Evgeny Chutchev  
Some Mathematicians Are Not Turing Machines

I. Introduction
A certain mathematician M, considering some hypothesis H, conclusion C and text P, can arrive at one of the following judgments:

- P does not convince M of the fact that since H, it follows that C (judgment of the type "Not Proved").
- P is the proof that since H, it follows that C (judgment of the type "Proved").

Let us examine the question: Is it possible to replace such a mathematician with an arbitrary Turing machine? In order to answer this question soundly, we shall formalize the concepts of "hypothesis", "conclusion" and "proof" and show that the answer to the question is negative under the two following conditions:

1. M is faultless, namely his judgment "Proved" always implies that since H, it actually follows that C.
2. M recognizes a certain P as the correct proof of the fact that for certain ℋ and C, if ℋ, then C (where P, ℋ, and C are stated below).

The reason for negative answer lies in the contents of ℋ and C, whose sense can be expressed informally in the following words:

- ℋ = "Mathematician M, who operates as Turing machine, is faultless"
- C = "Mathematician M, who operates as Turing machine, cannot recognize any proof that since ℋ, it follows that C".

The rest of the paper is organized as follows:
Sections II and III give definitions and introduce notations.
Sections IV and V contain the proof of the main statement.
Section VI provides examples related to the concept of "faultless mathematician".
Section VII presents the conclusion.
Section VIII contains a list of definitions and notations.

II. General definitions and notations
For the purposes of formalization, declared in Section I, we introduce the following definitions and notations.

II.1. Alphabets and symbols
1. Suppose 𝔹 is a finite alphabet and symbol θ does not belong to 𝔹. By 𝔹* denote all words over 𝔹. Put 𝔹 ≝ 𝔹 ∪ {θ}, and denote by 𝔽 the set of all words over 𝔽. We assume that all expressions, mentioned in this paper, are expressions over the alphabet 𝔽.
2. For 𝑉 ⊆ 𝔽 and 𝑈 ⊆ 𝔽*, put 𝑉 ∪ 𝑈 ≝ {𝑣θ𝑢 : 𝑣 ∈ 𝑉 , 𝑢 ∈ 𝑈}.

II.2. Turing machines
1. We consider Turing machines (TMs) with the tape alphabet 𝔽 and denote by ℳ the set of all TMs, assuming ℳ ⊆ 𝔽*.
2. For TM 𝔿 and 𝑏 ∈ 𝔽*, denote by [𝔿, 𝑏] the situation where TM 𝔿 starts on the initial contents of the tape equal to 𝑏. We write [𝔿, 𝑏] ≠ ∞, if [𝔿, 𝑏] will eventually halt. Otherwise, we write [𝔿, 𝑏] = ∞.

II.3. Enumerable sets
1. According to the conventional terminology, we call the set 𝔹 ⊆ 𝔽* enumerable, if there exists an arbitrary TM 𝔽E, such that ∀𝑥((𝑥 ∈ 𝔹) ⇔ ([𝔽E, 𝑥] ≠ ∞)).
   Note the following:
   - The sets 𝔹*, 𝔽*, and ℳ are enumerable.
   - If sets 𝑉 ⊆ 𝔹* and 𝑈 ⊆ 𝔽* are enumerable, then the set 𝑉 ∪ 𝑈 is enumerable.
2. For some symbol 𝔿 and enumerable set 𝔹, denote by 𝔿|&display formula| is the description of how TM 𝔿 works in situations [𝔿, 𝑏] for all 𝑏 ∈ 𝔹. In this case, the general description of 𝔿 is as follows:
   - Read the word 𝑏 ∈ 𝔽*, recorded at the beginning of the tape.
• Start up \([\mathcal{C}_E, b]\).
• If \(\mathcal{C}_E\) halts, start up \([\mathfrak{M}, b]\).
Note that \(\forall x \in \mathbb{E} \left( [\mathfrak{M}, x] = \infty \right)\) for TM \(\mathfrak{M}\) given by \(\mathfrak{M}|_E\).

II.4. Special Turing machines
1. Along with TMs, introduced in Subsection II.2.1, we consider special TMs (STMs). The only difference between TM and STM is that STM has two final states, one marked with "Proved" and another marked with "Not Proved".
2. With each STM \(\mathfrak{S}\) we associate TM \(\mathfrak{S}\), which is obtained by replacing the final state of \(\mathfrak{S}\), marked with "Not Proved", with an infinite loop.

II.5. Abridged notations
1. An expression of the type "TM \(\mathfrak{M}\) can be constructed" will mean that in the proof of assertion the algorithm of constructing is given.
2. In constructions of the form \(\text{Text} \Rightarrow \text{Text} \equiv \text{Text} \})\) and \(\text{Text} \equiv \text{Text} \})\), the "Text" will serve as the substantiation of the corresponding implication or equivalence. Moreover, if "Text" includes a certain designation \(d\) within the framework \([d]\), then this fragment of the "Text" should be read as "Owing to the definition of \(d\".

III. Theorems, mathematicians, and provability

III.1. Definitions
We introduce the following definitions, which formalize the actions of mathematicians.
1. Definitions of sets:
   • By abstract theorem (AT) denote any element of the set \(\mathbb{T} = \mathbb{H} \theta \mathbb{C}\), where \(\mathbb{H}\) and \(\mathbb{C}\) are stated below.
   • By hypothesis of AT denote any element of the set \(\mathbb{H} = \mathbb{A}\).
   • By conclusion of AT denote any element of the set \(\mathbb{C} = \mathbb{A}\).
   • By proof of AT denote any element of the set \(\mathbb{P} = \mathbb{B}\).
   We use the symbols \(\mathbb{H}, \mathbb{C}, \) and \(\mathbb{P}\) to emphasize that the set \(\mathbb{A}\) (or the set \(\mathbb{B}\)) is considered as the carrier, correspondingly, of hypotheses, conclusions, or proofs of ATs.
2. By abstract mathematician (AM) denote any map \(\mu: \mathbb{T} \theta \mathbb{P} \rightarrow \{0, 1\}\), assuming that for \(t \in \mathbb{T}\) and \(p \in \mathbb{P}\) the equality \(\mu(t \theta p) = 1\) means that \(\mu\) recognizes the word \(p\) as the proof of \(t\).
3. With every TM \(\mathfrak{M}\) associate the AM \(\mathfrak{S}\) according to the following rule:
\[
\forall t \in \mathbb{T} \forall p \in \mathbb{P} \left( [\mathfrak{M}(t \theta p) = 1] \Leftrightarrow ([\mathfrak{M}, t \theta p] \neq \infty) \right).
\]
4. Extending the concepts from Subsection II.2.2 to STM, associate with each STM \(\mathfrak{S}\) the AM \(\mathfrak{S}\) according to the following rule:
\[
\forall t \in \mathbb{T} \forall p \in \mathbb{P} \left( ([\mathfrak{S}(t \theta p) = 1] \Leftrightarrow \left\{ \begin{array}{l}
\text{In situation} [\mathfrak{S}, t \theta p] \text{STM} \mathfrak{S} \text{will reach the final state marked with "Proved"}
\end{array} \right. \right).
\]
5. We consider:
   • Word \(p \in \mathbb{P}\) as the proof of AT \(t\) for AM \(\mu\) and denote this by \(p \rightarrow t\), if \(\mu(t \theta p) = 1\).
   • AT \(t\) as provable by AM \(\mu\) and denote this by \(\mu \vdash t\), if \(\exists p \in \mathbb{P} (p \rightarrow t)\).

III.2. Faultless mathematicians
To introduce the concept of faultless AM we assume that every \(a \in \mathbb{A}\) has one of the following options:
• Word \(a\) is a true statement (in short notation, \(a = t\)).
• Word \(a\) is a false statement (in short notation, \(a = f\)).
• Word \(a\) can be neither true nor false (in short notation, \(a = u\)), that is, either \(a\) is not a statement or, being a statement, contains an internal contradiction.
Using the notation of the form \(a = x\) for a specific word \(a\), we assume that the meaning of such a word is determined by the context of this paper.
Now we call an AM \(\mu\) faultless if \(\phi_{\mu} = t\), where
\[
\phi_{\mu} \equiv \forall p \in \mathbb{P} \forall t \in \mathbb{T} \left( (p \rightarrow t) \Rightarrow ((H_t = t) \Rightarrow (C_t = t)) \right) \quad \text{(III.2.1)}
\]
and \( t = H_t \theta \mathcal{G}_t \) for \( t \in \mathbb{T} \).

Examples of TM \( \mathcal{W} \), for which \( \Phi_{\mathcal{W}} = t, \Phi_{\mathcal{W}} = \emptyset \), and \( \Phi_{\mathcal{W}} = u \), are given in Section VI.

**IV. Theorem**

The main statement of this paper is the corollary of the theorem proved below. To formulate the theorem and its corollary we denote by \( \sigma_{\mathcal{R}} \) the statement \([\mathcal{R}, \mathcal{R}] = \infty\) for an arbitrary TM \( \mathcal{R} \). To prove the theorem and its corollary we rely on the assertion \( \forall_{\mathcal{R} \in \mathcal{M}} \sigma_{\mathcal{R}} \neq u \).

**THEOREM.** For every TM \( \mathcal{W} \), a certain TM \( \mathcal{L}^{(\mathcal{W})} \), for which \( (\Phi_{\mathcal{W}} = t) \Rightarrow (\sigma_{\mathcal{L}^{(\mathcal{W})}} = t) \), can be constructed.

To prove the theorem we shall prove first the following lemmas.

**LEMMA 1.** For every TM \( \mathcal{X} \), a certain TM \( \mathcal{L}_{\mathcal{X}} \), for which \( (\sigma_{\mathcal{L}_{\mathcal{X}}} = \emptyset) \Rightarrow ([\mathcal{X}, \sigma_{\mathcal{L}_{\mathcal{X}}}] \neq \infty) \), can be constructed.

**PROOF.** Describe TM \( \mathcal{L} = \mathcal{L}_{\mathcal{X}} \) by means of \( \mathcal{L}|_{\mathcal{M}} \): in situation \([\mathcal{X}, \mathcal{D}]\), where \( \mathcal{D} \in \mathcal{M} \), TM \( \mathcal{L} \) simulates \([\mathcal{X}, \sigma_{\mathcal{D}}]\). Then \((\sigma_{\mathcal{L}} = \emptyset) \Leftrightarrow ([\mathcal{X}, \emptyset] \neq \infty) \Leftrightarrow ([\mathcal{X}, \sigma_{\mathcal{D}}] \neq \infty).

**LEMMA 2.** A certain TM \( \mathcal{R} \), for which \( \forall_{\mathcal{R} \in \mathcal{M}} \forall_{t \in \mathbb{T}} ([\mathcal{R}, \mathcal{R} \theta t] \neq \infty) \Leftrightarrow (\mathcal{R} \leftarrow t) \), can be constructed.

**PROOF.** Denote by \( \mathbb{N} \) the set of all natural numbers (including zero). We will not make any distinction between the elements of \( \mathbb{N} \) and their notations over the alphabet \( \mathcal{B} \), assuming \( \mathbb{N} = \mathcal{B}^* \). Now describe TM \( \mathcal{R} \) by means of \( \mathcal{R}|_{\mathcal{M} \theta \mathcal{G}} \):

1. The algorithm of \([\mathcal{R}, \mathcal{R} \theta c]\), where \( \mathcal{R} \in \mathcal{M} \) and \( t \in \mathbb{T} \), is divided into steps, enumerated by elements of \( \mathbb{N} \).
2. At the \( k \)-th step \( \mathcal{R} \) simulates \( k - p + 1 \) cycles of \([\mathcal{R}, t \theta p]\) for \( p = 0, \ldots, k \).
3. If in the process of simulation one of the models has reached a final state, \( \mathcal{R} \) halts. Otherwise, \( \mathcal{R} \) proceeds to the next step of its algorithm.

Then \((\mathcal{R}, \mathcal{R} \theta t] \neq \infty) \Leftrightarrow (\exists p \in \mathbb{P}([\mathcal{R}, t \theta p] \neq \infty)) \Leftrightarrow (\exists p \in \mathbb{P} (p \rightarrow t)) \Leftrightarrow (\mathcal{R} \leftarrow t).

**LEMMA 3.** For every TM \( \mathcal{W} \) and every hypothesis \( h \in \mathbb{H} \), a certain TM \( \mathcal{X}_{\mathcal{W}, h} \), for which \( \forall_{c \in \mathbb{C}} ([\mathcal{X}_{\mathcal{W}, h}, c] \neq \infty) \), can be constructed.

**PROOF.** Describe TM \( \mathcal{X} = \mathcal{X}_{\mathcal{W}, h} \) by means of \( \mathcal{X}|_{\mathcal{C}} \): in situation \([\mathcal{X}, c] \), where \( c \in \mathcal{C} \), TM \( \mathcal{X} \) starts up \([\mathcal{W}, \mathcal{W} \theta h \theta c]\). Then for every \( c \in \mathcal{C} \) we have

\[
(\mathcal{X} \leftarrow h \theta c) \Leftrightarrow ([\mathcal{W}, \mathcal{W} \theta h \theta c] \neq \infty) \Leftrightarrow ([\mathcal{X}, c] \neq \infty).
\]

**LEMMA 4.** For every TM \( \mathcal{W} \) and every hypothesis \( h \in \mathbb{H} \), a certain TM \( \mathcal{L}_{\mathcal{W} \theta h} \), for which \( (\sigma_{\mathcal{L}_{\mathcal{W} \theta h}} = \emptyset) \Rightarrow (\mathcal{R} \leftarrow h \theta \sigma_{\mathcal{L}_{\mathcal{W} \theta h}}) \), can be constructed.

**PROOF.** By the algorithm from the proof of Lemma 3 construct for \( \mathcal{W} \) and \( h \) TM \( \mathcal{X} = \mathcal{X}_{\mathcal{W}, h} \). Thereafter by the algorithm from the proof of Lemma 1 construct for \( \mathcal{X} \) TM \( \mathcal{L} = \mathcal{L}_{\mathcal{W}, h} \). Then \((\sigma_{\mathcal{L}} = \emptyset) \Leftrightarrow ([\mathcal{X}, \sigma_{\mathcal{L}}] \neq \infty) \Leftrightarrow (\mathcal{R} \leftarrow h \theta \sigma_{\mathcal{L}}).

**PROOF OF THE THEOREM.** For TM \( \mathcal{W} \) and hypothesis \( H = \Phi_{\mathcal{W}} \), construct by the algorithm from the proof of Lemma 4 TM \( \mathcal{L}^{(\mathcal{W})} = \mathcal{L}_{\mathcal{W} \theta h} \) and introduce the notation \( \sigma = \sigma_{\mathcal{L}^{(\mathcal{W})}} \). Then

\[
(\sigma = \emptyset) \xrightarrow{\text{Lemma 4}} (\mathcal{R} \leftarrow H \theta \sigma) \Leftrightarrow (\exists p \in \mathbb{P} (p \rightarrow H \theta \sigma)).
\]
Thus for $\sigma = t$ and $\pi$, for which $\exists \sigma : \pi \rightarrow H\theta \sigma$,

$$(H = t) \equiv \left(\left(\pi \rightarrow H\theta \sigma\right) \Rightarrow \left((H = t) \Rightarrow (\sigma = t)\right)\right) \quad \text{Owing to } H = t$$

\[ \text{V. Corollary of the theorem} \]

To formulate and prove the corollary of the theorem we introduce the following additional definitions:

1. Denote hypothesis and conclusion of the theorem from Section IV by Theorem.
2. Denote the proof of Theorem, given in Section IV, by The Proof of The Theorem.
3. We say that $\text{AM} \ \mu$ recognizes The Proof of Theorem, if $A_\mu = t$, where $A_\mu$ is denoted by

   the following chain of denotation:
   
   \[ \text{• } \mathcal{L}^{(3)} \text{ is } \text{TM from the statement of Theorem.} \]
   \[ \text{• } C_\mathcal{R} \equiv \sigma_q^{(3)}. \]
   \[ \text{• } P_{\mathcal{R}_3} \text{ is the text of The Proof of The Theorem for } \mathcal{R}. \]
   \[ \text{• } \mu \equiv \forall \mathcal{R}_3 \in \mathcal{R}_{3} \left(P_{\mathcal{R}_3} \rightarrow T_{\mathcal{R}_3}\right). \]

\[ \text{COROLLARY OF Theorem. } \forall_{\text{AM} \ \mu} \left( (\Phi_\mu = t) \land (A_\mu = t) \Rightarrow \forall_{\text{TM} \ \mathcal{R} \in \mathcal{R} (\mu \neq \mathcal{R})} \right). \]

\[ \text{PROOF. Consider } \text{AM} \ \mu, \text{ for which } (\Phi_\mu = t) \land (A_\mu = t), \text{ and assume } \mu \neq \mathcal{R} \text{ for some STM } \mathcal{R}. \]

Then $\mu = \mathcal{R}$, where $\mathcal{R} = \mathcal{S}$, and

$$(\sigma_q^{(3)} = t) \quad \text{(Lemma 4)} \quad \Rightarrow \quad (\mathcal{R} \neq \Phi_{3R} \theta \sigma_q^{(3)}) \quad \text{and} \quad \Rightarrow \quad (\mathcal{R} \neq T_{\mathcal{R}_3}) \quad \Rightarrow \quad (P_{\mathcal{R}_3} \rightarrow T_{\mathcal{R}_3}) \quad \Rightarrow \quad (A_{3R} \neq t).$$

The latter contradicts the assumptions $A_\mu = t$ and $\mu = \mathcal{R}$. Thus, $\sigma_q^{(3)} = f$, whence by The Theorem $\Phi_{3R} \neq t$. Then $\Phi_\mu \neq t$, from where $\mu$ fails to satisfy the assumption $\mu = \mathcal{S}$.

\[ \text{VI. Examples} \]

We shall give, as illustration, examples of TM $\mathcal{R}$, for which $\Phi_{3R} = t$, $\Phi_{3R} = f$, and $\Phi_{3R} = u$.

\[ \text{VI.1. Example of } \Phi_{3R} = t \]

Consider an arbitrary TM $\mathcal{R}$, for which $\forall_{\mathcal{R} \in \mathcal{E}} (\mathcal{R} = \infty)$. The definition of $\mathcal{R}$ implies $\Phi_{3R} = t$. Note that $A_{3R} = f$.

\[ \text{VI.2. Example of } \Phi_{3R} = f \]

Consider an arbitrary TM $\mathcal{B}$, for which $w = t$, where $w = \forall_{\mathcal{R} \in \mathcal{E}} (\mathcal{R} = \infty)$. We have

$$(p \rightarrow w \theta \rightarrow w) \Rightarrow ((w = t) \Rightarrow (-w = t)) = f \text{ for every } p \in \mathcal{P}, \text{ hence } \Phi_{3B} = t. \text{ Note that } A_{3B} = t.$$
Theorem \(\Phi_{\overline{\varepsilon}} = \varepsilon\) \(\iff\) \(\exists_{p \in P} \exists_{h \in H} \exists_{c \in C} \left(\neg \left(\left(p \rightarrow h \theta c\right) \Rightarrow \left(h = t\right) \Rightarrow \left(c = t\right)\right)\right)\),

but the latter contradicts The Theorem, therefore \(\Phi_{\overline{\varepsilon}} \neq \varepsilon\) and, as a result, \(\Phi_{\overline{\varepsilon}} = u\).

Note that equality \(\Phi_{\overline{\varepsilon}} = u\) is the corollary of the fact that the right side of (III.2.1), for \(\mu = \overline{C}\), depends on the truth of \(\Phi_{\overline{\varepsilon}}\).

### VII. Conclusion

Let us examine in more detail, what model can we offer for mathematician M, who is provided with the text P as a possible proof of the theorem \(T = (H, C)\), where H is the hypothesis of the theorem, and C is the conclusion. Mathematician M can make judgment “Not Proved” (M does not recognize P as the proof of T) or “Proved” (M recognizes P as the proof of T), and may change already made judgment. However, if M is human, then in the lifetime of M there will always be some last, and thus final, judgment relative to P and T. Therefore, it is possible to apply to M the model of an abstract mathematician, defined in Subsection III.1.2 as the map of pairs \(\{T, P\}\) to symbols 0 and 1, such that:

- \(\{T, P\}\) maps to 1 if M arrives at the final judgment “Proved”.
- \(\{T, P\}\) maps to 0 otherwise (M arrives at the final judgment “Not Proved”, expresses no judgment, does not at all examine the pair \(\{T, P\}\)).

Let us now consider how we can understand the equivalence of mathematician M and certain Turing machine. In the context of examination of pairs \(\{T, P\}\) this machine has to read the texts T and P and, according to the results of its work, it either reports “Proved” or “Not Proved”, or makes no report (the latter corresponds to the situation where M makes no judgment). However, exactly that kind of Turing machine we have introduced in Subsection III.1.4 as a possible way of generating an abstract mathematician.

Considering in the introduction (see Section I) the faultless mathematicians, we have involuntarily established the possibility that the concept of truth is suitable for some expressions. Indeed, we have named the mathematician M faultless, if his judgment “Proved” implies that since H, it actually follows that C, i.e. H can be true, C can be true and the truth of C actually arises from the truth of H. This informal definition is well harmonizes with the definition of faultless mathematician, introduced in Subsection III.2.

Further, M can recognize or not recognize the proof of the theorem from Section IV (hereinafter The Proof and The Theorem). In other words, if Turing machine, mentioned in The Theorem, is specified, and H is the hypothesis of The Theorem \((H = \Phi_{\overline{\varepsilon}})\), C is the conclusion of The Theorem \((C = \sigma_{\varepsilon}(H))\), and \(\mathcal{P}\) is The Proof, then:

- Either M recognizes The Proof, that is M arrives at the judgment “Proved” for \((H, C, \mathcal{P})\).
- Alternatively, M does not recognize The Proof, that is M arrives at the judgment “Not Proved” for \((H, C, \mathcal{P})\) or expresses no judgment or does not at all consider The Theorem.

Now we have a full opportunity to apply to mathematician M the corollary of The Theorem from Section V:

* A faultless mathematician, who recognizes The Proof of The Theorem, is not a Turing machine.

### VIII. Definitions and notations

In tables given below, the global definitions and notations (terms) are shown with specifying the number of subsection, where the corresponding definition or notation was denoted.

#### VIII.1. Definitions and notations

| Abridged notations | Turing machines | Theorems, mathematicians, and provability |
|--------------------|----------------|------------------------------------------|
| Text \(\iff\), Text \(\iff\) | \(M\) | \(T\) | |
| II.2 | II.2.1 | III.1.1 |
| Alphabets and symbols | [\mathfrak{M}, b] | III.2.2 | [\mathfrak{M}, b] \neq \infty | II.2.2 | [\mathfrak{M}, b] = \infty | II.2.2 | \mathbb{H} | III.1.1 |
|-----------------------|----------------|---------|-------------------------------|---------|-------------------------------|---------|--------|---------|
| A, A^*               | II.1.1        |         | \mathbb{C}                     | III.1.1 | \mathbb{P}                   | III.1.1 | \mathbb{S} | III.1.3 |
| \emptyset            | II.1.1        |         | \mathbb{F}                      | III.1.4 | \mu^{\mathfrak{M}}          | III.1.5 | p \rightarrow t | III.1.5 |
| \mathbb{B}, \mathbb{B}^* | II.1.1    |         | \mu \vdash t                    | III.1.5 | \phi_{\mu}                  | III.2   | \mathbb{V} | V       |
| \forall \emptyset U  | II.1.2        |         |                               |         |                               |         |        |         |

**The main statement**

| Term | Section |
|------|---------|
| \sigma_{\mathfrak{M}} | IV       |
| \Omega^{(\mathfrak{M})} | IV       |
| C_{\mathfrak{M}} | V        |
| P_{\mathfrak{M}} | V        |
| T_{\mathfrak{M}} | V        |
| A_{\mu} | V        |

### VIII.2 Terms

| Term | Section |
|------|---------|
| TM   | II.2.1  | AT | III.1.1 |
| TM that can be constructed | II.5.1 | AM | III.1.2 |
| STM  | II.4.1  | Faultless AM | III.2 |
| The Theorem | V | AM, recognizing | |
| The Proof (of The Theorem) | V | The Proof of The Theorem | V |