

Exercise [16.18]

My answer to this exercise must be "NO" (I could not establish the assertion by myself).

I therefore looked for the proof in Penrose's book "The Emperor's New Mind", where it is however expressed with another formalism (without Turing machines). The central idea seems to be a sentence like "*This sentence cannot be proven in the formal system F* ".

So here is my proposal how to express this with Turing machines:

Assume that \mathbf{P} is a Turing Machine¹ that proves all calculable Π_1 -sentences of the form

$$"T_w(w) \text{ does not terminate}",$$

i.e. the result of \mathbf{P} is "1" if the sentence is true. If the Π_1 -sentence is false or if \mathbf{P} cannot handle it, \mathbf{P} shall not terminate². This can be formulated as follows:

$$P(w) = \begin{cases} 1 & \Rightarrow T_w(w) \text{ does not terminate} \\ \text{does not terminate} & \Rightarrow T_w(w) \text{ terminates or not} \end{cases}$$

As \mathbf{P} is a Turing Machine, it must appear in the list of all Turing Machines, i.e. there must be some index k such that $\mathbf{P} = \mathbf{T}_k$. The Π_1 -sentence

$$"T_k(k) \text{ does not terminate}" \tag{\#}$$

is then my candidate for the desired $G(\mathbf{F})$ sentence:

- First, this sentence (#) cannot be proven by \mathbf{P} , i.e. we have $\mathbf{P}(k) \neq 1$, because $\mathbf{P}(k) = 1$ would require that $\mathbf{P}(k)$ terminates. As $\mathbf{P}(k) = \mathbf{T}_k(k)$, this would however be a contradiction to what was allegedly proven.
- Accordingly, $\mathbf{P}(k)$ does not terminate, i.e. the Proof-Machine \mathbf{P} yields no result for this Π_1 -sentence. However, the fact that " $\mathbf{P}(k)$ does not terminate" just shows the truth of the Π_1 -sentence (#) because $\mathbf{P}(k) = \mathbf{T}_k(k)$.

Is there any flaw in these considerations? I am grateful for any comments!

¹ I think that it is not necessary to consider several of such Turing Machines to cover all calculable Π_1 -sentences because these could be combined into one Machine.

² If \mathbf{P} should in some circumstances be able to find out that $T_w(w)$ terminates, this can be ignored and \mathbf{P} can be driven into an endless loop.