The fixed-point lemma says that for any formula $\psi(x)$, there is a sentence $\varphi$ such that $T \vdash \varphi \leftrightarrow \psi(\langle \varphi \rangle)$, provided $T$ extends $Q$. In the case of the liar sentence, we'd want $\varphi$ to be equivalent (provably in $T$) to "$\langle \varphi \rangle$ 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 $\alpha(\langle \pi \rangle)$, where $n = \# \alpha(x)$. (From now on, let’s abbreviate $\# \alpha(x)$ as $'\alpha(x)'$.) So if $\psi(x)$ is "is a falsehood," then "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., $\alpha(\langle \alpha(x) \rangle)$ or $\psi(\text{diag}(\langle \psi(\text{diag}(x)) \rangle))$. 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 $\alpha(\langle \alpha(x) \rangle)$. (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 $\psi(\text{diag}(\langle \psi(\text{diag}(x)) \rangle))$. Notice that

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

$$= \# \varphi.$$

Assuming $T$ can derive

$$\text{diag}(\langle \psi(\text{diag}(x)) \rangle) = '\varphi',$$

it can derive $\psi(\text{diag}(\langle \psi(\text{diag}(x)) \rangle)) \leftrightarrow \psi('\varphi').$ But the left hand side is, by definition, $\varphi.$
Of course, \( \text{diag} \) will in general not be a function symbol of \( T \), and certainly is not one of \( Q \). But, since \( \text{diag} \) is computable, it is \textit{representable} in \( Q \) by some formula \( \theta_{\text{diag}}(x, y) \). So instead of writing \( \psi(\text{diag}(x)) \) we can write \( \exists y (\theta_{\text{diag}}(x, y) \land \psi(y)) \). 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 \( Q \vdash \varphi \iff \psi(\varphi) \).

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

Since \( \theta_{\text{diag}} \) represents \( \text{diag} \), and \( \text{diag}(\alpha(x)) = \varphi \), \( Q \) can derive

\[
\theta_{\text{diag}}(\alpha(x), \varphi) \quad (1)
\]

\[
\forall y (\theta_{\text{diag}}(\alpha(x), y) \rightarrow y = \varphi) \quad (2)
\]

Now we show that \( Q \vdash \varphi \iff \psi(\varphi) \). We argue informally, using just logic and facts \textit{derivable} in \( Q \).

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

\[
\exists y (\theta_{\text{diag}}(\alpha(x), y) \land \psi(y)).
\]

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

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

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

**Photo Credits**

**Bibliography**