THE HITCHHIKER'S GUIDE TO THE INCOMPLETENESS THEOREM

MINGZHONG CAI

Contents

0. Metaintroduction

This work is a very **informal** guide to the incompleteness theorem, typically for high school students in the grade system on Planet Earth, or correspondingly any creature in **Third Level Mathematics**, according to Himesuku Fujuta's *Handbook of Mathematics in the Galaxy*([1, Chapter 2]).

> *Please read the citations at the beginning of the sections. Please disregard the citations at the end of the sections: they are there just because Suzumiya Haruhi wants them to be there.*
> *(*Kyon's comment on *The Hitchhiker's Guide to the Incompleteness Theorem.)*

Date: March 16, 2009.

This work is supported by NSF Grant DMS 0602193 from the MCTP component of EMSW21.

1. Introduction

> *Incompleteness Theorem: It is also called "Gödel's Incompleteness Theorem" on Planet Earth, or "Nico's Imperfectness Theorem" on Planet Elvia, etc.. After creatures discovered that their versions were virtually the same, they decided to simply call it the Incompleteness Theorem. 99.99% of the sixth level civilizations of our Galaxy had known the incompleteness theorem before they invented Automatic Computing Machines (**ACMs**). The only exception is the Kumalians. They developed **ACMs** 994 Kuma-years before they established logic foundations for them. The key is a special plant named* Hydrakuma Lipsolis *on Planet Kumala. Hydrakuma is known to display certain patterns according to the patterns of other Hydrakumae within a certain range.*
> *(*Incompleteness Theorem, *The Hitchhiker's Guide to the Incompleteness Theorem.)*

You might have heard of this term, but the "Incompleteness Theorem" really sounds weird: what is incomplete? Our very highly developed mathematics, or our foundation logic? Now let's have a look at the first (informal) version of the incompleteness theorem:

Incompleteness Theorem, First Version:
There is a true sentence that mathematics cannot prove.

All right, this is even worse: is it a theorem? It feels more like an assertion in philosophy rather than a real theorem in mathematics. Maybe we should first look at some normal theorems. Here are some examples of theorems, in our usual sense:

Theorem 1. *1+1=2.*

Theorem 2. *There is an even number which is not the sum of two prime numbers.*

Theorem 3. *Two triangles that have equal corresponding sides are congruent.*

But this is not usually recognized as a (mathematical) theorem:

Theorem 4. *There is a cow which is totally white.*

Why? As in the first three examples, we exactly know what the theorems say; in another words, we have clear definitions of the terminologies used in those theorems. However, in the fourth one we might have questions defining "cow", or "white". What is a cow? Adult female bovine. Ok, but what is adult? 1 year old? 1.5 years old? Bovine, what is that? A special kind of animal on Planet Earth? How about if we raise some on Planet Mars? And what is white? Moreover, if I draw a white cow on my paper, is it really a *white cow*?

Let us stop jabbering and get back to our main topic. For the incompleteness theorem we might also ask a lot of questions:

Question 1. *What is a sentence?*

Question 2. *What is a* true *sentence?*

Question 3. *What is mathematics?*

Question 4. *What is a proof?*

In some point of view we really have to define these terms before we have serious discussions about the incompleteness theorem.

2. Sentences

Let us answer this question: What is a (mathematical) sentence? As we see in the previous section, these terms in the sentence should be **defined**. But wait a moment: what is "0"? The first natural number. Then what is "first"? What is a "number" and what is "natural"?

This is bad. One question becomes three. And it is easy to see that we will never stop this process, as we can always ask "what is" questions. So the very first fundamental idea in logic is that we need to leave something undefined, whose meanings should be clear to most people and we can use them freely in any sentence. So maybe we could say "**Zero is Zero**". In another point of view, **Zero** is just a symbol, and anyone can be **Zero** if we define it to be.

If we work with elementary arithmetic, "0" will be a natural **constant symbol**, i.e. it stands for a certain element. Now let us have a look at our first theorem again:

Theorem 1. *1+1=2.*

One can regard "1" and "2" as constant symbols at this time, but what are "+" and "="? + is a **function symbol**, which takes some input and gives an output; = is a **relation symbol**, which also takes some input but will determine whether the input has some relation or not. In this case, it evaluates whether the two parts in the input are the same. Examples of functions are "$-$", "$\times$", "$\div$". Note that we can define subtraction from addition, and define division from multiplication; so usually we will omit subtraction and division, for simplicity. Another important relation symbol is "$<$", and it is easy to see that we can define all other inequality relations from "$<$" and "$=$".

We could regard "1" as a constant symbol, but then we need a lot of symbols for all natural numbers. A classical method is to introduce another function "S", the successor operator. For example, $S0$ is 1, $S1 = SS0$ is 2, etc. But for convenience we will still write numbers as usual, instead of the corresponding number of S's plus 0.

Now we can write a lot of sentences like $1 + 1 = 2$, but still it is not powerful enough to deal with the everyday usage of mathematics. For example, what is the solution set of the inequality $x \times x > 4$? Well, either $x > 2$ or $x < -2$. So we need to introduce **connectives**. There are several connectives used in logic:

$\wedge$**:** means "and".
$\vee$**:** means "or".
$\neg$**:** means "not".
$\Rightarrow$**:** means "imply".
$\Leftrightarrow$**:** means "if and only if".

So the solution for $x \times x > 4$ could be written as $(x > 2) \vee (x < -2)$, or $\neg((x \leq 2) \wedge (x \geq -2))$

Now we are powerful enough to deal with elementary arithmetic, but still something is missing. Let us try to write out the following theorem as a formal sentence:

Theorem 2. *There is an even number which is not the sum of two prime numbers.*

What is a prime? A number which is not divisible by any number less than it, except 1. That is, a number p is prime, if for every $x, y < p$, $x \times y \neq p$. Also, we need to talk about "there is", or "there exists" in the previous theorem. We call those **quantifiers**, i.e., they quantify the objects with some property. In mathematics we usually need the following two quantifiers:

$\exists$**:** means "exist".
$\forall$**:** means "for all".

For example, we can define a predicate for prime number:

$$\mathbf{Prime}(p) := \ \forall a \forall b((a < p \wedge b < p) \Rightarrow (a \times b \neq p))$$

So whenever we have $\mathbf{Prime}(p)$ we replace it by the thing after :=. Then we can write out the theorem in a formal sentence:

$$\exists x(\exists y(2 \times y = x) \ \wedge \ \forall p \forall q((\mathbf{Prime}(p) \wedge \mathbf{Prime}(q)) \Rightarrow (p + q \neq x)))$$

A **formula** is a reasonable sequence of constant symbols, function symbols, relation symbols, connectives, quantifiers and parentheses (parentheses are for readability, instead of introducing complicated priority rules or the Polish notation[1]). Here "reasonable" means that the combination of all these symbols must obey some obvious rules (for example, you can only connect two formulae by connectives, but it doesn't work for variables).

Notice that when we try to quantify a variable x, we need to make sure that x has not yet been quantified, i.e. $\forall x \exists x(x = 0)$ doesn't make sense. We call a variable

[1]Some terms (in this font) might take time and space to explain but they are not crucial to our main subject. See appendix for more information.

free if it is not quantified. So whenever we want to quantify some variable x, we want x to be free in the original formula.

A **sentence** is a formula without free variables. They do make a mathematical assertion and all theorems in mathematics are sentences. From now on we will focus on the formulae and sentences constructed by $(+, \times, =, <, S, 0)$ and $(\wedge, \vee, \neg, \Rightarrow, \Leftrightarrow, \exists, \forall)$, i.e. the language of arithmetic.

> *"A local noncorrosive and agglutinative countertime space was independently created in a modus of some limited conditions."*
> *I waited for a moment, but she didn't say anything. For this kind of sentence constructed by randomly picking words in the dictionary, how can I possibly understand it without a dictionary in hand?*
> *(*Kyon's reaction to a "sentence" of Nagato Yuki. *The Boredom of Suzumiya Haruhi.)*

3. Truth and Worlds

> *Geometry on Planet Viso: Urpes, the sun of Viso is so huge that it distorts the space around it. Visorians have experienced both positive and negative* curvatures *from time to time and have developed highly advanced geometry long before elementary calculus......In 478 A.D. when Visorians historically first contacted humans, dominating creatures on Planet Earth, they strongly opposed the Parallel Postulate in Euclid's Elements and believed that those little creatures would never evolve, especially in mathematics. As a result, they soon stopped their intercelestial teleport service between High Temple on Planet Viso and Delphi on Planet Earth. The next time that humans get in touch with other creatures in the Galaxy is, unfortunately, 4094 earth years later.*
> *(*Geometry on Planet Viso, *The Hitchhiker's Guide to the Incompleteness Theorem.)*

The fundamental idea in Model Theory is that, the **truth** of a sentence really depends on the **world** it is talking about. For example, consider the following sentence:

$$\exists x(Sx = 0),$$

which says that there is an element whose successor is 0. Of course it is false in our usual set of natural numbers $\{0, 1, 2, 3, ...\}$, but if we add -1 to this universe, then the previous sentence is true.

As another example, if we have the following set of numbers as a world:

$$\{0, 1, 2, 3\}$$

and define arithmetic in the usual way, then $1+1=2$ is true there. However, if we have this set but define $+$ in another way, for example, let $0+1=2$ and $1+1=3$, then $1+1=2$ is false there.

A **world** is a group of elements with the constant symbols, function symbols and relation symbols defined. When we talk about the natural numbers, we are actually talking about these elements, plus the constant (0), the functions ($+$,$\times$,S) and the relations ($=$,$<$) defined on these elements.

We say a sentence is **true** in a world if it holds in that world. In some sense, truth is actually defined, or determined by these constants, functions and relations. In another words, $1+1=2$ is true simply because we define it to be.

Of course we want our world to be *logical*, i.e., to obey simple logical rules. For example, if P, Q are true sentences in that world, then $P \wedge Q$ is also true. Any logically valid sentence can be regarded as a logic axiom. See appendix for details.

An interesting thing to note is that there are nice interpretations about sentences with worlds. As an example, if "$P \Rightarrow Q$" is always true in every world, then whenever P is true, Q is also true. One might also regard those interpretations as definitions of those connectives.

From now on, we will only focus on the truth of natural numbers. I.e., For a sentence of arithmetic, we say it is true if it is a true sentence in the world of natural numbers with those constants, functions and relations defined in the common way.

> *"Isn't that created by the Big Bang a long long time ago?"*
> *"That is the common saying. However, there is another possibility for us. This universe actually started three years ago."*
> *(*Kyon and Koizumi Itsuki discussed the origin of the universe. *The Melancholy of Suzumiya Haruhi.)*

4. Axioms of Mathematics

> *First Galaxy Assembly in Mathematics: the first Galaxy Assembly in Mathematics was held on Planet Myau in 5 A.G.P.. For the first time mathematicians from all over the Galaxy gathered together to talk about their research, and more importantly, set up a concrete list of Axioms and Semiaxioms for all creatures. ... Axioms are those sentences which are universally true among all creatures; semiaxioms are "locally true for some creatures, or independent from* ***our*** *knowledge of* ***our*** universe"*([3]). For example, Axiom 1.2.1 says $\forall x(x=x)$ and Semiaxiom B.2.14 says "every set of reals is Lebesgue measurable[2]".*

[2]The pure logical version is too long to write down here, so we have here a simplified version using terms in Fourth Level Mathematics. Its negative form is commonly used on Planet Earth.

(First Galaxy Assembly in Mathematics, *The Hitchhiker's Guide to the Incompleteness Theorem.)*

As we discussed in the previous section, the sentence "$1+1=3$" could be true in some worlds. However, that is not true in our world of common sense, or in another point of view, we really do not care about those worlds where $1+1=3$.

We shall start writing down some basic sentences which are commonly accepted as true sentences and only discuss the worlds where those sentences are true. To repeat, we don't mean that these sentences have to be true, but we do not care about the worlds where those sentences are false. Those sentences are so-called **axioms** of mathematics. One can choose axioms quite arbitrarily, and maybe that is why we need fixed versions of axioms for convenience of communication. Nowadays **Peano Arithmetic** (**PA**) is commonly accepted as a first order axiom system for natural numbers. See appendix for a complete list.

When we say **mathematics** in the incompleteness theorem, it really means "axioms of mathematics", and in terms of arithmetic, it stands for the Axioms of Peano Arithmetic. As we are communicating with each other, we will assume that people will follow those axioms, i.e., their worlds all contain **PA** as true sentences.

However, there are sentences in mathematics whose truthfulness is argued among mathematicians. Sometimes it is shown that one sentence φ is independent from other commonly accepted axioms, i.e., we can find two worlds accepting common axioms and respectively having φ and its negation as true sentences. The most famous example is the Axiom of Choice, which is equivalent to the negation of Semiaxiom B.2.14 under the axiom system Zermelo-Fraenkel.

$$x - y = (D-1) - z$$

$$x = \square, y = \square, z = \square$$

I stared at that formula, and soon got a headache. Mathematics is of course one of my headache-causing subjects.
(Kyon's reaction to a puzzle appeared on a door. *The Rashness of Suzumiya Haruhi.)*

5. Proof

What is a proof? Maybe we shall see a formal proof of "$1+1=2$". First we list the two axioms we need:

Axiom 1. $\forall x(x+0=x)$

Axiom 2. $\forall x \forall y(x+Sy=S(x+y))$

The first sentence says that for every x, $x + 0 = x$; the second one says that, for any x and y, the sum of x and the successor of y, equals the successor of the sum of x and y. Now we give a proof of $1 + 1 = 2$: note that $1 = S0$ and $2 = S1 = SS0$.

(1)	$1 + 1 = 1 + S0$	by definition of 1
(2)	$1 + S0 = S(1 + 0)$	by Axiom 2
(3)	$1 + 0 = 1$	by Axiom 1
(4)	$1 + S0 = S1$	by (2) and (3)
(5)	$1 + 1 = S1 = 2$	by (1) and (4)

For simplicity we omit some steps of logic and axioms of equality, but one can get some idea from this: it appears that every step in the proof should be either from an axiom, or from a definition, or from other steps one has obtained.

More formally, a **proof** is a sequence of sentences $\langle P_1, P_2, P_3, ..., P_n\rangle$ where each P_i is an axiom, or is a logical deduction from the previous sentences. Here logical deduction can be explicitly written down as the **Modus Ponens**:

Modus Ponens:
Once one has $P \Rightarrow Q$ and P, then one can have Q.

I.e., if there is P_i of form $P \Rightarrow Q$ and P_j of form P, then one can add Q as a P_k, $(k > i, j)$ into the list.

We say **PA** proves some sentence P if there is a proof $\langle P_1, P_2, ..., P_n\rangle$ where axioms are axioms of logic or axioms of **PA** and $P_n = P$. Also such sequence is called a **proof** for P (in **PA**).

OK, now we have finished explaining terms in the incompleteness theorem, and let's have a look at the formal version of it:

Incompleteness Theorem, Second Version:
There is true sentence in arithmetic which cannot be proved by axioms of Peano Arithmetic.

At the first glance, you may get a wrong idea: **PA** might be simply too **weak** to prove everything true. Ironically, as you will see later, the incompleteness theorem actually says that **PA** is too **strong** to be complete.

> *Schadenfreude is bad, but Haruhi looks so happy that she wants to sing a song. That is to say, this person is such a low-life, end of proof.* (Kyon "proves" what Haruhi is. *Boredom of Suzumiya Haruhi.)*

6. Coding Provability

*Galaxy Universal Code: The first version of the Galaxy Universal Code (**GUC**), Galaxy Universal Code Standard 1.0 was published by the Galaxy Congress of Information in 2 A.G.P.. The aim of that was to "set up universal binary codes for different languages" (*[4, Introduction]*). Thereafter creatures in our galaxy have used **GUC** as a standard to communicate with each other. The current version, **GUC** Standard 7.14, was published in 834 A.G.P., after the discovery of a new sixth level civilization, Gezilians on Planet Gez.*
*(*Galaxy Universal Code, *The Hitchhiker's Guide to the Incompleteness Theorem.)*

A fundamental idea in the proof of the incompleteness theorem is that **PA** is strong enough to "write out" or to "code" the sentence "$\langle P_1, P_2, ..., P_n\rangle$ is a proof of P_n".

Codes are usually for privacy or simplicity, but here, as well as in modern computer science, we use codes for **expressibility**, i.e., we want to have a fixed way to use a natural number to code a formal sentence in our language. Then we can talk about sentences, truth and provability in terms of natural numbers (with plus, times and inequalities).

First we try to code two numbers into one: let

$$f(x, y) = \frac{1}{2}(x^2 + 2xy + y^2 + 3x + y).$$

It is not difficult to prove that $f(x, y)$ is a bijection, i.e., given any natural number z, one can uniquely find an ordered pair (x, y) s.t. $f(x, y) = z$. We will write $\langle x, y\rangle$ for this function. In some sense, we can "talk" about pairs of natural numbers by one single number. For example, we can write:

$$\begin{aligned}\mathbf{Code}_0(x, z) :=& \quad \exists y(z = \frac{1}{2}(x^2 + 2xy + y^2 + 3x + y))\\ \mathbf{Code}_1(y, z) :=& \quad \exists x(z = \frac{1}{2}(x^2 + 2xy + y^2 + 3x + y))\end{aligned}$$

to mean that x, y are the first and second components of $z = \langle x, y\rangle$, respectively.

Now we can define triples, quadruples, quintuples and so on:

$$\begin{aligned}\langle a, b, c\rangle &= \langle a, \langle b, c\rangle\rangle\\ \langle a, b, c, d\rangle &= \langle a, \langle b, \langle c, d\rangle\rangle\rangle\\ \langle a, b, c, d, e\rangle &= \langle a, \langle b, \langle c, \langle d, e\rangle\rangle\rangle\rangle\\ &\cdots\cdots\end{aligned}$$

Well, how about if we want to talk about all n-tuples? For example, we could use

$$g(x_1, x_2, ..., x_n) = \langle n, \langle x_1, x_2, ..., x_n \rangle\rangle$$

to code all of them, i.e. we first read the first bit of the code, then decode the remaining part accordingly. Now fix such a coding. Indeed we can write out a sentence $\mathbf{Code}(i, n, x)$ which says that "x is the i-th component of n". The proof of this fact is **difficult**, or we can have a much more complicated coding with an easier proof of this fact. See appendix for details.

We can code a formula in a similar way: first we fix the list of variables $v_1, v_2, v_3, ...$ and use the following table:

$+$	$\times$	S	$=$	$<$	$\vee$	$\wedge$	$\neg$	$\Rightarrow$	$\Leftrightarrow$	$\exists$	$\forall$	v_i
3	5	7	9	11	13	15	17	19	21	23	25	2i

For example, if we write $v_1 = v_2$, we will code this formula as $\langle 9, 2, 4\rangle$; if we write $\exists v_1 P(v_1)$ where P is a formula with code n, then we code this as $\langle 23, 2, n\rangle$. As a concrete example, let us look at:

Axiom 1.2.1: $\forall x(x = x)$

Change of variables does not change the meaning, so it is the same to say $\forall v_1(v_1 = v_1)$. $v_1 = v_1$ is coded as $\langle 9, 2, 2\rangle = 29649$. And the whole axiom is coded as $\langle 25, 2, \langle 9, 2, 2\rangle\rangle = \langle 25, 2, 29649\rangle$, which is about 4.66835×10^{33}. You might guess that long formulae will have huge codes. That is true, but you don't have to worry about it. Computers can easily handle such coding and decoding processes. In another point of view, this coding is not designed for practical use so we really do not care about its efficiency.

In this way we can code every formula into a number, and we can write out a sentence $\mathbf{Fm}(n)$ which says "n is (a code of) a formula" (we again omit the proof here).

Even better we can write out sentences saying the following:

$\mathbf{Ax}(n)$**:** "n is an axiom (logic or **PA**)".

$\mathbf{MP}(l, m, n)$**:** "l is of form $P \Rightarrow Q$, m is of form P and n is of form Q".

Details of the proof are left to the reader, or see reference [7].

As you might guess, we now can actually code a proof in a similar way: a proof is simply a sequence of formulae, each of which is coded into a number and each of which is either an axiom or a consequence by **Modus Ponens**. Following this we can write out:

$$\begin{aligned}\mathbf{Proof}(p, n) := \ & \exists i(\forall j \le i(\exists x(\mathbf{Code}(j, p, x) \wedge (\mathbf{Ax}(x) \vee \\ & \exists u, v < j(\mathbf{Code}(u, p, y) \wedge \mathbf{Code}(v, p, z) \wedge \mathbf{MP}(y, z, x))))) \\ & \wedge \mathbf{Code}(i, p, n))\end{aligned}$$

which says that the i-th component of p is n and for every component x before that, either it is an axiom ($\mathbf{Ax}(x)$), or it is from **Modus Ponens** ($\mathbf{MP}(y, z, x)$). So actually this formula codes the definition "p is a proof of a sentence n".

Finally we can write:

$$\mathbf{Pv}(n) := \quad \exists p(\mathbf{Proof}(p, n))$$

i.e. sentence n is provable (in **PA**).

> *There are some puzzling geometric figures on one side of it. What the hell is this? Sumerian language? I guess even if you try to decode that in Enigma (a cypher machine mostly used during WWII), it won't read anything.*
> *(*Kyon's comment on a piece of memo from Nagato Yuki. *The Boredom of Suzumiya Haruhi.)*

7. Final Trick and Discussions

> *Incompleteness Theorem, continued: According to Mye Postula, Arbiter of the First Galaxy Assembly in Mathematics, "the main consequence of the incompleteness theorem, ... ,is to guarantee that mathematicians will never lose their jobs". From another point of view, mathematicians are obliged to study mathematics forever to complete this subject, although "the completion of mathematics" is simply an illusion.*
> *(*Incompleteness Theorem, *The Hitchhiker's Guide to the Incompleteness Theorem.)*

The rest part of the proof is a bit tricky. First for any formula $\varphi(x)$ with only one free variable x, we can code φ as a number, which is usually denoted as $\ulcorner \varphi \urcorner$. If we plug this number into $\varphi(x)$, then we get a sentence $\varphi(\ulcorner \varphi \urcorner)$. We usually confuse φ with its code n, so we will write $n(n)$ to simplify notation. Now we put:

$$\mathbf{G}(n) := \quad \neg \mathbf{Pv}(n(n))$$

$\mathbf{G}(n)$ somehow says that, if we regard n as a code for a formula with one free variable, and plug in n as a number "into itself" to get a sentence $n(n)$, then this sentence $n(n)$ is not provable. All right, it is not difficult to see that $\mathbf{G}(n)$ is a formal formula in our language. Wait, $\mathbf{G}(n)$ is **a formal formula with one free variable**! That is to say, we can code $\mathbf{G}(n)$ into a number, say g, and plug g into $\mathbf{G}(n)$. In this process we get $\mathbf{G}(g)$, or $g(g)$, a formal sentence, which says, by definition of $\mathbf{G}(n)$, $g(g)$ is not provable.

OK, we get a sentence $P = g(g)$ which says that P itself is not provable. What does it mean? First of all, P is true. If it were false, P would be provable, but then it would mean that P is true. Contradiction! Secondly, since P is true, P is then not provable, by what P says! This finishes the proof of the incompleteness theorem.

Now let us look at the role of **PA** in the proof:
1. **PA** provides a coding method for formulae and proofs.
2. **PA** (with the axioms of logic) is an **effective** list of axioms, which can be described using codes of formulae.

Therefore any axiom system which is stronger than **PA** and which is an effective list of axioms can replace the role of **PA** in the proof above. I.e., we have a stronger form:

Incompleteness Theorem, Third Version:
For any axiom system which is powerful enough to express the arithmetic of natural numbers and consists of an effective list of axioms, it is either inconsistent (contradicts itself), or there is a true sentence which is not provable in that axiom system.

From the proof we get a sentence $P = g(g)$ which says that P itself is not provable (in **PA**). How about if we add P as an axiom into **PA**? Now in $\mathbf{PA} + P$, P is surely true and provable (since it is an axiom). But the process in our proof will generate a new P' which says that P' is not provable in $\mathbf{PA} + P$. And we can never end this process if we try to add such new sentences into **PA**.

Also note that we really need the condition "effective" in the theorem. For example, we can find all possible sentences true in the natural numbers, and define them as "axioms". Then it is easy to see that the conclusion of the incompleteness theorem does not hold, simply because "all possible true sentences about natural numbers" do not form an effective list, i.e. we do not have an effective way of telling which sentences are true.

Finally, there is a so-called "Second Incompleteness Theorem". We won't go through the details, but the proof of the incompleteness theorem can be written as a formal proof as in our definition. As a consequence, if we could prove the consistency of **PA** (i.e. it does not contradict itself) within **PA**, then for that P, we would have proved, **formally** in **PA**, that P is true, which is an contradiction to the incompleteness theorem. So **PA** cannot prove the consistency of itself. Furthermore we have a similar extension to theories stronger than **PA**:

Incompleteness Theorem, Fourth Version:
Any axiom system which is powerful enough to express the arithmetic of natural numbers and consists of an effective list of axioms cannot prove its own consistency, unless it is inconsistent.

> *"A consistent axiomatic set theory cannot prove its own consistency."*
> (Nagato Yuki's comment on Kyon's confusion with a seeming contradiction in his **own** world line. *The Boredom of Suzumiya Haruhi.)*

8. An Example: Goodstein's Theorem

You might be somewhat unsatisfied: the true sentence we get in the previous proof of the incompleteness theorem is really not that "mathematical" in the common sense. In this section we will see a real mathematical example which is not provable in **PA**. Both its proof and the proof that it is not provable in **PA** are far beyond our discussion here.

First let us define a notion of *base* n *representation of* m: first write m as the sum of powers of n. For example, if $m = 70$ and $n = 2$, then we write $70 = 2^6 + 2^2 + 2$.

Next write each exponent as the sum of powers of n, then repeat until the numbers are either n or 1. So as in our example $70 = 2^{2^2+2} + 2^2 + 2$. This is called the base n representation of m.

Now define a number $G_n(m)$: if $m = 0$ let $G_n(m) = 0$. Otherwise write m as a base n representation, and change every n by $n + 1$, then minus 1. So $G_2(70) = 3^{3^3+3} + 3^3 + 3 - 1 = 205891132094660$.

Define the Goodstein sequence of m as follows:

$$m_0 = m, m_1 = G_2(m_0), m_2 = G_3(m_1), m_3 = G_4(m_2), ...$$

This sequence will increase **dramatically**. For example, the sequence $\{4_k\}$ will reach a maximum $3 \times 2^{402653210} - 1$, which is about $3.44754 \times 10^{121210694}$.

Goodstein's theorem says that, for every m, the sequence $\{m_k\}$ will eventually reach 0. However, this theorem is not provable in **PA**. It turns out that the sequence increases so fast that **PA** cannot "see" it.

That is to say, we can find a group of elements with $(0, +, \times, <, S)$ defined and they "pretend" to be natural numbers in the sense that they satisfy every sentence that can be proved by **PA**, and Goodstein's theorem is false in that world. As there will be $0, 1, 2, 3, ...$ in that world (since they are defined in our language), there will also be some elements which are greater than any "real" natural number. Those are called "nonstandard" numbers. It is not too surprising that the Goodstein sequence of a nonstandard number may not reach 0, although this result is not as trivial as one might think.

Appendix

Polish Notation. Also called the *prefix notation*. Usually we write $5 + 4$ where $+$ is recognized as a function taking two variables. So if we write $+(5, 4)$, this looks fine. Then we can omit the parentheses: $+54$. Now for example, if we want to write $5 \times (4 + 7)$, we will write, in Polish notation, $\times 5 + 4\ 7$; if we want to write $5 \times 4 + 7$, we shall have $+ \times 5\ 4\ 7$. Similarly it works for logic connectives and quantifiers. One can prove that any such formula in Polish notation has a unique way of parsing.

However, Polish notation is not reader-friendly, so we usually write formulae in the common notation with parentheses.

Curvature. Curvature describes how "curved" a space is. For example, the Euclidean space (geometry) has curvature 0, while in other geometries the curvature varies. Parallel Postulate is false in those geometries with nonzero curvature.

Model Theory. Model Theory is a branch of mathematics studying **models**, or in our jargon, worlds. It has some strong connection with algebra.

Logic Axioms. Here is a list of logic axioms:

Group 1, axioms of connectives (one can define other connectives from $\Rightarrow$ and $\neg$, so we only need to take care of those two).

Axiom 1.1.1: $P \Rightarrow (P \Rightarrow Q)$
Axiom 1.1.2: $(P \Rightarrow (Q \Rightarrow R)) \Rightarrow ((P \Rightarrow Q) \Rightarrow (P \Rightarrow R))$
Axiom 1.1.3: $(\neg Q \Rightarrow \neg P) \Rightarrow (P \Rightarrow Q)$

Group 2, axioms of equality. Equality $(=)$ is usually considered as a logic symbol rather than a relation symbol, and we need equality in every world we care about.

Axiom 1.2.1: $\forall x(x = x)$
Axiom 1.2.2: $\forall x \forall y(x = y \Rightarrow y = x)$
Axiom 1.2.3: $\forall x \forall y \forall z((x = y \wedge y = z) \Rightarrow x = z)$
Axiom 1.2.4: $\forall x \forall y((x = y \wedge P(x)) \rightarrow P(y))$

Group 3, axioms of quantifiers.

Axiom 1.3.1: $\forall x P(x) \Rightarrow P(c)$
Axiom 1.3.2: $P(c) \Rightarrow \exists x P(x)$

First Order vs. Second Order. First order theory only talks about elements in a world. Second order theory also talks about sets of elements in a world. Usually the second order theory is more powerful than the first order one. One can similarly consider higher order theories but these two are the most common theories used in practice.

Peano Arithmetic. The original version of Peano Axioms is in second order theory. Commonly we adapt it to a first order version and call it Peano Arithmetic.

Axiom 2.1.1: $\forall x(\neg Sx = 0)$
Axiom 2.1.2: $\forall x \forall y(Sx = Sy \Rightarrow x = y)$
Axiom 2.1.3: $\forall x(\neg x < 0)$
Axiom 2.1.4: $\forall x \forall y(x < Sy \Leftrightarrow (x < y \vee x = y))$
Axiom 2.1.5: $\forall x(x + 0 = x)$
Axiom 2.1.6: $\forall x \forall y(x + Sy = S(x + y))$
Axiom 2.1.7: $\forall x(x \times 0 = 0)$

Axiom 2.1.8: $\forall x \forall y(x \times Sy = x \times y + x)$
Axiom 2.1.9: $\forall x \forall y((x < y) \vee (x = y) \vee (y < x))$
Axiom 2.1.10: $(P(0) \wedge \forall x(P(x) \Rightarrow P(x+1))) \Rightarrow (\forall x P(x))$

The last one is the induction axiom: if $P(0)$ is true, and for all x, $P(x)$ implies $P(x+1)$, then actually $P(x)$ holds for all x.

Zermelo-Fraenkel Axioms. Zermelo-Fraenkel Axiom system is a theory about sets. It is much stronger than **PA** and it is now accepted as a common foundation of mathematics.

Axiom of Choice. Axiom of Choice says that given a collection of nonempty sets of elements, one can pick one element from each set. It really sounds intuitively true, but it has a lot of results in mathematics which are not intuitive. It has been proved that Axiom of Choice is independence from Zermelo-Fraenkel Axioms.

Some Coding Issue. Here we give another way of coding and define **Code**(i, n, x) directly. See [7, Section 4.2] for proofs.

Given a sequence $x_1, x_2, x_3, ..., x_k$, we code them into a number:

$$Sq(x_1, x_2, x_3, ..., x_k) = \langle u, n! \rangle$$

where

$$q_i = \langle x_i, i \rangle, n = \max\{q_i : i \leq k\}, u = \prod_{i \leq k}(1 + (q_i \times n!)).$$

Define

$$P(i, n, x) := \exists u < n \exists v < n(n = \langle u, v \rangle \wedge \exists y(1 + (\langle x, i \rangle + 1) \times v) \times y = u)$$

and finally put:

$$\mathbf{Code}(i, n, x) := P(i, n, x) \wedge \forall y < x(\neg P(i, n, x))$$

References

[1] Himesuku Fujuta, *Handbook of Mathematics in the Galaxy*, Area 11 Central Press, 2013 A.T.B..
[2] Mingzhong Cai, *The Hitchhiker's Guide to the Incompleteness Theorem*, 2009 A.D., available online.
[3] Mye Postula, Jaume Talus, et al. *Declaration of the First Galaxy Assembly in Mathematics*, English version, Galrev-Regnirps, 3422 A.G.P..
[4] Committee of Standards, Galaxy Congress of Information, *Galaxy Universal Code Standard 1.0*, Universal Press, 2 A.G.P..
[5] Mye Postula, personal correspondence.
[6] Nagaru Tanigawa, Noizi Ito, *Suzumiya Haruhi* (series), Kadokawa Shoten, 2003 A.D.-present.
[7] Peter G. Hinman, *Fundamentals of Mathematical Logic*, A K Peters, 2005 A.D..
[8] Laurie Kirby, Jeff Paris, *Accessible Indepence Results For Peano Arithmetic*, Bulletin London Mathematical Society, 1982 A.D..

Department of Mathematics, Cornell University, Ithaca NY 14853
E-mail address: yiyang@math.cornell.edu