The fixed-point lemma says that for any formula \( \psi(x) \), there is a sentence \( \varphi \) such that \( T \vdash \varphi \leftrightarrow \psi(\varphi^\gamma) \), provided \( T \) extends \( Q \). In the case of the liar sentence, we’d want \( \varphi \) to be equivalent (provably in \( T \)) to “\( \varphi^\gamma \) 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’ 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 \( \alpha(\#) \), where \( \# = \#\alpha(x)^\# \). (From now on, let’s abbreviate \( \#\alpha(x)^\# \) as \( \#\alpha^\# \).) 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 \( \text{diag} \) 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 \( \psi(\text{diag}(\#)) \). And Quine’s version of the liar sentence would then be the diagonalization of it, i.e., \( \alpha(\#\alpha^\#) \) or \( \psi(\text{diag}(\#\psi(\text{diag}(\#)))) \). 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 \( \alpha(\#\psi(\text{diag}(\#))) \). (Recall, \( \#\alpha(x)^\# \) is the standard numeral of the Gödel number of \( \alpha(x) \), i.e., \( \#\alpha^\# \).) If \( \text{diag} \) were a function symbol in \( T \) representing 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 \( \psi(\text{diag}(\#)) \). Notice that

\[
\text{diag}(\#\psi(\text{diag}(\#))) = \#\psi(\text{diag}(\#\psi(\text{diag}(\#))))^\#
\]

Assuming \( T \) can derive

\[
\text{diag}(\#\psi(\text{diag}(\#)))^\gamma = \varphi^\gamma,
\]

it can derive \( \psi(\text{diag}(\#\psi(\text{diag}(\#)))) \leftrightarrow \psi(\#\varphi^\gamma) \). 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 \( \exists y \left( \theta_{\text{diag}}(x, y) \land \psi(y) \right) \). Otherwise, the proof sketched above goes through, and in fact, it goes through already in \( Q \).

\begin{lemma} \text{inp.1.} Let \( \psi(x) \) be any formula with one free variable \( x \). Then there is a sentence \( \varphi \) such that \( Q \vdash \varphi \iff \psi(\varphi^\gamma) \).
\end{lemma}

\begin{proof}
Given \( \psi(x) \), let \( \alpha(x) \) be the formula \( \exists y \left( \theta_{\text{diag}}(x, y) \land \psi(y) \right) \) and let \( \varphi \) be its diagonalization, i.e., the formula \( \alpha(\varphi^\gamma) \).

Since \( \theta_{\text{diag}} \) represents \( \text{diag} \), and \( \text{diag}(\# \alpha(x)) = \# \varphi^\gamma \), \( Q \) can derive
\begin{align*}
\theta_{\text{diag}}(\alpha^\gamma, \varphi^\gamma) \quad &\text{(1)} \\
\forall y \left( \theta_{\text{diag}}(\alpha^\gamma, y) \to y = \varphi^\gamma \right) \quad &\text{(2)}
\end{align*}

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

First, suppose \( \varphi \), i.e., \( \alpha(\varphi^\gamma) \). Going back to the definition of \( \alpha(x) \), we see that \( \alpha(\varphi^\gamma) \) just is
\[ \exists y \left( \theta_{\text{diag}}(\alpha^\gamma, y) \land \psi(y) \right). \]

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

Now suppose \( \psi(\varphi^\gamma) \). By eq. (1), we have
\[ \theta_{\text{diag}}(\alpha^\gamma, \varphi^\gamma) \land \psi(\varphi^\gamma). \]

It follows that
\[ \exists y \left( \theta_{\text{diag}}(\alpha^\gamma, y) \land \psi(y) \right). \]

But that’s just \( \alpha(\varphi^\gamma) \), i.e., \( \varphi \).
\end{proof}

\begin{problem} \text{inp.1.} A formula \( \varphi(x) \) is a \textit{truth definition} if \( Q \vdash \psi \iff \varphi(\psi^\gamma) \) for all sentences \( \psi \). Show that no formula is a truth definition by using the fixed-point lemma.
\end{problem}

\section*{Photo Credits}

\section*{Bibliography}