Lecture 14
Rosser's Theorem, the length of proofs, Robinson's Arithmetic, and Church's theorem

Michael Beeson

The hypotheses needed to prove incompleteness

The question immediate arises whether the incompleteness of $\mathbf{PA}$ can be "fixed" by extending the theory.

Gödel's original publication was about Russell and Whitehead's theory in *Principia Mathematica*. First-order logic was not yet a standard concept– that really didn't happen until the textbook of Hilbert-Bernays appeared in the last thirties, and because WWII arrived soon on the heels of Hilbert-Bernays, it was at least 1948 before the concept was widespread.

Gödel's methods were new and confusing: first-order logic was ill-understood, Gödel numbering was completely new, the diagonal method was confusing, the arithmetization of syntax was confusing. There were doubts about the generality of the result. Maybe it was a special property of *Principia Mathematica*? Then (as now) few people were familiar with that dense and difficult theory.

Maybe we just "forgot some axioms"?

An examination of the proof puts that idea to rest. Clearly the proof applies to any theory T such that

- T contains $\mathbf{PA}$
- The proof predicate $\mathrm{Prf}_T(k, x)$ ("k is a proof in T of the formula with Gödel number x") is recursive (so representable)
- the axioms of T are true in some model whose "integers" are isomorphic to the structure defined by $\mathbb{N}$ with standard successor, addition, and multiplication.

It works for all the axiom systems we know

- In particular those conditions apply to the strongest axioms accepted by most mathematicians, for example Zermelo-Frankel set theory with the axiom of choice, **ZFC**.
- So there are true sentences of arithmetic that are not provable even with the aid of all the axioms known to mathematics.
- And if someone discovers more axioms, there will still be true sentences that the new axioms don't prove.

Weakening the hypotheses

- Another reason for examining the hypotheses is to try to formalize the proof of the incompleteness theorem.
- In the proofs of the incompleteness theorem given above, we have freely assumed that all theorems of $\mathbf{PA}$ are true, i.e. satisfied in $\mathbb{N}$.
- But we already know that truth in $\mathbb{N}$ is not an arithmetical predicate, so the proofs we gave above of the First Incompleteness Theorem cannot be formalized in $\mathbf{PA}$.
- For various reasons, both philosophical and mathematical, we would like to weaken the hypotheses of the incompleteness theorem to the point where we could formalize the proof in $\mathbf{PA}$.

ω-consistency

- Gödel defined a theory to be "ω-inconsistent" if for some formula A, each instance $A(\bar{n})$ is provable but also $\vdash \neg\forall x\, A(x)$.
- Then a theory is ω-consistent if it is not ω-inconsistent.
- Omega consistency implies consistency, but it is stronger (see Exercise 13.1).
- The right way to think of ω-inconsistency is given in the following lemma:

Lemma

if T is ω-consistent and T extends $\mathbf{PA}$, and R is a bounded arithmetic predicate with only k free, and $T \vdash \exists k\, R(k)$, then for some k, $T \vdash R(\bar{k})$.

ω-consistency made understandable

Lemma

if T is ω-consistent and T extends $\mathbf{PA}$, and R is a bounded arithmetic predicate with only k free, and $T \vdash \exists k\, R(k)$, then for some k, $T \vdash R(\bar{k})$.

Proof. Suppose T and R are as stated, and suppose $T \vdash \exists k\, R(k)$. Then

$$T \vdash \neg\forall k \neg R(k).$$

By ω-consistency, for some k, T does not prove $\neg R(\bar{k})$. Since R is a bounded arithmetic formula, the predicate defined by R is represented by the formula R. Then either $T \vdash R(\bar{k})$ or $T \vdash \neg R(\bar{k})$, depending on whether $R(k)$ is true or false. Since in this case T does not prove $\neg R(\bar{k})$, it follows that $T \vdash R(\bar{k})$. That completes the proof.

Recursively axiomatizable theories

A theory T is called "recursively axiomatizable" if some set of axioms for it is recursive. (Technically, the set of Gödel numbers of axioms is recursive). That will result in its having a recursive proof predicate Prf_T.

Theorem (Gödel)

Let T be an ω-consistent recursively axiomatizable theory extending $\mathbf{PA}$. Then T is incomplete.

Proof. We go over the proof of the First Incompleteness Theorem to check that the assumption of truth can be replaced by the assumption of ω-consistency. Let Prf_T be the proof predicate for the theory T. Choose ϕ by the self-reference lemma to say "I am not provable." Specifically, choose ϕ so that

$$\mathbf{PA} \vdash \phi \equiv \neg \exists k \, \mathrm{Prf}_T(\ulcorner \phi \urcorner, k).$$

Now we claim that if T is consistent, ϕ is not provable, and if T is ω-consistent, then $\neg\phi$ is not provable.

With $\mathbf{PA} \vdash \phi \equiv \neg\exists k\, \mathrm{Prf}_T(\ulcorner\phi\urcorner, k)$

First suppose T is consistent and ϕ is provable in T. Let k be the Gödel number of a proof of ϕ. Then $\mathrm{Prf}_T(\ulcorner\phi\urcorner, k)$ is true.
Recall that Prf_T not only defines, but also represents, the proof predicate of T. Therefore we have

$$\mathbf{PA} \vdash \mathrm{Prf}_T(\overline{\ulcorner\phi\urcorner}, \bar{k})$$

Then $\mathbf{PA}$ also proves, in one more inference, $\exists k\, \mathrm{Prf}_T(\overline{\ulcorner\phi\urcorner}, \bar{k})$. But that is provably equivalent to $\neg\phi$. Since T extends $\mathbf{PA}$, T also proves $\neg\phi$. But now T proves both ϕ and $\neg\phi$, contradicting the consistency of T. Hence ϕ is unprovable, as claimed, if T is consistent.
Now suppose that T is ω-consistent and that $\neg\phi$ is provable. Then since T extends $\mathbf{PA}$, and using the formula in the slide title,

$$T \vdash \exists k\, \mathrm{Prf}_T(\overline{\ulcorner\phi\urcorner}, k).$$

By ω-consistency and the lemma above, there is a k such that

$$T \vdash \mathrm{Prf}_T(\overline{\ulcorner\phi\urcorner}, \bar{k}).$$

We showed there is a k such that

$$T \vdash \mathrm{Prf}_T(\overline{\ulcorner \phi \urcorner}, \bar{k}).$$

The formula Prf_T has been chosen (we suppose) to both define and represent the proof predicate of T, so k really is a proof of ϕ. Then $T \vdash \phi$. But then T proves both ϕ and $\neg\phi$, and hence T is inconsistent. But by Exercise 14.1, ω-inconsistency implies consistency, contradiction. That completes the proof.

Rosser's Theorem

Kleene's student Rosser was able to eliminate the hypothesis of ω-inconsistency, replacing it with simple consistency.

Theorem (Rosser, 1936)

Let T be a recursively axiomatizable consistent theory extending $\mathbf{PA}$. Then T is incomplete.

Proof. Choose ϕ by the self-reference lemma to say, "for every proof of me, there is a shorter proof of my negation." More formally: let $A(k, z)$ be the bounded arithmetic formula

$$\mathrm{Prf}_T(z, k) \supset \exists j < k\,(\mathrm{Prf}_T(\mathrm{Neg}(z), j))$$

and choose ϕ by the self-reference lemma so that

$$\mathbf{PA} \vdash \phi \equiv \forall k\, A(k, \overline{\ulcorner \phi \urcorner}) \tag{1}$$

Then

$$\mathbf{PA} \vdash \phi \equiv \mathrm{Prf}_T(\overline{\ulcorner \phi \urcorner}, k) \supset \exists j < k\,(\mathrm{Prf}_T(\overline{\ulcorner \neg\phi \urcorner}, j))$$

Rosser's theorem continued

$$\mathbf{PA} \vdash \phi \equiv \mathrm{Prf}_T(\overline{\ulcorner\phi\urcorner}, k) \supset \exists j < k\,(\mathrm{Prf}_T(\overline{\ulcorner\neg\phi\urcorner}, j)) \qquad (2)$$

Now suppose ϕ is provable in T. Let k be (the Gödel number of) a proof of ϕ in T. Since Prf_T represents the proof predicate of T,

$$\mathbf{PA} \vdash \mathrm{Prf}_T(\overline{\ulcorner\phi\urcorner}, \bar{k})$$

Since T proves ϕ, and T extends $\mathbf{PA}$, it follows from (2) that

$$T \vdash \exists j < \bar{k}(\mathrm{Prf}_T(\overline{\ulcorner\neg\phi\urcorner}, j)) \qquad (3)$$

The formula on the right defines a bounded arithmetic predicate, and hence it represents that same predicate. Since T is consistent, the formula $\exists j < \bar{k}(\mathrm{Prf}_T(\overline{\ulcorner\neg\phi\urcorner}, j))$ is false (if it were true, that would contradict the fact that T proves ϕ). Since that formula represents a primitive recursive predicate, it is refutable. But that, with (3), contradicts the consistency of T. Therefore ϕ is not provable in T.

Rosser's theorem continued

On the other hand, if $\neg\phi$ is provable in T, then let j be the Gödel number of the smallest proof of $\neg\phi$. Then $\mathrm{Prf}_T(\overline{\ulcorner\neg\phi\urcorner}, \bar{j})$ is true. Since Prf_T represents the proof predicate of T, we have

$$\mathbf{PA} \vdash \mathrm{Prf}_T(\overline{\ulcorner\neg\phi\urcorner}, \bar{j}).$$

Hence $\mathbf{PA} \vdash \forall k > \bar{j}\, A(k, \overline{\ulcorner\phi\urcorner})$.
Therefore $\mathbf{PA} \vdash \phi \equiv \forall k < \bar{j}\, A(k, \overline{\ulcorner\phi\urcorner})$.
Since $\neg\phi$ is provable, and T extends $\mathbf{PA}$, we have

$$T \vdash \exists k < \bar{j}\, \neg A(k, \overline{\ulcorner\phi\urcorner}).$$

This is a bounded arithmetic formula, so it represents the predicate it defines, in this case a predicate of zero arguments, but still: if that predicate is false, it is refutable, which would contradict the consistency of T. Hence it is true, i.e. for some $k < j$, $A(\bar{k}, \overline{\ulcorner\phi\urcorner})$ is false. But A is an implication, so if it is false, the part before the implication sign is true, which means k is proof of ϕ. But since $\neg\phi$ is provable, that contradicts the consistency of T. That completes the proof of Rosser's theorem.

On the lengths of proofs

Gödel himself in a later publication studied the question of lengths of proofs. Here we present a simple result in that direction.

Theorem (Gödel)

Let f be any primitive recursive function of one variable. Then there is a formula ϕ of one free variable such that $\forall x \phi(x)$ is true, but for each n, $\phi(\bar{n})$ has no proof with fewer than $f(n)$ steps.

Proof. Choose ϕ by the self-reference lemma to say "I have no proof shorter than $f(n)$". Then, if $\phi(\bar{n})$ is false, it does have a proof, and so it is true, contradiction. Hence $\phi(\bar{n})$ is true. Hence it has no proof with fewer than $f(n)$ steps.

On the next slide we will give the proof in more detail.

Details of the no-short-proof theorem

To make this proof more precise, we choose ϕ so that

$$\mathbf{PA} \vdash \phi \equiv \forall k(length(k) < f(n) \supset (\neg \mathrm{Prf}(\overline{\ulcorner \phi \urcorner}, k))$$

where $length$ is the length of a sequence (a proof is a sequence of steps with justification), and "$j < f(n)$" means $\exists y \, R(n, y) \wedge j < y)$, where R represents f.

Now we can check the steps of the proof as sketched before:

- If $\phi(\bar{n})$ is false, it does have a proof, and so it is true, contradiction.
- Hence $\phi(\bar{n})$ is true.
- Hence it has no proof with fewer than $f(n)$ steps.
- That completes the proof.

Lengths of proofs and incompleteness

We will show that the theorem on the lengths of proofs is actually a generalization of the incompleteness theorem, or put another way, the First Incompleteness Theorem is a corollary of Gödel's theorem about long proofs.

- If $\forall x\, \phi(x)$ is provable, then all the instances $\phi(\bar{n})$ have proofs obtained from that fixed proof by substituting $\bar{n}$ for x.
- These proofs are longer by about the length of $\bar{n}$, which is about n.
- so taking f in the long-proofs theorem to be, say n^2, or any function growing faster than n, we see that $\forall x\, \phi(x)$ is not provable, where ϕ is the formula from the long-proofs theorem.

What is the smallest theory for which the Incompleteness Theorem works?

- We showed that it works for any recursively axiomatizable consistent extension of $\mathbf{PA}$.
- It is natural to consider replacing $\mathbf{PA}$ in this result by the smallest possible subtheory of $\mathbf{PA}$.
- So we examine what is actually used, and try to find the minimum required axiom set.
- The answer was found by Raphael Robinson, Berkeley 1950.

Robinson Arithmetic

Here are the seven axioms of Robinson Arithmetic $\mathbf{RA}$:

$$x' \neq 0 \tag{1}$$
$$x' = y' \supset x = y \tag{2}$$
$$y \neq 0 \supset \exists x(x' = y) \tag{3}$$
$$x + 0 = x \tag{4}$$
$$x + y' = (x + y)' \tag{5}$$
$$x \cdot 0 = 0 \tag{6}$$
$$x \cdot (y') = (x \cdot y) + x \tag{7}$$

Axiom (3) is provable by induction in $\mathbf{PA}$; the others are axioms of $\mathbf{PA}$.
Thus $\mathbf{RA}$ is $\mathbf{PA}$ without induction, plus the axiom $y \neq 0 \supset \exists x(x' = y)$.

Raphael Robinson

Robinson was a professor at UC Berkeley. After Tarski arrived there, he worked with Tarski for many years. He died in 1995.

$\mathbf{RA}$ is a very weak theory

- $\mathbf{RA}$ does not even prove the commutativity of multiplication
- I have not proved that fact for you, but at least you can see that in $\mathbf{RA}$, we don't have induction available, which is how we proved commutativity in $\mathbf{PA}$.
- We will see that $\mathbf{RA}$ nevertheless represents all primitive recursive functions.
- This highlights again the difference between representing functions, and proving that those functions satisfy their recursion equations (with a free variable), and proving properties of those functions.

Presburger Arithmetic

If we drop multiplication (keeping only the first five axioms of $\mathbf{RA}$) we obtain Presburger Arithmetic.

- Presburger Arithmetic is decidable.
- This was proved by the Polish student Mojżesz Presburger in 1929, when he was 25 years old.
- Presburger was Jewish, so could not obtain an academic position and died in the Holocaust.

Mojżesz Presburger

Proof of Robinson's Theorem

- We need to check that $\mathbf{RA}$ can represent all primitive recursive predicates.
- We did that in Lecture 11 for $\mathbf{PA}$.
- In order to prove Robinson's theorem, we need to check that we do not need induction to prove that all primitive recursive predicates are representable. We only need Axiom 3 (nonzero numbers are successors), which in $\mathbf{PA}$ is proved by induction.
- This is not completely obvious.

On the definition of $x < y$

There is an issue about what is the right definition of $x < y$ to use in $\mathbf{RA}$. Kleene defines it (§17, and again on pages 196 and 229, with discussion on p. 196) as

$$\exists z(z' + x = y).$$

The alternative would be

$$\exists z(x + z' = y).$$

- These can be proved equivalent in $\mathbf{PA}$, using induction, but not (apparently) in $\mathbf{RA}$.
- Which one should be used in $\mathbf{RA}$?

A possible issue with $x < y$

We need to prove two things about $x < y$.

- That $x < y$ is representable by the formula defining it.
- That works with Kleene's definition, see p. 196, and see also Lecture 10, where we verified that we did not need formal induction, except to show that nonzero numbers are successors, which is Axiom 3 of $\mathbf{RA}$.
- that for each numeral $\bar{n}$,

$$\mathbf{RA} \vdash x < \bar{n}' \equiv x = 0 \vee \ldots \vee x = \bar{n}.$$

- That works straightforwardly with the alternative definition, as shown in the exercises for Lecture 10.
- With more difficulty, it also works for Kleene's definition, see Kleene p. 198 and the exercises.

Representability of β in $\mathbf{RA}$

We need to check that the proof that β is representable works in $\mathbf{RA}$; that is, we never needed induction in the formal system.

- Indeed in the lecture slides we never needed induction.
- But we referred to Kleene p. 203 for the rest of the details.
- On that page, there's also no formal induction.

The first incompleteness theorem works for $\mathbf{RA}$

- We've proved it now.
- Kleene states it as Lemma 18a, p. 204, the implication being that he has already proved it without stating it.
- Never mind that we didn't prove

$$\mathbf{PA} \vdash x < \bar{n}' \equiv x = 0 \vee \ldots \vee x = \bar{n}$$

  in $\mathbf{RA}$. Kleene never claims that you can prove that in $\mathbf{RA}$.
- Study guide: Note carefully the starred items near the top of p. 204, that are *not* claimed to be provable in $\mathbf{RA}$. Make sure you understand that those formulas, involving variables instead of numerals, say much more than that the functions involved are representable. They say that those functions are provably total.
- $\mathbf{RA}$ seems almost obvious today, but before Robinson, Kleene was using another finitely axiomatizable theory, as Kleene says on p. 533. Robinson's abstract was available in time for the printing of Kleene's 1952 book, but not his publication; so he must have changed his book to do things using only $\mathbf{RA}$.

The *Entscheidungsproblem*, or Decision Problem

The problem is, does there exist an algorithm to determine if a formula ϕ of predicate calculus is, or is not, provable in the predicate calculus.

- By the completeness theorem, that's the same as asking whether there is an algorithm to determine the validity of ϕ, i.e., is ϕ true in all models of the language used in ϕ?
- There is no such algorithm, as we now prove. Let A be the conjunction of the axioms of $\mathbf{RA}$. If there were such an algorithm, we could decide whether

$$A \supset \exists k\, \mathbb{T}(\bar{e}, \bar{e}, k)$$

  is provable in predicate calculus.
- But that is if and only if $\mathbf{RA} \vdash \exists k\, \mathbb{T}(\bar{e}, \bar{e}, k)$, which is true if and only if $\varphi_e(e)$ halts, since $\mathbf{RA}$ is sound and proves every true Σ_1^0 sentence.
- So if the decision problem for predicate calculus were recursively solvable, we could recursively solve the halting problem.

$\mathbf{RA}$ and the *Entscheidungsproblem*

The proof we just gave depends on the facts that

- $\mathbf{RA}$ is finitely axiomatizable, and
- $\mathbf{RA}$ represents every primitive recursive predicate, in particular, the $\mathbf{T}$ predicate, and therefore proves every true Σ_1^0 sentence.
- But *any* finitely axiomatized theory with those two properties would have worked, it did not need to be the absolutely minimal theory $\mathbf{RA}$.

Historical note

- Turing showed the *Entscheidungsproblem* could not be solved by a Turing machine, in his original paper on Turing machines.
- He did not have the $\mathbf{T}$-predicate or make any use of $\mathbf{PA}$ or Gödel numbers.
- Instead, he directly formulated a theory in predicate calculus whose models are Turing machine computations.
- The *Entscheidungsproblem* was a famous problem, put forward in Hilbert-Ackermann's 1928 logic textbook.
- It was solved independently by Turing and Church in 1936, and its solution made both of them famous; but Church was already well-known and Turing was only 24.
- Note the ambiguity: The *Entscheidungsproblem* refers both to the problem of deciding validity in predicate calculus, and the problem of whether or not there exists an algorithm to do that. So Turing solved the *Entscheidungsproblem* by showing that the *Entscheidungsproblem* cannot be solved by a Turing machine.

Church's theorem

There is no algorithm for deciding if a formula ϕ is a theorem of $\mathbf{PA}$.

Proof. $\mathbf{PA} \vdash \exists k\, \mathbb{T}(\bar{e}, \bar{e}, k)$ if and only if $\phi_e(e)$ halts. So if we could decide the former by an algorithm, we could recursively solve the halting problem.

- Church's proof showed that there is no algorithm to decide the equivalence or non-equivalence of two terms in the λ-calculus, plus the definability of the λ-calculus in arithmetic.
- In other words, he used λ-calculus instead of Turing machines.
- This result also solves the *Entscheidungsproblem*.

Alonzo Church 1903–1995

Essentially Undecidable Theories

A theory T is called essentially undecidable if every consistent extension of T is undecidable.

$\mathbf{RA}$ is essentially undecidable.

Proof. Let T be a consistent extension of $\mathbf{RA}$.

- if T is decidable, then its theorems constitute a recursive axiomatization. So if T is not recursively axiomatizable, it is undecidable.
- Therefore we may assume T is recursively axiomatizable.
- Since T contains $\mathbf{RA}$, all primitive recursive functions and predicates are representable in T.
- Then the proof of Church's theorem given above applies to T, so T is undecidable.
- That is (to review) $T \vdash \exists k\, \mathbb{T}(\bar{e}, \bar{e}, k)$ if and only if $\varphi_e(e)$ halts.