Diagonalization

- You knew this was going to come up! What does diagonalization look like in this context?
- Suppose we are given a formula φ with one free variable x; remember for this we write $\varphi(x)$.
- We can compute the Gödel number of this formula $\ulcorner\varphi(x)\urcorner$ and then we can plug it back into the formula itself to form

  $$\varphi(\ulcorner\varphi(x)\urcorner).$$

  We call this new formula the diagonalization of φ.
- To be fair, we aren't really plugging the number into φ because that is something that happens when we interpret the formula in N.
- We are actually plugging the term associated to the number $\ulcorner\varphi(x)\urcorner$ into the formula. For instance, if the Gödel number of φ is 2 (very unlikely) then we plug in the term $SS0$. Smith puts bars over the numbers when he does this but the notation is going to get cluttered enough.

The Gödel sentence

- Let's use the diagonalization right away. Fix a theory T as before, one for which it is possible to determine if a given formula is an axiom of T. We are going to create a formula $Gdl_T(m, n)$.
- $Gdl_T(m, n)$ is supposed to hold if n is the Gödel number of some formula $\varphi(x)$ and m is the Gödel number of a proof from T of $\varphi(\lceil\varphi(x)\rceil)$, the diagonalization of φ.
- Now create the formula

$$\psi(y) := \forall x \neg Gdl_T(x, y) \text{ and let } G_T := \psi(\lceil\psi(x)\rceil).$$

- This is all very convoluted and there are many things to say but probably the most important is: if T is consistent (can't prove false) then T cannot prove G_T and G_T is true!

What the ...?

- Let me quickly sketch a proof of that last claim. Suppose T could prove G_T. Then there is a proof in our proof system of this fact and it has some Gödel number n. That means that $Gdl_T(n, \ulcorner\psi(x)\urcorner)$ holds.
- But T proves $\psi(\ulcorner\psi(x)\urcorner)$ which means that T proves $\forall x \neg Gdl_T(x, \ulcorner\psi(x)\urcorner)$ and then, in particular, T proves $\neg Gdl_T(n, \ulcorner\psi(x)\urcorner)$. This is a contradiction to the consistency of T.
- However, if G_T is false then $\exists x\, Gdl_T(x, \ulcorner\psi\urcorner)$ holds i.e. there is a proof of G_T from T which was just showed was false if T is consistent. So G_T must be true!
- Amazing!

What did we just prove?

- We assumed that we had a consistent theory of arithmetic T (all its axioms were true in N), like Peano arithmetic. Implicitly we were using the fact that T was strong enough to prove the axioms for the little system Q. We needed this so that T did the right thing for Σ_1-formulas. We were also assuming there was some p.r. way that we could recognize the axioms of T.
- To be quite precise, what we want is that there is a Σ_1-formulua $Prov_T(x, y)$ such that if $Prov_T(m, n)$ holds then
  1. m codes a sequence of formulas $\varphi_0, \ldots, \varphi_k$,
  2. n codes φ_k, and
  3. $\varphi_0, \ldots, \varphi_k$ is a T-proof of φ_k.
- So if $\varphi_0, \ldots, \varphi_k$ is a T-proof of φ_k with codes as above then $Prov_T(m, n)$ will hold and T will prove $Prov_T(m, n)$ since $Prov_T$ is a Σ_1-formula.