of « $G$  is provable» =  $\neg G$ .<sup>3</sup> But we're assuming  $\neg G$  isn't provable. We have arrived at a genuine contradiction, not just an inconsistency. So this can't happen. We'll use the shorthand " $\zeta$ " for this in the future.

**Suppose  $\neg G$  is provable:**

This is what it means to disprove  $G$ . But beware:  $\neg G$  isn't a self-referential sentence. In particular, it's not equal to « $\neg G$  isn't provable». Instead, it's « $G$  is provable». These are very different sentences.  $\neg G$  is really more like an evil twin that's telling lies about its sibling (assuming the system  $\mathcal{F}$  as a

---

<sup>1</sup>Note to the veteran reader: this entire presentation is focused on the standard interpretation of arithmetic. No funky transfinite numbers.

<sup>2</sup>For the purpose of this discussion every formal system is effectively axiomatized by definition. This basically just boils down to the fact that proofs are computer checkable.

<sup>3</sup>Whereas regular quotes (" ") perform their usual function, guillemets (« ») surround formal sentences denoted by their English description. Also "=" doesn't always have to mean exact equality. I also use it to relate sentences which are merely logically equivalent, provided this is obvious.

whole is honest). Furthermore, mind the fact that this case and the previous aren't logical opposites. That's where I use either/or.

**Either  $G$  is also provable:**

$\Rightarrow \mathcal{F}$  is inconsistent.

**Or  $G$  isn't provable:**

$\Rightarrow$  Since  $\neg G$  is provable there exists some concrete proof of  $\neg G$  which we can write out *inside*  $\mathcal{F}$ .

$\Rightarrow$  But in order to force a contradiction we'll have to find a proof of  $G = \langle\langle G \text{ isn't provable} \rangle\rangle$ , not  $\langle\langle \neg G \text{ is provable} \rangle\rangle$  (uh-oh).

$\Rightarrow$  We can't find a proof of  $G$  despite the fact that its existence is exactly what  $\neg G = \langle\langle G \text{ is provable} \rangle\rangle$  asserts.

$\Rightarrow$  We haven't convicted  $\mathcal{F}$  of an outright inconsistency. But something is still very, very wrong with it.

These deductions, as well as the ones in the proofs to follow, are deliberately written in such a painstaking style in order to make the similarities between the first and second set of proofs as plain as possible. These are the decisive moments of our battle.

Now, it would be nice to say that we've proven that, if our formal system  $\mathcal{F}$  is consistent, then  $G$  is undecidable. Unfortunately our second case wasn't strong enough. So, just as Gödel did in 1931, we'll have to settle for a weaker formulation:

**Theorem 1** (First Incompleteness Theorem—Original Version). *Let  $\mathcal{F}$  be an **honest**<sup>4</sup> formal system capable of reasoning about elementary arithmetic. Then  $\mathcal{F}$  is incomplete; it contains a sentence that can neither be proven nor disproven in  $\mathcal{F}$ .*

*Proof.* We just did that. □

Okay, I pulled a fast one there. Crucially, I skipped over a rather important concept: *honesty*. What *exactly* do I mean by that?

As the name suggests, an honest formal system is a formal system that speaks the truth. We can believe the claims it makes. In particular, if our  $\mathcal{F}$  is honest and proves  $\neg G = \langle\langle G \text{ is provable} \rangle\rangle$ , then, just as  $\neg G$  says, it really is the case that  $G$  is provable. This is how we can force a contradiction out of the second case.

But honesty is a stronger assumption than consistency. An honest formal system is always consistent. After all, an inconsistent formal system contains a sentence  $S$  such that both  $S$  and  $\neg S$  is provable. One of these must be a lie. However, a consistent formal system isn't necessarily honest. Here's a completely hypothetical example to illustrate:

---

<sup>4</sup>The traditional term is "*sound*". But when explaining this concept to my mother I had to admit that "honesty" is much more fitting. Also, I've been burned by overloaded usage of the term "soundness" in the past.

Imagine I'm the leader of an idyllic communist utopia. Everything is going according to current five-year plan. Except one day a nuclear reactor of mine suddenly catches fire. If I then go on to tell the world community that there's nothing to worry about, this will merely be a lie, not a contradiction. Even once Swedish scientists begin smelling the fire my excuses, while fabricated, can still retain their *internal* consistency. It's only when I finally confess to my mishap that I contradict myself. At that point I have given two directly contradictory accounts. Now even those without a Geiger counter can tell that I've been dishonest.

Disorienting as it may be, this dishonest-yet-consistent state can also occur in formal systems. Most important for our purposes, it is conceivable that  $\mathcal{F}$  can prove  $\neg G = \langle G \text{ is provable} \rangle$  while  $G$  in fact *isn't* provable. This would lead to a curious situation I'll call *sub-inconsistency*<sup>5</sup> where for all  $n \in \mathbb{N}$  the sentence

$\langle G \text{ doesn't have a proof shorter than } n \rangle$

can be proven in  $\mathcal{F}$  by just going through all possible proofs up to length  $n$  and pointing out that none of them do the trick. Yet the sentence with the "for all" on the inside,

$\langle G \text{ doesn't have a proof shorter than } n \text{ for all } n \in \mathbb{N} \rangle = G,$

is not only unprovable but flat out contradicted by  $\neg G$ . This is why Gödel had to assume honesty (or at least *super-consistency*) in his original proof. Basically, he had to press  $\mathcal{F}$  up against the wall and shout: "No games, dammit!"

Okay, but isn't there a less violent way to prevent this kind of sub-terfuge? Indeed there is. In 1936 J. Barkley Rosser found a way [2] to demonstrate the improved

**Theorem 2** (First Incompleteness Theorem—Rosser's Version). *Let  $\mathcal{F}$  be a **consistent** formal system capable of reasoning about elementary arithmetic. Then  $\mathcal{F}$  is incomplete.*

*Proof.* In hindsight Rosser's trick is quite simple. Instead of formalizing "I am not provable" he formalized the sentence "For every proof of me there exists a shorter disproof". Let's refer to this sentence as  $R$  and see how  $R$  can help us.

**Suppose  $R$  is provable:**

$\Rightarrow$  There exists some shortest proof  $p$  of  $R$ .

**Either there is a proof of  $\neg R$  that's shorter than  $p$ :**

$\Rightarrow \mathcal{F}$  is inconsistent. <sup>6</sup>

**Or there isn't:**

---

<sup>5</sup>Also known under the name " *$\omega$ -inconsistency*". Let's spare our horses.

<sup>6</sup>Inconsistencies can be promoted to genuine contradictions if they collide with the assumption of consistency.

- ⇒ We can go through all strings shorter than  $p$  and determine that none of them are a proof of  $\neg R$ .
- ⇒ We can write out this list *inside*  $\mathcal{F}$  and point out that all longer strings are at least as long as  $p$ .
- ⇒ This allows us to construct a new (longer) proof of «There exists a proof of  $R$  with no shorter disproof (namely  $p$ )» =  $\neg R$ .
- ⇒  $\mathcal{F}$  is inconsistent.  $\zeta$

**Suppose  $\neg R$  is provable:**

- ⇒ There exists some shortest proof  $p$  of  $\neg R$ .

**Either there is a proof of  $R$  that's shorter than  $p$ :**

- ⇒  $\mathcal{F}$  is inconsistent.  $\zeta$

**Or there isn't:**

- ⇒ We can go through all strings shorter than  $p$  and determine that none of them are a proof of  $R$ .
- ⇒ We can write out this list *inside*  $\mathcal{F}$  and point out that all longer strings are at least as long as  $p$ .
- ⇒ This allows us to construct a new (longer) proof of «For every proof of  $R$  there exists a shorter disproof (namely  $p$ )» =  $R$ .
- ⇒  $\mathcal{F}$  is inconsistent.  $\zeta$  □

Whereas Gödel's original proof was limping on one leg, Rosser's version is perfectly balanced, as all things should be, yielding the stronger result.

Finally, there's one more summit to conquer in this mathematical Himalaya. That's the Second Incompleteness Theorem.

**Theorem 3** (Second Incompleteness Theorem). *Again, let  $\mathcal{F}$  be a consistent formal system capable of reasoning about elementary arithmetic. Then  $\mathcal{F}$  can't prove its own consistency.*

Perhaps the fact that this theorem speaks of consistency, not honesty, makes you a little suspicious. Is this another one of Rosser's upgrades? No. For the Second Incompleteness Theorem Gödel only needed half of his First Incompleteness Theorem. And, as luck would have it, this was exactly the half that spoke of consistency.

*Proof.* Let's assume for the purpose of contradiction that  $\mathcal{F}$  can prove its own consistency

- ⇒ The first half of the proof for the First Incompleteness Theorem can be read as a proof by contradiction that if  $\mathcal{F}$  is consistent, then  $G$  isn't provable.
- ⇒ Using the fact that « $G$  isn't provable» =  $G$  this can be stated more concisely as " $\mathcal{F}$  is consistent  $\implies G$ ".

- ⇒ As Gödel himself pointed out with some hand-waving, there wasn't actually anything in the reasoning we used to prove this that we couldn't carry out just as well in  $\mathcal{F}$ . So « $\mathcal{F}$  is consistent  $\implies G$ » is actually a theorem in  $\mathcal{F}$ .
- ⇒ We're assuming that  $\mathcal{F}$  can prove « $\mathcal{F}$  is consistent». So by modus ponens<sup>7</sup>  $\mathcal{F}$  can also prove  $G$ .
- ⇒ But  $G$  can't be proven. ⚡ □

Moral of the story: if you declare yourself a very stable genius, you're not [3].

## 2 The Rise Of The Machines

We now come to the dramatic turn in our story. Namely, we're about to demand that our formal system  $\mathcal{F}$  can reason about algorithms, not just elementary arithmetic. And then we'll reprove all three of the previous results on top of that.

What could possibly provoke such heresy? Well, you may recall that in the prequel we relied quite heavily on the fact that both  $G$  and  $R$  could *somehow* be constructed in  $\mathcal{F}$ . But we never actually performed these constructions. We also liberally made assertions of the form: “we can list out all proofs up to length  $n$  inside  $\mathcal{F}$ ”. But  $\mathcal{F}$  is about numbers, not formal sentences. How is all of this ultimately accomplished?

The reality is that I hand-waved these details for a reason. They are non-trivial in the extreme (ideal exercises for the reader!). That's because, from a modern perspective, Gödel basically had to teach programming to formal systems that were about arithmetic. And he had to do it half a decade before the first models of computation were even devised. This is where all the tough technical work had to be done. The incompleteness theorems themselves were mere victory laps at the end.

This delimitation between the different parts of Gödel's work is frequently passed over without comment but will become even more apparent as we move on. In fact, setting the record straight on this matter is really the main point of this essay. By placing computation—for example in the form of Turing machines—directly into the bedrock of our formal system our task will become orders of magnitude simpler. Gödel had to scale a vertical cliff to convince his peers that “ $p$  is a proof of  $S$ ” could be expressed within the confines of arithmetic. We, on the other hand, aided by the ski lift built by Turing and his apostles, can see at a glance that the same could be checked by a computer. Let's exploit that.

Anybody who still harbors a nostalgic longing for “elementary arithmetic” can then take on the separate task of showing how to ponder computation in the realm

---

<sup>7</sup>Modus ponens is the inference rule that from  $A$  and « $A \implies B$ » follows  $B$ .



of the natural numbers. Perhaps one can also study computation starting from a different base camp such as knot theory. Who knows? (I'm completely clueless regarding knot theory.) Meanwhile, we will have the incompleteness theorems above the clouds all to ourselves.

### 3 To Halt Or Not To Halt

Alright, here we go. The only ingredient we need to prepare is the unsolvability of the Halting Problem due to Alan Turing in 1936 [4].

**Lemma 1** (Unsolvability Of The Halting Problem). *There's no algorithm that implements the function*

$$A \mapsto \begin{cases} \text{accept} & \text{if } A(A) \text{ halts} \\ \text{reject} & \text{if } A(A) \text{ doesn't halt.} \end{cases}$$

(“ $A(A)$ ” denotes an algorithm  $A$  running on its own source code.)

*Proof.* Suppose  $H$  solves the Halting Problem. Then one could create the following algorithm:

---

```

 $\overline{H}(A)$ 
  if  $H(A)$  rejects
    halt
  if  $H(A)$  accepts
    loop

```

---

“halt” just... halts. We don't even need to bother accepting or rejecting. It doesn't matter. And “loop” is a simple infinite loop without exits. Now, question: does  $\overline{H}(\overline{H})$  halt?

**Either  $\overline{H}(\overline{H})$  halts:**

$\Rightarrow H(\overline{H})$  accepts.  
 $\Rightarrow \overline{H}(\overline{H})$  doesn't halt.  $\zeta$

**Or  $\overline{H}(\overline{H})$  doesn't halt:**

$\Rightarrow H(\overline{H})$  rejects.  
 $\Rightarrow \overline{H}(\overline{H})$  halts.  $\zeta$

Either way we get a contradiction. □

Easy peasy. As you'd expect of a mere lemma. We're now ready for the First Incompleteness Theorem.

**Theorem 4** (First Incompleteness Theorem—Original Version By Computation). *Let  $\mathcal{F}$  be an honest Turing complete formal system. Then  $\mathcal{F}$  is incomplete.*

For a formal system to be *Turing complete* simply means that it can reason about algorithms. And our arguments won't be overly technical anyway. So let me just leave it there.

*Proof.* Suppose  $\mathcal{F}$  was complete. Then we could solve the Halting Problem using the following algorithm:

---

$H(A)$   
**for**  $x \in$  all possible strings<sup>8</sup>  
    **if**  $x$  proves « $A(A)$  doesn't halt»  
        reject  
    **if**  $x$  proves « $A(A)$  halts»  
        accept

---

Because  $\mathcal{F}$  is complete this search will eventually hit upon a proof of either « $A(A)$  halts» or « $A(A)$  doesn't halt». And because  $\mathcal{F}$  is honest we'll be able to trust its judgment. That determines whether or not  $A(A)$  halts. But, as we saw only a moment ago, it's impossible to solve the Halting Problem. So  $\mathcal{F}$  couldn't have been complete.  $\zeta$  □

That certainly went by a lot faster than my last reading of *Gödel, Escher, Bach* [5]. Though, if you've ever had the incompleteness theorems explained to you by a computer scientist, then you probably saw this proof coming a mile away. And for good reason. It's a very elegant little proof; deserving of its popularity.

## 4 Welcome To The World Of Mirrors

Nevertheless, this isn't the proof I want to use going forward. The reasons are threefold:

- The jump from honesty to consistency is rather tricky.
- The Second Incompleteness Theorem doesn't follow naturally at all.
- The parallels to Gödel's and Rosser's original proofs are lost. (This one's the biggie.)

We can do better. Watch this:

*Second Proof Of Theorem 4.* Consider the following algorithm:

---

$P(A)$   
**for**  $x \in$  all possible strings  
    **if**  $x$  proves « $A(A)$  doesn't halt»  
        halt

---

---

<sup>8</sup>In case it wasn't clear: we're iterating in lexicographical order. This procedure is sometimes also jokingly referred to as the *British Museum algorithm* because it's akin to using chimps in front of typewriters in an attempt to reproduce all the books in the British Museum.

I now claim that « $P(P)$  doesn't halt» is undecidable in  $\mathcal{F}$ . Let's call this sentence  $\overset{\infty}{G}$  because it's the perfect analogue of Gödel's  $G$ . The critical new feature of  $\overset{\infty}{G}$  is that its construction has become just a small matter of programming. If we go through the alternatives, we find:

**Suppose  $\overset{\infty}{G}$  is provable:**

**Either  $\neg\overset{\infty}{G}$  is also provable:**

$\Rightarrow \mathcal{F}$  is inconsistent.  $\zeta$

**Or  $\neg\overset{\infty}{G}$  isn't provable:**

$\Rightarrow$  Since  $\overset{\infty}{G}$  is provable  $P(P)$  will eventually find such a proof after some sequence of steps which we can write out inside  $\mathcal{F}$ .

$\Rightarrow$  Because this sequence leads to a terminal state this allows us to construct a new proof of « $P(P)$  halts» =  $\neg\overset{\infty}{G}$ .  $\zeta$

**Suppose  $\neg\overset{\infty}{G}$  is provable:**

**Either  $\overset{\infty}{G}$  is also provable:**

$\Rightarrow \mathcal{F}$  is inconsistent.  $\zeta$

**Or  $\overset{\infty}{G}$  isn't provable:**

$\Rightarrow$  Since  $\neg\overset{\infty}{G}$  is provable  $P(P)$  will eventually find such a proof after some sequence of steps which we can write out inside  $\mathcal{F}$ .

$\Rightarrow$  But in order to halt  $P(P)$  will have to find a proof of  $\overset{\infty}{G}$  = « $P(P)$  doesn't halt», not  $\neg\overset{\infty}{G}$  (uh-oh).

$\Rightarrow$   $P(P)$  doesn't halt despite the fact that this is exactly what  $\neg\overset{\infty}{G}$  = « $P(P)$  halts» asserts.

$\Rightarrow \mathcal{F}$  is dishonest.  $\zeta$  □

Is anybody else experiencing déjà vu? Note that, whereas in the previous proof we only established the *existence* of undecidable sentences in  $\mathcal{F}$ , here we actually have a concrete undecidable sentence on our hands. And we didn't even need the Halting Problem!

Moreover, almost miraculously, the two cases for  $\overset{\infty}{G}$ 's provability run into the exact same problems as  $G$ . In the first case  $\mathcal{F}$  plants its  $\mathcal{F}$ ace squarely in inconsistency. In the second one the precise correlation between the information  $\mathcal{F}$  communicated and the facts insofar as they can be determined and demonstrated is such as to cause epistemological problems of sufficient magnitude as to lay upon the logical and semantic resources of the English language a heavier burden than they can reasonably be expected to bear. It told a lie [6].

Upon closer inspection we notice that we're even dealing with the same kind of lie:

⌘ sub-inconsistency ⌘

After all, if  $\overset{\infty}{G}$  isn't provable, then one can still prove that  $P(P)$  hasn't halted after  $n$  steps for all  $n \in \mathbb{N}$  by just writing out  $n$  steps of the computation in  $\mathcal{F}$  and

noting that it's still running. But the summarizing sentence « $P(P)$  hasn't halted after  $n$  steps for all  $n \in \mathbb{N}$ » =  $\bar{G}$  is mysteriously unprovable. If  $\mathcal{F}$  then goes on to prove  $\neg \bar{G}$  = « $P(P)$  halts», this will be a false promise. Yet we're unable to catch it red-handed.

And if you thought that was neat, just wait until you see Rosser's trick. As you may recall, this one worked by patching the asymmetry between the two cases in Gödel's original argument. Let's see what happens if we try the same here.

**Theorem 5** (First Incompleteness Theorem—Rosser's Version By Computation). *Let  $\mathcal{F}$  be a consistent Turing complete formal system. Then  $\mathcal{F}$  is incomplete.*

*Proof.* Consider the following algorithm:

---

$B(A)$ <sup>9</sup>  
**for**  $x \in$  all possible strings  
    **if**  $x$  proves « $A(A)$  doesn't halt»  
        halt  
    **if**  $x$  proves « $A(A)$  halts»  
        loop

---

In a sense we've seen this algorithm before. Just take  $H$  as it's defined in the first proof of Theorem 4, substitute it into  $\bar{H}$ , and simplify. It's  $\mathcal{F}$ 's most valiant attempt to implement  $\bar{H}$ . This time our undecidable sentence is (you guessed it)  $\bar{R}$  = « $B(B)$  doesn't halt».

**Suppose  $\bar{R}$  is provable:**

⇒ There exists some shortest proof  $p$  of  $\bar{R}$ .

**Either there is a proof of  $\neg \bar{R}$  that's shorter than  $p$ :**

⇒  $\mathcal{F}$  is inconsistent.  $\zeta$ <sup>10</sup>

**Or there isn't:**

⇒  $B(B)$  will go through all strings shorter than  $p$  and determine that none of them are a proof of  $\neg \bar{R}$ .

⇒ We can write out this computation inside  $\mathcal{F}$  and point out that  $B(B)$  will then find  $p$  and enter a terminal state.

⇒ This allows us to construct a new (longer) proof of « $B(B)$  halts» =  $\neg \bar{R}$ .

⇒  $\mathcal{F}$  is inconsistent.  $\zeta$

**Suppose  $\neg \bar{R}$  is provable:**

⇒ There exists some shortest proof  $p$  of  $\neg \bar{R}$ .

**Either there is a proof of  $\bar{R}$  that's shorter than  $p$ :**

---

<sup>9</sup>I picked the names  $P$  and  $B$  because there's a slight visual resemblance between the letters and the programs they denote. (Also this is why I placed the "doesn't halt" case first.)

<sup>10</sup>I might add that in this case  $B(B)$  never reaches  $p$  but finds the shorter proof first. We could dwell on this. But we already have an inconsistency which is all we need.

⇒  $\mathcal{F}$  is inconsistent.  $\zeta$

**Or there isn't:**

⇒  $B(B)$  will go through all strings shorter than  $p$  and determine that none of them are a proof of  $\check{R}$ .

⇒ We can write out this computation inside  $\mathcal{F}$  and point out that  $B(B)$  will then find  $p$  and enter an infinite loop.<sup>11</sup>

⇒ This allows us to construct a new (longer) proof of « $B(B)$  doesn't halt» =  $\check{R}$ .

⇒  $\mathcal{F}$  is inconsistent.  $\zeta$  □

I feel like I'm watching a reboot (in more than one sense). We can even recognize the phenomenon of finding either shorter or longer disproofs and proofs; pure satisfaction.

Alright, time for the grand finale.

**Theorem 6** (Second Incompleteness Theorem—By Computation). *One last time, let  $\mathcal{F}$  be a consistent Turing complete formal system. Then  $\mathcal{F}$  can't prove its own consistency.*

We could continue using  $B$  for the Second Incompleteness Theorem. But that would be ugly. Gödel didn't need Rosser's help for his proof. Neither shall we.

*Proof.* Let's assume for the purpose of contradiction that  $\mathcal{F}$  can prove its own consistency

⇒ The first half of our new proof for the First Incompleteness Theorem can be read as a proof by contradiction that if  $\mathcal{F}$  is consistent, then  $P(P)$  doesn't halt.

⇒ Using the fact that « $P(P)$  doesn't halt» =  $\check{G}$  this can be stated more concisely as “ $\mathcal{F}$  is consistent  $\implies \check{G}$ ”.

⇒ Now we simply observe (with some faith) that there wasn't actually anything in the reasoning we used to prove this that we couldn't carry out just as well in  $\mathcal{F}$ . So « $\mathcal{F}$  is consistent  $\implies \check{G}$ » is actually a theorem in  $\mathcal{F}$ .

⇒ We're assuming that  $\mathcal{F}$  can prove « $\mathcal{F}$  is consistent». So by modus ponens  $\mathcal{F}$  can also prove  $\check{G}$ .

⇒ But then  $P(P)$  will eventually find such a proof and halt.  $\zeta$  □

---

<sup>11</sup>Proving that a program runs forever raises the specter ( $\mathcal{L}$ ) of sub-inconsistency. But rest assured. Any formal system capable of reasoning about algorithms can prove that “while true do nothing” runs forever. Otherwise it ought to be ashamed of itself.

## 5 Taking Off The Glasses

At this point I can't even tell the difference between the classical approach and the algorithmic retelling anymore. Can you? All these proofs are exactly the same as before! In fact, the relation is so strong that I was able to write the latter ones by copy-pasting wholesale and then tinkering a little. Just line them up and see for yourself. Every single line matches. Clearly, this can't be coincidence. There must be a deeper reason for this overabundance of similarities. And there is. Perhaps you've already caught on long ago. Or perhaps Clark Kent's glasses fooled you too. Drum whirl please.

Recall once more what  $\tilde{G}$  says. It says " $P(P)$  doesn't halt". But  $P(P)$  isn't just executing some simple infinite loop. It's looking for something. It's scouring the proofs in  $\mathcal{F}$  for a proof of  $\langle P(P) \text{ doesn't halt} \rangle = \tilde{G}$ .  $\tilde{G}$ 's claim that  $P(P)$  will never find a proof of  $\tilde{G}$  is echoing  $G$ 's claim that the original computers, *you and me*, will never find a proof of  $G$ . In other words,  $\tilde{G}$  is also proclaiming to the world:

"I am not provable!"

Exercise: Take off  $\tilde{R}$ 's glasses as well.

This means I could've organized this presentation very differently. I could've just made these observations right away and then said: "Now reread the first section.  $\square$ ". I've been wasting your time!

## 6 Refactoring The Incompleteness Theorems

Gödel was programming the integers. That is that. And he didn't even know it at the time; truly impressive. I think this feat deserves to be encapsulated with its own theorem. Perhaps we can come together and save posterity from the grim fate of also only realizing this much later in their lives. It's for the kids—and for the horses.

**Theorem 7.** *Elementary arithmetic is a Turing complete domain.*

I'll have to leave the exact formulation of this theorem to smarter people. After all, I did recently score in the bottom 7% of the GRE's analytical writing section (come on!). The upshot is that this allows us to easily recover the incompleteness theorems for elementary arithmetic from a more general insight:


**Theorem 8** (Turing Incompleteness Theorem). *Any consistent formal system  $\mathcal{F}$  capable of reasoning about a Turing complete domain can reason about algorithms. That makes  $\mathcal{F}$  a consistent Turing complete formal system. Thereby, the incompleteness theorems apply to it. In slogan form:*

*If it's complete, then it's incomplete.*

*Proof.* I think the idea is best conveyed with an example. Take the famously Turing complete Game of Life. For any algorithm one can create a configuration in the Game of Life which mimics that algorithm. Now suppose somebody came up with a formal system  $\mathcal{F}$  to reason about the Game of Life. This formal system might be able to prove assertions such as: «Starting with a configuration of the form [...] square  $(0, 0)$  will never be occupied».

The incompleteness theorems are now immediate. All we have to do is to encode the algorithms  $P$  and  $B$  into a Game of Life configuration and reiterate the previous arguments. Conclusion:  $\mathcal{F}$  will never be able to decide whether the encoding of  $B$  “halts”. And it won't be able to prove its own consistency.  $\square$

As a consequence computational undecidability and logical undecidability go hand in hand. And elementary arithmetic is just one further entry in the long list containing the Game of Life, Fractran, Post Tag Systems, Magic the Gathering, the human brain, Rule 110, the Lambda Calculus, musical notation, ...

(Thank you so much for your attention! The cake I promised in the abstract is on its way. Please stand by.) 

## References

- [1] Kurt Gödel, *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I*, Monatshefte für Mathematik und Physik, Issue 38, Dec. 1931 pp. 173–198
- [2] John Barkley Rosser, *Extensions of Some Theorems of Gödel and Church*, Journal of Symbolic Logic, Volume 1, No. 3, Sep. 1936, pp. 87–91
- [3] Donald Trump, Twitter, 6. Jan. 2018 (I just want to be able to say that Donald Trump has been cited when discussing the fundamentals of mathematics.)
- [4] Alan Turing, *On Computable Numbers, with an Application to the Entscheidungsproblem*, Proceedings of the London Mathematical Society, Volume s2–42, Issue 1, 1937, pp. 230–265
- [5] Douglas Hofstadter, *Gödel, Escher, Bach*, 1979
- [6] Yes, Prime Minister, *The Tangled Web* (it's a British television sitcom from the 80s)
- [7] Algorithm  $B$  also appears in Scott Aaronson's *Quantum Computing since Democritus* where he uses it to give a pseudo proof of the Second Incompleteness Theorem. It was in plugging the holes in Aaronson's argument that these ideas took their origin.