The second incompleteness theorem

- One aspect of the proof of Gödel's first incompleteness theorem was the mixture of arguments both inside and outside the axiomatic system we were working in. This would have disturbed Hilbert and others who were looking for an entirely self-contained demonstration that arithmetic was consistent.
- Gödel also thought that they might latch onto this aspect of the proof to see a possible way around his incompleteness theorem. So he decided to do more.
- The question is going to be: how much does a theory T of the kind we have been looking at - theory of arithmetic, consistent, containing Q - have to know in order to prove its own consistency? What are the ramifications of it knowing about its own consistency?

Coding up the first proof

- At some level, we have been playing with the formula $Prf_T(x)$ given by
$$\exists y \, Prov_T(y, x)$$
which means "there is a T-proof of the formula encoded by x". Notice that this is Σ_1
- Recall that we had two formulas which we used in the proof of the first theorem; $\psi(x)$ and G_T given by
$$\forall y \, \neg Gdl(y, x) \text{ and } \psi(\ulcorner \psi \urcorner).$$
$\psi(x)$ in particular is logically equivalent to
$$\forall y \, (Diag(x, y) \rightarrow \neg Prf_T(y)).$$
- Now Diag is a p.r. relation captured by Q and so T proves $Diag(\ulcorner \psi \urcorner, \ulcorner G_T \urcorner)$.
- The conclusion is that T proves that G_T is equivalent to $\neg Prf_T(\ulcorner G_T \urcorner)$.

Consistency

- We now introduce a sentence $Con_T := \neg Prf_T(0 = 1)$. Definitely if T is consistent and contains Q then T does not prove 0 = 1.
- Now comes the key point: we know that if T is an effectively enumerable, consistent theory of arithmetic containing Q then T does not prove G_T. We would like to formalize this *inside* the theory of T itself. That is, we would like T to be strong enough so that
$$T \vdash Con_T \leftrightarrow \neg Prf_T(\lceil G_T \rceil).$$
- We will come back to this but let's finish from here: Suppose that $T \vdash Con_T$ i.e. T knows that it is consistent. Then $T \vdash \neg Prf_T(\lceil G_T \rceil)$ and so T proves G_T which would mean that T is inconsistent! This is often said succinctly as "Any theory strong enough to know its own consistency is itself inconsistent."

Working inside the system

- We now need to go back and figure out what it means to work "inside" T to prove the first incompleteness theorem. We will only sketch this and you can read the details in Smith.
- How did the proof of the first incompleteness theorem go at a high level? From an early stage we knew we needed the theory to be sufficiently nice and ultimately we settled on the idea that the theory captured all Σ_1 formulas. This allowed us to build a Gödel sentence.
- We argued outside the system that our theory T captured various Σ_1-formulas. How did we do that? Again, at a high level, we mostly argued by induction that this or that formula or the numbers we substituted had certain requisite properties.

It is not then a big leap to believe that what we need to do is to formalize the notion that we can do induction on Σ_1-formulas. We define the fragment of Peano arithmetic which involves only Σ_1-formulas by $I\Sigma_1$; it has the axiom schema:
If $\varphi(x)$ is Σ_1 then we assume the axiom

$$(\varphi(0) \land \forall x\,(\varphi(x) \rightarrow \varphi(S(x)))) \rightarrow \forall x\,\varphi(x).$$

Theorem (Gödel's second incompleteness theorem)

If T is effectively enumerable, consistent and contains Q and $I\Sigma_1$ then $T \nvdash Con_T$.