Lecture 15
The Second Incompleteness Theorem

Michael Beeson

The Second Incompleteness Theorem

- Let Con_{PA} be the formula

$$\forall k \neg \mathrm{Prf}(k, \overline{\ulcorner \bar{0} = \bar{1} \urcorner})$$

- Then Con_{PA} expresses the consistency of $\mathbf{PA}$.
- The second incompleteness theorem is this:
  $\mathbf{PA} \nvdash \mathit{Con}_{PA}$
- That is, $\mathbf{PA}$ does not prove its own consistency.

Death knell for Hilbert's program

Hilbert's program, developed by Hilbert in the 1920s, called for

- Dividing mathematics into the "finitistic part" and the "infinitistic part".
- The finitistic part should use only completely unquestionable principles, e.g. primitive recursion and quantifier-free induction.
- The consistency of the infinitistic part should be proved in the finitistic part.
- That would establish that it is "safe" to use infinitistic methods in mathematics.
- But Gödel's second incompleteness theorem showed that one cannot even prove the consistency of the *finitistic* part, let alone the consistency of the infinitistic part, using only finitistic methods.

First incompleteness theorem reconsidered

Kleene (bottom of p. 208) states the theorem this way:

If $\mathbf{PA}$ is (simply) consistent then neither $\vdash A_q(\bar{q})$ nor $\vdash \neg A_q(\bar{q})$. Here $A_q(\bar{q})$ is the Rosser sentence that says "For every proof of me, there's a shorter proof of my negation." In Kleene's notation, A_q is the formula with Gödel number q, and here q is a particular integer defined on p. 208.

Why does Kleene put in the hypotheses about consistency?

- **$\mathbf{PA}$ is consistent**, so these hypothesis are superfluous.
- The reason he puts it in anyway: he has in mind to formalize the theorem in $\mathbf{PA}$, where (as it turns out) the hypothesis Con_{PA} is not provable, so it must be stated explicitly in a version that we hope to prove in $\mathbf{PA}$.

Formalizing the First Incompleteness Theorem

We will formalize the statement of
If $\mathbf{PA}$ is (simply) consistent then neither $\vdash A_q(\bar{q})$ nor $\vdash \neg A_q(\bar{q})$.

$$\mathrm{Con}_{PA} \supset \forall k\,(\neg \mathrm{Prf}(k, Subst(Num(q), \overline{\ulcorner x \urcorner}, q)$$
$$\wedge \neg \mathrm{Prf}(k, Neg(Subst(Num(q), \overline{\ulcorner x \urcorner}, q))))$$

- Technically we should use formulas representing $Subst$ and Num.
- The function Neg produces the Gödel number of $\neg E$ from $\ulcorner E \urcorner$.
- x is a certain variable, first in the list of variables.
- This shows that the theorem can at least be *expressed* in $\mathbf{PA}$.

Proving the second incompleteness theorem

Suppose we could formalize the proof of the first incompleteness theorem. Then

$$\vdash \mathrm{Con}_{PA} \supset \forall k\,(\neg \mathrm{Prf}(k, Subst(Num(q), \overline{\ulcorner x \urcorner}, q)$$

$$\wedge \neg \mathrm{Prf}(k, Neg(Subst(Num(q), \overline{\ulcorner x \urcorner}, q))))$$

Now suppose, for proof by contradiction, that $\vdash \mathrm{Con}_{PA}$. Then

$$\vdash \forall k\,(\neg \mathrm{Prf}(k, Subst(Num(q), \overline{\ulcorner x \urcorner}, q)$$

$$\wedge \neg \mathrm{Prf}(k, Neg(Subst(Num(q), \overline{\ulcorner x \urcorner}, q))))$$

Then we argue in $\mathbf{PA}$ to prove $A_q(\bar{q})$ as follows: it suffices to show that for every proof of $A_q(\bar{q})$ there's a shorter proof of $\neg A_q(\bar{q})$. But there is no proof of $\neg A_q(\bar{q})$; so that proves $A_q(\bar{q})$. That contradiction proves the second incompleteness theorem.

Formalization is difficult

The proof then comes down to going through the proof of the first incompleteness theorem, and formalizing every step in $\mathbf{PA}$.

- We referred to many syntactic objects; these now must be mentioned by Gödel number.
- Proofs by induction on the complexity of terms or formulas now become proofs by mathematical induction in $\mathbf{PA}$.
- We need course-of-values induction: if something is true for all numbers less than k implies it's true for k, then it holds for all numbers. (You showed in an exercise how to prove this in $\mathbf{PA}$.)
- All the things we prove by induction need to be expressible by arithmetical formulas.
- We saw that "every theorem of $\mathbf{PA}$ is true" is not so expressible. That shows that the usual proof that $\mathbf{PA}$ is consistent is not directly formalizable.

Remarks

- Formalizing the proof in $\mathbf{PA}$ requires much more than what we did when we checked what conditions on a theory T were needed for the first incompleteness theorem to hold.
- Then we only had to check that $x < y$, cutoff division, remainder, and β are representable.
- For example we did not need to formalize the Chinese remainder theorem or the main properties of β.
- We did not even need to prove that rem and β are total functions.
- That's why we could get by with the tiny theory $\mathbf{RA}$.
- But we may well need more for formalizing the proof.

Formalizing a metamathematical proof

Consider, for example, the theorem that for closed terms t, if $Val(t) = m$ then $\vdash t = \bar{m}$.

The following predicates and functions are primitive recursive:

- $Term(x)$, true when x is the Gödel number of a term.
- $ClosedTerm(x)$, true when x is the Gödel number of a term with no free variables.
- $functor$, $arity$, arg_1, and arg_2, considered as defined on Gödel numbers of terms.
- $V(x)$, the value of the closed term whose Gödel number is x.
- Since arg_1 and arg_2 are decreasing functions, V is defined by course-of-values recursion. For example if $functor(x) = \overline{43}$ (43 is ascii for $+$) then $Val(x) = Val(arg_1(x)) + Val(arg_2(x))$.
- So $Val(x)$ is defined by cases on $functor(x)$, with recursive calls to Val at smaller arguments, grounded by $Val(48) = 0$ (48 is ascii for zero.)

Formalizing that closed terms are provable equal to their values

- Since Val is primitive recursive, it is represented by a formula V.
- Given the Gödel numbers u and v of terms α and β, we can construct the Gödel number $eq(u, v)$ of the term $\alpha = \beta$.
- The function eq is also primitive recursive.
- Now the theorem we are trying to formalize is

  $$ClosedTerm(x) \supset \exists k, m\,(V(x, m) \wedge Prf(k, eq(x, Num(m))))$$

  because, if x is the Gödel number of a closed term t, $V(x, m)$ says m is the value of term t, and $eq(x, Num(m))$ is the Gödel number of the formula $t = \bar{m}$.

A sample formalized lemma

In the course of proving that closed terms are provably equal to their values, we need to formalize a proof of

$$\vdash \bar{\alpha} + \bar{\beta} = \overline{\alpha + \beta}.$$

Formally that becomes

$$\exists k \, \mathrm{Prf}(k, eq(sum(Num(a), Num(b)), Num(a + b)))$$

where sum constructs the Gödel number of $\alpha + \beta$ given Gödel numbers of α and β.

- We proved that lemma by induction on β. So, the proof can be formalized in $\mathbf{PA}$.
- But even that would be a long proof. Remember how long the proof of $\bar{2} + \bar{2} = \bar{4}$ turned out to be.

Summary of the lecture so far

- We formalized the statement of the first incompleteness theorem.
- We showed that if the first incompleteness theorem can be proved in $\mathbf{PA}$, that implies the second incompleteness theorem.
- It seems likely that the proof *can* be formalized, but to achieve certainty would require exhibiting and checking a very long formal proof.
- Even if that could be done, it would still be good to have a shorter proof.
- Therefore we wish to isolate a few (three as it turns out) simple properties such that, if we verify those three properties, the second incompleteness theorem follows.

What is needed for the Second Incompleteness Theorem?

- We specify conditions on the formula Pr_T that represents the provability predicate of a theory T, sufficient to guarantee that the second incompleteness theorem holds for T.
- Such conditions were first written down by Hilbert-Bernays in their 1939 textbook. The conditions were simplified by the Dutch logician Löb in a 1955 publication.
- So now, those conditions D1, D2, and D3 are known by all three names. Wikipedia omits Löb's name, but the Stanford Encyclopedia of Philosophy gets it right, as do some other books.
- Kleene's 1952 book was pre-Löb, and he states the Second Incompleteness Theorem, but refers to Hilbert-Bernays for the proof for "a slightly different system" than $\mathbf{PA}$.

The Hilbert-Bernays-Lob provability conditions

Let $\mathrm{Pr}(x)$ be $\exists k\, \mathrm{Prf}_T(k, x)$, i.e. "$x$ is the Gödel number of a theorem of T." We write $\mathrm{Pr}(x)$ instead of $\mathrm{Pr}_T(x)$ just for brevity. Let $implies(x, y)$ produce the Gödel number of $A \supset B$ from Gödel numbers x and y of A and B.

Hilbert and Bernays showed that the second incompleteness theorem holds for any recursively axiomatizable theory T that represents satisfies the conditions for the first incompleteness theorem, plus the following. We drop the numeral overlines over Gödel numbers for simplicity.

- (D1) if $T \vdash \phi$ then $T \vdash \mathrm{Pr}(\ulcorner \phi \urcorner)$.
  "provable implies provably provable."
- (D2) $T \vdash \mathrm{Pr}(\ulcorner \phi \urcorner) \supset \mathrm{Pr}(\ulcorner \mathrm{Pr}(\ulcorner \phi \urcorner) \urcorner)$
  (This is just the formalization of the first condition.)
- (D3) $T \vdash \mathrm{Pr}(\ulcorner \phi \supset \psi \urcorner) \wedge \mathrm{Pr}(\ulcorner \phi \urcorner) \supset \mathrm{Pr}(\ulcorner \psi \urcorner)$

A lemma using D1-D3

Suppose $\vdash A \equiv B$. Then $\vdash \Pr(\ulcorner A \urcorner) \equiv \Pr(\ulcorner B \urcorner)$.

- It suffices to prove it with $\supset$ instead of $\equiv$.
- Suppose $\vdash A \supset B$. Then by D1, $\vdash \Pr(\ulcorner A \supset B \urcorner)$.
- Then by D3, $\vdash \Pr(\ulcorner A \urcorner) \supset \Pr(\ulcorner B \urcorner)$.
- QED

Löb's theorem

Löb used his conditions D1-D3 to answer the question (which you were asked to think about in a homework exercise) about the fixed points of Pr. His theorem shows that the only fixed points are the theorems of T.

Theorem (Löb)

If $T \vdash \mathrm{Pr}(\overline{\ulcorner \psi \urcorner}) \supset \psi$ *then* $T \vdash \psi$.

- This theorem might seem like just a curiosity, but we will see below that it quickly implies the Second Incompleteness Theorem.
- That is how we prove that D1-D3 imply the Second Incompleteness Theorem.

Proof of Löb's theorem

Suppose $T \vdash \Pr(\ulcorner\psi\urcorner) \supset \psi$. We must show $T \vdash \psi$. Choose ϕ by the self-reference lemma so that

$$T \vdash \phi \equiv (\Pr(\ulcorner\phi\urcorner) \supset \psi).$$

$$\vdash \Pr(\ulcorner\phi\urcorner) \equiv \Pr(\ulcorner \Pr(\ulcorner\phi\urcorner) \supset \psi\urcorner) \qquad \text{by the lemma}$$

Now $\vdash \Pr(\ulcorner \Pr(\ulcorner\phi\urcorner) \supset \psi\urcorner) \wedge \Pr(\ulcorner \Pr(\ulcorner\phi\urcorner)\urcorner) \supset \Pr(\ulcorner\psi\urcorner)$ is an instance of D3. And $\Pr(\ulcorner\phi\urcorner)$ provably implies both conjuncts on the left. Therefore

$$\vdash \Pr(\ulcorner\phi\urcorner) \supset \Pr(\ulcorner\psi\urcorner).$$

But at the top of the slide, we assumed $\vdash \Pr(\ulcorner\psi\urcorner) \supset \psi$. That gives us

$$\vdash \Pr(\ulcorner\phi\urcorner) \supset \psi.$$

But then by the defining property of ϕ we have $\vdash \phi$. Hence by D1 we have $\vdash \Pr(\ulcorner\phi\urcorner)$. But since $\vdash \Pr(\ulcorner\phi\urcorner) \supset \psi$, by D3 we have $\vdash \psi$. That completes the proof.

Löb's theorem implies the Second Incompleteness Theorem

All we have to do is put $0 = \bar{1}$ in for the formula in Löb's theorem, and the Second Incompleteness Theorem drops out.

This observation was made by Kreisel in 1965; I do not know if he was the first to observe it. Let $\bot$ abbreviate $\overline{\ulcorner 0 = \bar{1} \urcorner}$, so Con_T is provably equivalent to $\neg \Pr(\bot)$. Then

- If $0 = \bar{1}$ is not provable, then by Löb's theorem, $\Pr(\bot) \supset 0 = \bar{1}$ is not provable.
- But $A \supset 0 = \bar{1}$ is provably equivalent to $\neg A$.
- Hence, if $0 = \bar{1}$ is not provable, then $\neg \Pr(\bot)$ is not provable.
- That is, if $0 = \bar{1}$ is not provable, then Con_T is not provable.
- That is, if T is consistent, then T does not prove Con_T.
- That is the Second Incompleteness Theorem.

Verification of D1 for recursive extensions T of $\mathbf{PA}$

- D1 is "provable implies provably provable"
- This is a special case of the fact that every true Σ_1^0 sentence is provable.
- You proved this in an exercise.

Verification of D3

D3 is easier to verify than D2, so we do it first.

$$T \vdash \Pr(\ulcorner \phi \supset \psi \urcorner) \wedge \Pr(\ulcorner \phi \urcorner) \supset \Pr(\ulcorner \psi \urcorner)$$

Replacing $\Pr(y)$ by $\exists x \operatorname{Prf}(x, y)$ this becomes

$$T \vdash \operatorname{Prf}(x, \ulcorner \phi \supset \psi \urcorner) \wedge \operatorname{Prf}(z, \ulcorner \phi \urcorner) \supset \exists w \operatorname{Prf}(w, \ulcorner \psi \urcorner).$$

Since modus ponens is one of the rules of inference, there will be a primitive recursive function mp that gets a Gödel number of a proof of ψ from Gödel numbers of proofs of $\phi \supset \psi$ and ϕ. So we just need to show

$$T \vdash \operatorname{Prf}(x, \ulcorner \phi \supset \psi \urcorner) \wedge \operatorname{Prf}(z, \ulcorner \phi \urcorner) \supset \operatorname{Prf}(mp(x, z), \ulcorner \psi \urcorner).$$

In other words, we need to show that the formula representing the proof predicate satisfies its defining recursion equations.

Verification of D2

D2 is the formalization of D1:

$$T \vdash \Pr(\ulcorner \phi \urcorner) \supset \Pr(\ulcorner \Pr(\ulcorner \phi \urcorner) \urcorner)$$

Putting in the definition of $\Pr(y)$ as $\exists x \operatorname{Prf}(x, y)$, this amounts to

$$T \vdash \operatorname{Prf}(x, \ulcorner \phi \urcorner) \supset \exists z \operatorname{Prf}(z, \ulcorner \Pr(\ulcorner \phi \urcorner) \urcorner).$$

- This says that within T, proofs (encoded by Gödel number) can be verified to be proofs.
- We need to tell how to find z from x: given a proof of ϕ, how do we convert it to a proof of $\Pr(\ulcorner \phi \urcorner)$?
- We need to define a primitive recursive function F ("F for formalize") such that

$$T \vdash \operatorname{Prf}(x, \ulcorner \phi \urcorner) \supset \operatorname{Prf}(F(x), \ulcorner \Pr(\ulcorner \phi \urcorner) \urcorner).$$

- This we could prove in T by course of values induction on x, once F is defined.

Verification of D2 continued

Working on

$$T \vdash \mathrm{Prf}(x, \ulcorner \phi \urcorner) \supset \mathrm{Prf}(F(x), \ulcorner \mathrm{Pr}(\ulcorner \phi \urcorner) \urcorner).$$

- For example, one of the equations for F says that if the last step in proof x was modus ponens, then $F(x) = mp(F(arg_1(x)), F(arg_2(x)))$.
- so if x is a proof of B by modus ponens from $A \supset B$ and A, then $arg_1(x)$ is a proof of $A \supset B$, and $arg_2(x)$ is a proof of A, so by induction hypothesis, $F(arg_1(x))$ is a proof of $\mathrm{Pr}(\ulcorner A \supset B \urcorner)$ and $F(arg_2(x))$ is a proof of $\mathrm{Pr}(\ulcorner A \urcorner)$, so $F(x) = mp(F(arg_1(x)), F(arg_2(x)))$ is a proof of $\mathrm{Pr}(\ulcorner B \urcorner)$, by the defining equation of mp.

Verification of D2 continued

- There is a similar case for each rule of inference
- The "stopping cases" for F are when x is a one-step proof by an axiom.
- The induction step says that within T, one step of a proof (as encoded by Gödel numbers) does follow the rules for proofs.
- That is, the formula representing the proof predicate and the function F are related, by certain recursion relations.
- The only way this could fail is a "bug" in the definition of F or the definition of the Prf predicate. The condition D2 expresses a fundamental property of the formula Prf that it must satisfy in order to be what we think of as a "formalization of the proof predicate."

What we needed to verify D1-D3

- The formula representing the proof predicate satisfies its defining recursion equations.
- Proofs can be verified to be proofs.
- Both those are proved by course-of-values induction.
- So, we need a couple of simple instances of course-of-values induction plus $\mathbf{RA}$.
- It seems nobody has isolated an exactly "minimal" system, partly out of concern that if you take too few axioms away, you can't really ensure that your Prf predicate represents what you have in mind by proofs, so maybe the formula you have written down doesn't even adequately reflect the second incompleteness theorem.
- The interest of the second incompleteness theorem is that it applies to the *strongest* theories we know. It's not so interesting exactly how *weak* a theory it will still work for.

What about a computer-checked proof?

- There does exist a computer-checked proof of the First Incompleteness Theorem (by Natarajan Shankar in the proof-checker PVS), but for technical reasons it can't be automatically converted to a proof in $\mathbf{PA}$.
- So we don't yet have a computer-checked proof of the Second Incompleteness Theorem.
- That may change soon: I heard a rumor that one will be announced this summer in Vienna, where the Summer of Logic will take place.