Incompleteness and Provability

Laura Pierson

September 22, 2015

1 Gödel's Theorems

Essentially, we're going to be talking about the limits of what statements a system of axioms can prove. Intuitively, it seems like any logical statement should be either true or false, and we should be able to determine which it is just using the axioms of our systems. In fact, this is actually true for geometry: Euclidean geometry can be formalized with a system of axioms so that there are no unprovable statements, i.e. the system is complete. This is basically because geometry is not strong enough to make the kinds of statements that cause problems in other theories, specifically, in any axiom system that contains the positive integers (formalized with a set of axioms called Peano Arithmetic).

A model of a system of axioms is a specific set of things that satisfy these axioms. However, these models might not all be the same, i.e., there might be other statements that are true in some models but not in others. Gödel's First Incompleteness Theorem tells us that in any system that can contains arithmetic, there are statements that are not provable. The Second Incompleteness Theorem says that such a system cannot prove its own consistency, and the Completeness Theorem says that if a statement is true in every model of a theory, then it is can be proved from the axioms (thus for any unprovable statement we find from the Incompleteness Theorems, there are models in which it is false).

1.1 The Incompleteness Theorems

The basic idea of Gödel's proof of the incompleteness theorems is to figure out how to encode self-referential statements and statements about provability using positive integers, which then lets us talk about these statements within our theory itself. The standard example used in the proof of the First Incompleteness Theorem is the statement "This statement is not provable." This statement must be true, since if it were false it would be provable and thus true. However, it cannot be provable, because then it would be false. Thus, it must be true but not provable from the system's axioms.

Let's look a bit at the specific mechanism used here:

Encoding of formulas We start by taking all the symbols that make up our language, including variables, constants, quantifiers ($\forall$, $\exists$), logical operators ($\wedge, \vee, \neg$), and functions/relations, and assign each one to a positive integer (its Gödel number). Now, any well-formed formula in our system can be represented as some finite sequence of these symbols, so we can now find an integer that corresponds to it as a function of the symbols that make it up. We could do this using concatenation, or using prime factorization (raise 2 to the Gödel number for the first symbol, 3 to the number for the second, 5 to the number for the 3rd, and so on). Now, since we can represent any formula in terms of just positive integers, we can represent relations between these formulas as relations between positive integers.

Provability Now we have to make sure this encoding can formalize the notion of being provable within our axiomatic system. There are a finite number of axioms and deduction rules, and a proof is some finite sequence of applications of these axioms and principles of deduction, all of which have Gödel numbers. We can thus write a statement that there exists a finite sequence of applications of the system's axioms that imply the truth of a given statement, and then find a Gödel number for this, which lets us encode statements about provability.

Self-Reference Okay, so we have some Gödel number for the formula "statement x is not provable," where x is a free variable. We can now plug the Gödel number of anything into this formula, *including the Gödel number for the formula itself.* This now gives us a new number to represent our desired statement, "This statement is not provable."

Given the First Theorem, the basic idea for the proof of the Second Theorem is simple (although the specifics are more complicated). If our axiomatic system can prove its own consistency, then we can formalize our whole *proof of the First Theorem* inside the theory itself, using a similar kind of assignment. But our proof of the First Theorem showed that the statement "This statement is provable" must be true, so if this proof itself could be written within our theory, the theory would prove this statement true, but then the statement would be provable, and this would give us a contradiction and make our system inconsistent.

1.2 The Completeness Theorem

Again, we won't go into all the details here, but essentially, Gödel's proof of the Completeness Theorem shows that, given any statement in our language, either the statement can be proven false from the axioms, or there is some model of the axiomatic system in which the statement is true. For any statement, basically what Gödel did was to change the quantifiers in the statement one by one to make it easier to work with. First, he moved all the quantifiers to the beginning of the statement, then successively changed some of the $\forall$ statements

to $\exists$ statements. to decrease the number of distinct "blocks" one by one, until eventually we just have a string of $\forall$ and then a string of $\exists$ statements, and this can be made into a string of $\exists$ only, which now lets us find a specific model that satisfies something that are not refutable.

2 Goodstein Sequences

2.1 Construction

Now let's look at an example of a simpler and more intuitive statement that is unprovable within Peano Arithmetic. We define the **hereditary base** n representation of a positive integer as follows: Write the number in base n. Now, if all our exponents are at most n, we stop, and if not, we write any exponent greater than n in base n, and we keep doing this until we have no numbers in our expression that are less than n. For example, suppose we want to write 772 in hereditary base 2 notation. We would get the following:

$$775 = 2^9 + 2^8 + 2^2 = 2^{2^3+1} + 2^{2^3} + 2^2 = 2^{2^{2+1}+1} + 2^{2^{2+1}} + 2^2.$$

Now let's define a **Goodstein sequence** as follows: Starting with any positive integer, write it in hereditary base 2 notation. Now change all the 2's to 3's (i.e. change the number to base 3) and then subtract 1. Then change all the 3's to 4's and subtract 1 again. For instance, if we start with 4, we get the following:

$$2^2 = 4 \to 3^3 - 1 = 2 \cdot 3^2 + 2 \cdot 3 + 2 = 26 \to 2 \cdot 4^2 + 2 \cdot 4 + 2 - 1 = 2 \cdot 4^2 + 2 \cdot 4 + 1 = 41$$

$$\to 2 \cdot 5^2 + 2 \cdot 5 = 60 \to 2 \cdot 6^2 + \cdot 6 + 5 = 83...$$

If we keep going like this, what happens? The sequences grow really fast, but it seems like they always hit zero eventually. However, can we prove this? Well, no, not in Peano Arithmetic. This is an example of Gödel's First Incompleteness Theorem, and one that doesn't involve complicated, obscure theorems resulting from a lengthy encoding process! But what axioms do we need to prove Goodstein's Theorem? What we need is to build the ordinals, assuming the well-ordering principle, and then we can bound these sequences above with a decreasing sequence of ordinals.

2.2 The Ordinals

Basically, the ordinals are an extension of the natural numbers to include infinite things, while still being able to be nicely written out in a list ("well-ordered"). We define the first infinite ordinal, ω, to be the first thing bigger than all the natural numbers. Then we just keep adding 1: $\omega + 1, \omega + 2, \omega + 3, ...$ The first thing bigger than all these numbers is $\omega \cdot 2$. Similarly, we can have $\omega \cdot 2 + 1, \omega \cdot 3, \omega \cdot 4, ...$, and the limit of these ordinals is ω^2. We can keep going to construct $\omega^3, \omega^4 + \omega^2 \cdot 3 + 17, \omega^\omega, \omega^{\omega^\omega}, ...$

There are two kinds of ordinals: successor ordinals and limit ordinals. Successor ordinals are just one more than the one before them (like the natural numbers, $\omega + 73, \omega^\omega \cdot 7 + \omega + 5, ...$), while limit ordinals are not one bigger than anything, but rather are the limit of an infinite sequence of smaller ordinals (e.g. $\omega, \omega^5 + \omega, \omega^\omega, ...$). Note that we cannot subtract ordinals: this would interfere with our ability to list them in this nice way.

In set theory, the ordinals can actually be described just in terms of sets: 0 is $\emptyset$, and every other ordinal is the set of all the ordinals smaller than it:

$$0 = \emptyset, 1 = \{0\} = \{\emptyset\}, 2 = \{0, 1\} = \{\emptyset, \{\emptyset\}\}, 3 = \{0, 1, 2\} = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}, ...$$

$$\omega = \{0, 1, 2, 3, ...\}, \omega + 1 = \{0, 1, 2, ..., \omega\}$$

Now we can state a very important property of the ordinals:

Well-Ordering Principle Any subset of an ordinal has a smallest element. Equivalently, any decreasing sequence of ordinals must reach zero in a finite amount of time. (Why are these equivalent? And why does this follow from our construction of the ordinals?)

2.3 Proof of Goodstein's Theorem

Okay, so now that we have the ordinals, how does this help us with Goodstein sequences? Essentially, we're going to bound any Goodstein sequence above with a decreasing sequence of ordinals. Then, by the Well-Ordering Principle, this sequence will eventually reach zero.

Let's take the Goodstein sequence starting with 4 as our example. In the first term we replace all 2's by ω's to get a bigger ordinal. Similarly, in the second term, we replace all 3's by ω's, in the third term we replace all 4's by ω's, and so on:

$$2^2 < \omega^\omega$$

$$3^2 \cdot 2 + 3 \cdot 2 + 2 < \omega^2 \cdot 2 + \omega \cdot 2 + 2$$

$$4^2 \cdot 2 + 4 \cdot 2 + 1 < \omega^2 \cdot 2 + \omega \cdot 2 + 1$$

$$5^2 \cdot 2 + 5 \cdot 2 < \omega^2 \cdot 2 + \omega \cdot 2$$

$$6^2 \cdot 2 + 6 + 5 < \omega^2 \cdot 2 + \omega + 5...$$

While the numbers are getting bigger, our upper bounds for them are getting smaller, because at each step we are either subtracting one or replacing some of the ω's by finite things, which are necessarily smaller. Thus, we have a decreasing sequence of ordinals, and since there are no infinite decreasing sequences of ordinals, we must eventually hit zero, and so the Goodstein sequence must as well.

2.4 Goodstein's Theorem and Peano Arithmetic

So we know now (if we believe in ordinals) that all Goodstein sequences terminate, and we won't find a counterexample to this, at least not in the model of Peano Arithmetic we're used to. So why can't we prove this from Peano Arithmetic? Well, basically, because Goodstein's Theorem actually implies the consistency of Peano Arithmetic. This is due to a theorem by Gerhard Gentzen. The termination of Goodstein sequences actually implies the Well-Ordering Principle, at least for small ordinals. (Suppose not. Then there exists some infinite decreasing sequence of ordinals. Then we can always find a Goodstein sequence that is "big enough" to be bigger than the ordinals in our sequence when we replace all the ωs by finite things.)

What Gentzen did was essentially to define a set of allowed reductions on derivations in the theory, which lead to simpler derivations but preserve any contradictions in the theory. These derivations form a tree, eventually leading down to the atomic statements (axioms) of the theory, and by assigning ordinal bounds to the heights of these axioms in the tree, he showed by the Well-Ordering Principle that some sequence of these reductions would eventually lead back to the axioms of the theory in a finite number of steps.