The fixed-point lemma says that for any formula $\psi(x)$, there is a sentence $\varphi$ such that $T \vdash \varphi \leftrightarrow \psi(\#\varphi)$, provided $T$ extends $Q$. In the case of the liar sentence, we’d want $\varphi$ to be equivalent (provably in $T$) to “$\#\varphi$ 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 $\varphi$ 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’ 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 $\alpha(\#\alpha(x))$, where $\#\alpha(x)$ is the standard numeral of the Gödel number of $\alpha(x)$, i.e., $\#\alpha(x)$. 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 $\text{diag}$ 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., $\alpha(\#\alpha(x))$ or $\psi(\text{diag}(\#\psi(\text{diag}(x)))$. 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 not derivable in $T$.” Then $\alpha(x)$ would be “yields a sentence not derivable in $T$ when applied to the Gödel number of its diagonalization.”

To formalize this in $T$, we have to find a way to formalize $\text{diag}$. 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 $\alpha(\#\psi(\text{diag}(x)))$. (Recall, $\#\alpha(x)$ is the standard numeral of the Gödel number of $\alpha(x)$, i.e., $\#\alpha(x)$. If $\text{diag}$ were a function symbol in $T$ representing the function $\text{diag}$, we could take $\varphi$ to be the formula $\psi(\text{diag}(\#\psi(\text{diag}(x))))$. Notice that

$$\text{diag}(\#\psi(\text{diag}(x))) = \#\psi(\text{diag}(\#\psi(\text{diag}(x))))$$

Assuming $T$ can derive

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

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

*fixed-point-lemma rev: f9d72b0 (2019-05-06) by OLP / CC–BY*
Of course, $diag$ will in general not be a function symbol of $T$, and certainly is not one of $Q$. But, since $diag$ is computable, it is \emph{representable} in $Q$ by some formula $\theta_{diag}(x, y)$. So instead of writing $\psi(diag(x))$ we can write $\exists y (\theta_{diag}(x, y) \land \psi(y))$. Otherwise, the proof sketched above goes through, and in fact, it goes through already in $Q$.

**Lemma inp.1.** Let $\psi(x)$ be any formula with one free variable $x$. Then there is a sentence $\varphi$ such that $Q \vdash \varphi \leftrightarrow \psi(\langle \varphi \rangle)$.

\textbf{Proof.} Given $\psi(x)$, let $\alpha(x)$ be the formula $\exists y (\theta_{diag}(x, y) \land \psi(y))$ and let $\varphi$ be its diagonalization, i.e., the formula $\alpha(\langle \alpha(x) \rangle)$.

Since $\theta_{diag}$ represents $diag$, and $diag(\alpha(x)) = \varphi$, $Q$ can derive
\begin{align*}
\theta_{diag}(\langle \alpha(x) \rangle, \langle \varphi \rangle) \\
\forall y (\theta_{diag}(\langle \alpha(x) \rangle, y) \rightarrow y = \langle \varphi \rangle). \tag{1}
\end{align*}

Now we show that $Q \vdash \varphi \leftrightarrow \psi(\langle \varphi \rangle)$. We argue informally, using just logic and facts \emph{derivable} in $Q$.

First, suppose $\varphi$, i.e., $\alpha(\langle \alpha(x) \rangle)$. Going back to the definition of $\alpha(x)$, we see that $\alpha(\langle \alpha(x) \rangle)$ just is
\[
\exists y (\theta_{diag}(\langle \alpha(x) \rangle, y) \land \psi(y)).
\]
Consider such a $y$. Since $\theta_{diag}(\langle \alpha(x) \rangle, y)$, by eq. (2), $y = \langle \varphi \rangle$. So, from $\psi(y)$ we have $\psi(\langle \varphi \rangle)$.

Now suppose $\psi(\langle \varphi \rangle)$. By eq. (1), we have $\theta_{diag}(\langle \alpha(x) \rangle, \langle \varphi \rangle) \land \psi(\langle \varphi \rangle)$. It follows that $\exists y (\theta_{diag}(\langle \alpha(x) \rangle, y) \land \psi(y))$. But that’s just $\alpha(\langle \alpha(x) \rangle)$, i.e., $\varphi$. \hfill \square

\textbf{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 \emph{statement} in terms of itself, whereas there we wanted to define a \emph{function} in terms of itself; this difference aside, it is really the same idea.

\textbf{Photo Credits}

\textbf{Bibliography}