

inp.1 The Fixed-Point Lemma

inc:inp:fix:
sec The fixed-point lemma says that for any formula $\psi(x)$, there is a sentence φ explanation such that $\mathbf{T} \vdash \varphi \leftrightarrow \psi(\ulcorner \varphi \urcorner)$, provided \mathbf{T} extends \mathbf{Q} . In the case of the liar sentence, we'd want φ to be equivalent (provably in \mathbf{T}) to “ $\ulcorner \varphi \urcorner$ is false,” i.e., the statement that $\# \varphi^\#$ is the Gödel number of a false sentence. To understand the idea of the proof, it will be useful to compare it with Quine's informal gloss of φ as, “yields a falsehood when preceded by its own quotation” yields a falsehood when preceded by its own quotation.” The operation of taking an expression, and then forming a sentence by preceding this expression by its own quotation may be called *diagonalizing* the expression, and the result its diagonalization. So, the diagonalization of ‘yields a falsehood when preceded by its own quotation’ is “yields a falsehood when preceded by its own quotation” yields a falsehood when preceded by its own quotation.” Now note that Quine's liar sentence is not the diagonalization of ‘yields a falsehood’ but of ‘yields a falsehood when preceded by its own quotation.’ So the property being diagonalized to yield the liar sentence itself involves diagonalization!

In the language of arithmetic, we form quotations of a formula with one free variable by computing its Gödel numbers and then substituting the standard numeral for that Gödel number into the free variable. The diagonalization of $\alpha(x)$ is $\alpha(\bar{n})$, where $n = \# \alpha(x)^\#$. (From now on, let's abbreviate $\# \alpha(x)^\#$ as $\ulcorner \alpha(x) \urcorner$.) So if $\psi(x)$ is “is a falsehood,” then “yields a falsehood if preceded by its own quotation,” would be “yields a falsehood when applied to the Gödel number of its diagonalization.” If we had a symbol *diag* for the function $\text{diag}(n)$ which computes the Gödel number of the diagonalization of the formula with Gödel number n , we could write $\alpha(x)$ as $\psi(\text{diag}(x))$. And Quine's version of the liar sentence would then be the diagonalization of it, i.e., $\alpha(\ulcorner \alpha \urcorner)$ or $\psi(\text{diag}(\ulcorner \psi(\text{diag}(x)) \urcorner))$. Of course, $\psi(x)$ could now be any other property, and the same construction would work. For the incompleteness theorem, we'll take $\psi(x)$ to be “ x is unprovable in \mathbf{T} .” Then $\alpha(x)$ would be “yields a sentence unprovable in \mathbf{T} when applied to the Gödel number of its diagonalization.”

To formalize this in \mathbf{T} , we have to find a way to formalize *diag*. The function $\text{diag}(n)$ is computable, in fact, it is primitive recursive: if n is the Gödel number of a formula $\alpha(x)$, $\text{diag}(n)$ returns the Gödel number of $\alpha(\ulcorner \alpha(x) \urcorner)$. (Recall, $\ulcorner \alpha(x) \urcorner$ is the standard numeral of the Gödel number of $\alpha(x)$, i.e., $\# \alpha(x)^\#$). If *diag* were a function symbol in \mathbf{T} representing the function *diag*, we could take φ to be the formula $\psi(\text{diag}(\ulcorner \psi(\text{diag}(x)) \urcorner))$. Notice that

$$\begin{aligned} \text{diag}(\# \psi(\text{diag}(x))^\#) &= \# \psi(\text{diag}(\ulcorner \psi(\text{diag}(x)) \urcorner))^\# \\ &= \# \varphi^\#. \end{aligned}$$

Assuming \mathbf{T} can prove

$$\text{diag}(\ulcorner \psi(\text{diag}(x)) \urcorner) = \ulcorner \varphi \urcorner,$$

it can prove $\psi(\text{diag}(\ulcorner \psi(\text{diag}(x)) \urcorner)) \leftrightarrow \psi(\ulcorner \varphi \urcorner)$. But the left hand side is, by definition, φ .

Of course, $diag$ will in general not be a function symbol of \mathbf{T} , and certainly is not one of \mathbf{Q} . But, since $diag$ is computable, it is *representable* in \mathbf{Q} by some formula $\theta_{diag}(x, y)$. So instead of writing $\psi(diag(x))$ we can write $\exists y (\theta_{diag}(x, y) \wedge \psi(y))$. Otherwise, the proof sketched above goes through, and in fact, it goes through already in \mathbf{Q} .

Lemma inp.1. *Let $\psi(x)$ be any formula with one free variable x . Then there is a sentence φ such that $\mathbf{Q} \vdash \varphi \leftrightarrow \psi(\ulcorner \varphi \urcorner)$.*

[inc:inp:fix:](#)
[lem:fixed-point](#)

Proof. Given $\psi(x)$, let $\alpha(x)$ be the formula $\exists y (\theta_{diag}(x, y) \wedge \psi(y))$ and let φ be its diagonalization, i.e., the formula $\alpha(\ulcorner \alpha(x) \urcorner)$.

Since θ_{diag} represents $diag$, and $diag(\# \alpha(x) \#) = \# \varphi \#$, \mathbf{Q} can prove

$$\theta_{diag}(\ulcorner \alpha(x) \urcorner, \ulcorner \varphi \urcorner) \tag{1} \quad \text{inc:inp:fix:}$$

$$\forall y (\theta_{diag}(\ulcorner \alpha(x) \urcorner, y) \rightarrow y = \ulcorner \varphi \urcorner). \tag{2} \quad \text{repdiag1}$$

[inc:inp:fix:](#)
[repdiag1](#)
[inc:inp:fix:](#)
[repdiag2](#)

Now we show that $\mathbf{Q} \vdash \varphi \leftrightarrow \psi(\ulcorner \varphi \urcorner)$. We argue informally, using just logic and facts provable in \mathbf{Q} .

First, suppose φ , i.e., $\alpha(\ulcorner \alpha(x) \urcorner)$. Going back to the definition of $\alpha(x)$, we see that $\alpha(\ulcorner \alpha(x) \urcorner)$ just is

$$\exists y (\theta_{diag}(\ulcorner \alpha(x) \urcorner, y) \wedge \psi(y)).$$

Consider such a y . Since $\theta_{diag}(\ulcorner \alpha(x) \urcorner, y)$, by eq. (2), $y = \ulcorner \varphi \urcorner$. So, from $\psi(y)$ we have $\psi(\ulcorner \varphi \urcorner)$.

Now suppose $\psi(\ulcorner \varphi \urcorner)$. By eq. (1), we have $\theta_{diag}(\ulcorner \alpha(x) \urcorner, \ulcorner \varphi \urcorner) \wedge \psi(\ulcorner \varphi \urcorner)$. It follows that $\exists y (\theta_{diag}(\ulcorner \alpha(x) \urcorner, y) \wedge \psi(y))$. But that's just $\alpha(\ulcorner \alpha(x) \urcorner)$, i.e., φ . \square

[digression](#)

You should compare this to the proof of the fixed-point lemma in computability theory. The difference is that here we want to define a *statement* in terms of itself, whereas there we wanted to define a *function* in terms of itself; this difference aside, it is really the same idea.

Photo Credits

Bibliography