EÖTVÖS LORÁND TUDOMÁNYEGYETEM
Bölcsészettudományi Kar

DIPLOMAMUNKA

Logikai paradoxonok és eldönthetetlen mondatok

Logical Paradoxes and Undecidable Sentences

Témavezető:
Csaba Ferenc
Tudományos munkatárs

Készítette:
Cheng, Ka Yue
Logic and Theory of Science

2015

A HKR 346. § ad 76. § (4) c) pontja értelmében:
„… A szakdolgozathoz csatolni kell egy nyilatkozatot arról, hogy a munka a hallgató saját szellemi terméke…”

SZERZŐSÉGI NYILATKOZAT

Alulírott Cheng, Ka Yue (K2VXS8) ezennel kijelentem és aláírásommal megerősítem, hogy az ELTE BTK Logic and Theory of Science angol nyelvű mesterszakján írt jelen diplomamunkám saját szellemi termékem, melyet korábban más szakon még nem nyújtottam be szakdolgozatként és amelybe mások munkáját (könyv, tanulmány, kézirat, internetes forrás, személyes közlés stb.) idézőjel és pontos hivatkozások nélkül nem építettem be.

Budapest, 2015-04-15

……………………………
a hallgató aláírása

A nyilatkozatot a diplomamunkához kell csatolni.

Acknowledgements

I would like to thank my supervisor Ferenc Csaba for his useful comments and pointing out some mistakes in the draft. Despite this, the responsibility for the remaining typos and errors is mine.

I would like to express my gratitude to all my teachers, especially Professor András Máté, Professor István Németi, Professor Hajnal Andréka, Professor Péter Mekis, Professor László E. Szabó, Attila Molnár and Zalán Gyenis, as well as everyone from the Department of Logic. This thesis would not exist without them, I am grateful for studying logic here.

Last but not least, I wish to express my sincere thanks to my family for their support, and to my lovely cat Alice.

Contents

1 Introduction

According to Chaitin (1995), Gödel once told him “it doesnt matter which paradox you use [to prove the first incompleteness theorem]”. However, there is no proof that for every (logical) paradox there is a corresponding undecidable sentence in Peano arithmetic. Therefore it is worth to investigate what will happen if we formalize paradoxes in Peano arithmetic.

We begin by presenting a proof of the Gödel’s first incompleteness theorem, then a proof of Löb’s theorem and the Second Incompleteness Theorem.

After that, we investigate what happens when formalizing the paradoxes to obtain undecidable sentences. The first two are Grelling’s paradox and Curry’s paradox, which are not difficult to formalize and the resulting undecidable sentences are not really different from Gödel’s one.

Then we turn to the Berry paradox. There are already two proofs of the first incompleteness theorem involving this paradox in the literature, one is from Gregory Chaitin, another one is from George Boolos.

The fourth paradox is the Yablo’s paradox, a paradox about an infinite sequence of sentences. Starting from this section, we use a generalized version of the Diagonal Lemma heavily.

Three interrelated paradoxes follow, I developed them from the preface paradox, but they are significantly different from it. The first two of them have both an infinite and finite version, while the third one is paradoxical mainly in the infinite case.

The last paradox discussed in this paper is the surprise examination paradox. An infinite version of this paradox proposed by Sorensen, called the earliest class inspection paradox, is also discussed. We first have a look in the formalization of the surprise examination paradox by Frederic Fitch, from which he concluded that the surprise examination paradox is not a real paradox but only self-contradictory. Then I give my own formalization of both the infinite and finite version of the paradox, showing that Fitch’s claim is not entirely justified.

Finally there are some remarks concerning what is done in this thesis and how should we interpret the results.

2 A Proof of Gödel Incompleteness Theorems

In this paper, we will work in a formal system called **T**. The main idea of the proof presented here is from Gödel (1992), which is an English translation form Gödel's classic 1931 paper, Smith (2007), and Smullyan (1992).

2.1 Syntax

> "The details of an encoding are fascinating to work out and boring to read. The author wrote the present section for his own benefit and his feelings will not be hurt if the reader chooses to skip it." (Smoryński, 1977, p.829)

Symbols

- logical symbols: $\forall, \neg, \rightarrow, (,), =$;
- infinitely many variable symbols: v_i, for $i \in \mathbb{N}$;
- an unary function symbol S;
- two binary function symbols $+, \times$;
- a constant symbol: 0.

Abbreviations

For readability, in the rest of this proof the following abbreviations may be used (x, y, z are variables):

- $(\varphi \vee \psi)$ for $(\neg\varphi \rightarrow \psi)$
- $(\varphi \wedge \psi)$ for $\neg(\neg\varphi \vee \neg\psi)$
- $(\varphi \leftrightarrow \psi)$ for $((\varphi \rightarrow \psi) \wedge (\psi \rightarrow \varphi))$
- $\exists x(\varphi)$ for $\neg\forall x(\neg\varphi)$
- Sx for $S(x)$
- $x + y$ for $+(x, y)$
- $x \times y$ for $\times(x, y)$

- 1 for $S0, 2$ for $S1, 3$ for $S2$, etc.
- $x \leq y$ for $\exists z\,(x + z = y)$
- $x < y$ for $(x \leq y) \wedge \neg(x = y)$
- $x \leq y \leq z$ for $(x \leq y) \wedge (y \leq z)$, similarly we will write $x < y \leq z$, $x \leq y = z$ etc.
- $(\forall x \leq t)\,\varphi$ for $\forall x\,((x \leq t) \rightarrow \varphi)$, where t is a term.
- $(\exists x \leq t)\,\varphi$ for $\exists x\,((x \leq t) \wedge \varphi)$, where t is a term.
- $(\mu x < t)\varphi(x) = t'$ for $[(t' < t) \wedge \varphi(t') \wedge \forall z\,(\varphi(z) \rightarrow t' \leq z)] \vee [(\forall z < t)\neg\varphi(z) \wedge (t' = t)]$, where t and t' are terms.

We will also abbreviate variables by single letters, e.g. x, y, z, w, u, v, n etc. When there are different letters in a formula, it is always assumed that they are abbreviation of different variables.

When there is no ambiguity, brackets maybe skipped, e.g.$\neg 0 = 1$ instead of $\neg(0 = 1)$. Also, different kinds of brackets like $[,]$ and $\{, \}$ may be used instead of $(,)$ to make formulas more readable.

Terms

An expression is a *term* if it belongs to the following recursively defined set of expressions:

T1 0 is a term.

T2 Every variable v_i is a term.

T3 If t_1, t_2 are terms, then so are $St_1, (t_1 + t_2)$, and $(t_1 \times t_2)$.

T4 No other expressions are terms.

A term is a *closed* term if it contains no variables.

Formulas

An expression is an *atomic formula* if it is of the form $t_1 = t_2$, where t_1 and t_2 are terms.

An expression is a formula if it belongs to the following recursively defined set of expressions:

F1 Every atomic formula is a formula.

F2 If φ and ψ are formulas, then $\neg\varphi$ and $(\varphi \to \psi)$ are both formulas.

F3 If φ is a formula and v_i a variable, then $\forall v_i \varphi$ is a formula.

F4 No other expressions are formulas.

Free and bounded variables

Let v_i be a variable. We define free and bounded variables in formulas by the following rules:

1. For any atomic formula φ, v_i is a free variable in φ if and only if it occurs in φ. No variables are bounded variables in any atomic formula.

2. For any formulas φ, v_i is a bounded variable in $\forall v_i \varphi$.

3. For any formulas φ and ψ, v_i is a free variable in $(\varphi \to \psi)$ if and only if it is a free variable in φ or ψ; v_i is a free variable in $\neg\varphi$ if and only if it is a free variable in φ.

4. For any formulas φ and ψ, v_i is a bounded variable in $(\varphi \to \psi)$ if and only if it is a bounded variable in φ or ψ; v_i is a bounded variable in $\neg\varphi$ if and only if it is a bounded variable in φ.

5. For $i \neq j$ and any formula φ, v_i is a free variable in $\forall v_j \varphi$ if and only if it is a free variable in φ.

Sometimes we may write "v_i is free in φ" for "v_i is a free variable in φ". A variable v_i can be both free and bounded in a formula. If v_i is free in a formula φ, the free occurrences of v_i in φ are the occurrences of v_i such that v_i is free in φ. Bounded occurrences of a variable in a formula is defined similarly.

A *sentence* is a formula that has no free variable.

Substitutions

Let t be a term, v_i a variable and φ a formula. t is *free for v_i in φ* if no free occurrence of v_i in φ is within the scope of a quantifier $\forall v_j$ where v_j is any variable occurring in t.

If t is a term and v_i is a variable, $\varphi(t/v_i)$ is the formula formed by replacing every free occurrences of v_i by t. If φ has only one variable x, then we may write $\varphi(t)$ instead of $\varphi(t/x)$.

2.2 Axioms and inference rules

Axiom schemata for First-order logic

1. $(\varphi \to (\psi \to \varphi))$
2. $((\varphi \to (\psi \to \chi)) \to ((\varphi \to \psi) \to (\varphi \to \chi)))$
3. $((\neg\psi \to \neg\varphi) \to (\varphi \to \psi))$
4. $\forall v_i \varphi \to \varphi(t/v_i)$, where t is a term and free for v_i in φ.
5. $\forall v_i \, (v_i = v_i)$
6. $(t = t') \to (\varphi \to \varphi(t//t'))$ where t, t' are terms, $\varphi(t//t')$ is a formula obtained by substituting an occurrence of t by t', and t is not in the scope of any quantifier $\forall v_j$ where v_j is a variable in t' but not in t.

Inference Rules

MP From φ and $(\varphi \to \psi)$ infer ψ.

$\forall$-Intro From $\varphi \to \psi$ infer $\varphi \to \forall x \psi$ if x is not free in φ.

Axioms for arithmetic

7. $\forall v_0 \, (\neg S v_0 = 0)$
8. $\forall v_0 \forall v_1 (S v_0 = S v_1 \to v_0 = v_1)$
9. $\forall v_0 \, (v_0 + 0 = 0)$
10. $\forall v_0 \forall v_1 \, (v_0 + S v_1 = S(v_0 + v_1))$

11. $\forall v_0 (v_0 \times 0 = 0)$

12. $\forall v_0 \forall v_1 \, (v_0 \times S v_1 = (v_0 \times v_1) + v_0)$

13. For every formula $\varphi(v_i)$ with one free variable v_i, the following formula is an axiom: $(\varphi(0) \land \forall v_i \, (\varphi(v_i) \to \varphi(S v_i))) \to \forall v_i (\varphi(v_i))$

2.3 Proofs and theorems

A sequence of formulas is a *proof* if every formula in this sequence is either an axiom, an instance of an axiom schema, or obtained by applying an inference rule to some formulas occurring before it.

A formula is a *theorem* if it is in a proof. A proof of a formula φ is a proof that ends with φ. A formula is said to be derivable, or *provable*, denoted by $\vdash \varphi$, if it is a theorem. A formula φ is *refutable* if the negation of the formula, $\neg\varphi$ is provable. A formula is *decidable* if it is either provable or refutable, and *undecidable* if it is neither provable nor refutable.

2.4 Gödel numbering

We now define assign each expression E to a unique natural number $\ulcorner E \urcorner$, which is called the Gödel number of E.

First we define it for all basic symbols:

$$\ulcorner 0 \urcorner = 1, \ulcorner S \urcorner = 3, \ \ulcorner \neg \urcorner = 5, \ \ulcorner \to \urcorner = 7, \ \ulcorner \forall \urcorner = 9, \ulcorner (\urcorner = 11, \ \ulcorner) \urcorner = 13,$$
$$\ulcorner + \urcorner = 15, \ \ulcorner \times \urcorner = 17, \ \ulcorner v_i \urcorner = 19^{i+1} \text{ where } i \in \mathbb{N}, \ \ulcorner = \urcorner = 21$$

For every finite sequence of numbers $e_1, e_2, \dots e_n$ we encode the sequence into the number:

$$\prod_{i=1}^{n} p_i^{e_i+1} = 2^{e_1+1} \times 3^{e_2+1} \times \dots \times p_n^{e_n+1}$$

where p_i is the i^{th} prime number.

We call this the code number of the sequence $e_1, e_2, \dots e_n$. By the fundamental theorem of arithmetic, every finite sequence corresponds to a unique code number.

If x is the code number of a sequence s, we will say *s is represented by x* or *x represents s*.

Since an expression is a finite sequence of symbols and each symbol corresponds to a number, every expression corresponds to a finite sequence of number. The code number of an expression is called the Gödel number of it. For example, an expression with n symbols, $\alpha_1, \alpha_2, \ldots, \alpha_n$ with Gödel numbers $a_1, a_2, \ldots, a_n$ respectively, corresponds to the following Gödel number:

$$\prod_{i=1}^{n} p_i^{a_i+1} = 2^{a_1+1} \times 3^{a_2+1} \times \ldots \times p_n^{a_n+1}$$

Similarly, the Gödel number of a sequence of expressions (for example, a proof) is the code number of the sequence of Gödel numbers of those expressions in the original order. We will denote the Gödel number of the sequence $s_1, s_2, \ldots s_n$ as $\ulcorner s_1, s_2, \ldots s_n \urcorner$, which is $2^{\ulcorner s_1 \urcorner + 1} \times 3^{\ulcorner s_2 \urcorner + 1} \times \ldots \times p_n^{\ulcorner s_n \urcorner + 1}$.

2.5 Primitive recursive functions and relations

We now recursively define the class of *primitive recursive functions*:

PR1 The zero function, $Z(x) = 0$ for any x, is primitive recursive.

PR2 The successor function $S(x)$ is primitive recursive.

PR3 For any natural numbers i, n such that $0 \leq i < n$, the projection function $P_i^n(x_0, x_1, \ldots, x_{n-1}) = x_i$ is primitive recursive.

PR4 If $G(x_0, x_1, \ldots, x_{m-1}), H_0(x_0, x_1, \ldots, x_{n-1}), H_1(x_0, x_1, \ldots, x_{n-1}), \ldots,$ $H_{m-1}(x_0, x_1, \ldots, x_{n-1})$ are primitive functions, then $F(x_0, x_1, \ldots, x_{n-1})$ $= G(H_0(x_0, x_1, \ldots, x_{n-1}), H_1(x_0, x_1, \ldots, x_{n-1}), \ldots, H_{m-1}(x_0, x_1, \ldots, x_{n-1}))$ is also a primitive recursive function.

PR5 If $G(x_1, \ldots, x_n), H(x_0, x_1, \ldots, x_n)$ are primitive functions, then $F(x_0, x_1, \ldots, x_n)$ defined by the following recursion is also a primitive recursive function:

$$\begin{cases} F(0, x_1, \ldots, x_n) = G(x_1, \ldots, x_n) \\ F(S(x), x_1, \ldots, x_n) = H(F(x, x_1, \ldots, x_n), x_1, \ldots, x_n) \end{cases}$$

PR6 Only functions obtained from the above rules, i.e. from **PR1** to **PR5**, are primitive recursive functions.

A n-ary relation R is primitive recursive if its characteristic function $\chi_R(x_0, x_1, \ldots, x_{n-1})$ is primitive recursive.

It is well known that the class of primitive relations is closed under substitution by primitive recursive functions, conjunction, disjunction, negation, bounded quantification and bounded minimization.

In other words, if R, R' are n-ary relations, S is an $(n+1)$-ary relation, $f_0, f_1, \ldots, f_{n-1}$ are m-ary functions, g is an $n+1$-ary functions, and all of them are primitive recursive, then:

1. $R(f_0(x_0, x_1, \ldots, x_{m-1}), f_1(x_0, x_1, \ldots, x_{m-1}), \ldots, f_{n-1}(x_0, x_1, \ldots, x_{m-1}))$ is an m-ary primitive recursive relation.

2. $R(x_0, x_1, \ldots, x_{n-1}) \wedge R'(x_0, x_1, \ldots, x_{n-1})$ is an n-ary primitive recursive relation.

3. $R(x_0, x_1, \ldots, x_{n-1}) \vee R'(x_0, x_1, \ldots, x_{n-1})$ is an n-ary primitive recursive relation.

4. $\neg R(x_0, x_1, \ldots, x_{n-1})$ is an n-ary primitive recursive relation.

5. $(\forall y \leq x_0) S(y, x_1, \ldots, x_n)$ is an $(n+1)$-ary primitive recursive relation.

6. $(\exists y \leq x_0) S(y, x_1, \ldots, x_n)$ is an $(n+1)$-ary primitive recursive relation.

7. The $(n+1)$-ary function $f(x_0, x_1, \ldots, x_n)$ defined by $f(x_0, x_1, \ldots, x_n) = (\mu y < x_0) S(y, x_1, \ldots, x_n)$ is primitive recursive.

8. The $(n+1)$-ary function $f(x_0, x_1, \ldots, x_n)$ defined by $f(x_0, x_1, \ldots, x_n) = (\mu y < g(x_0, x_1, \ldots, x_n)) S(y, x_1, \ldots, x_n)$ is primitive recursive.

It is also well known that $x + y$, $x \times y$ are primitive recursive functions, and $x \leq y$, $x = y$ are primitive recursive relations.

The proofs of the above results can be found in Smoryński (1991).

2.6 Pseudo-terms

Since we have only three function symbols $S, +$ and $\times$, all terms are construct by them and the only constant symbol 0. However in the following we will define some new "functions", for example $Pr(n)$, and use them as a function.

To make things precise, please note that those “terms” such as $2^5, Pr(3), R(9)$ are not real terms, but *pseudo-terms*, which are defined by the following:

Let $F(\mathbf{x}, y)$ abbreviates the formula $F(x_0, x_1, \ldots x_{n-1}, y)$ with n free variables $x_0, x_1, \ldots, x_{n-1}, y$. $F(\mathbf{x}, y)$ is a pseudo-term with respect to the variable y if the formula

$$\forall x_0 \forall x_1 \ldots \forall x_{n-1} \exists y \big[F(\mathbf{x}, y) \wedge \forall z \big(F(\mathbf{x}, z) \rightarrow y = z\big)\big]$$

is provable.

When we write something like $f(\mathbf{x}) = y$, it should be understood as “there is a formula $F(\mathbf{x}, y)$ which is a pseudo-term”. Also when we are using pseudo-terms, for example in the form $\varphi\big(f(x)\big)$ where $\varphi(x)$ is an open formula with one variable, it should be read as $\exists y\big(\varphi(y) \wedge f(x) = y\big)$. But we will skip all the proofs that there are such pseudo-terms.

2.7 β-function and primitive recursion

An n-place function $b(x_1, \ldots x_{n-1}, y)$ is called a β-function if for each finite sequence of natural numbers $k_0, k_1, \ldots, k_m$ there are numbers $c_0, c_1, \ldots, c_{n-1}$ such that for every $i \leq m$, $b(c_0, c_1, \ldots, c_{n-1}, i) = k_i$.

From Gödel (1992), we know there is a ternary β-function $\beta(x, y, z)$. To define the corresponding relation, we need only bounded quantification, a proof can be found in Smith (2007).

The existence of a β-function ensures that we can encode and decode finite sequences by numbers, this helps us to define functions (which should be pseudo-terms) obtained from primitive recursion in the object language.

For example, we can define the exponentiation in the following way:

$$exp(x, y, z) \leftrightarrow \exists c \exists d \big[\beta(c, d, 0) = S0 \wedge (\forall u < y)\big(\beta(c, d, Su) = x \times \beta(c, d, u)\big) \wedge \beta(c, d, y) = z\big]$$

The idea is that, to say $x^y = z$, we can describe the existence of a sequence $1, x, x^2, \ldots, x^y$, where the first term is 1 and each later term is the previous term multiplied by x, and the $y + 1^{st}$ term is z.

2.8 Some primitive recursive relations and functions

Now we can define the following functions and relations. To make the formulas slightly more understandable, below each formula there is a short description of it.

1. $y \mid x \longleftrightarrow (\exists z \leq x)\,(x \times z = y)$

   y divides x, or equivalently x is divisible by y.

2. $x^0 = S0$

   $x^{Sy} = x \times x^y$

   x^y is the exponential function.

3. $Prime(x) \longleftrightarrow (\forall y \leq x)\,(y \mid x \rightarrow (y = 1 \vee y = x))$

   x is a prime number.

4. $Pr(0) = 0$

   $Pr(Sk) = \big(\mu y \leq (Pr(k) + Pr(k))\big) Prime(y)$

   $Pr(k)$ is the n^{th} prime number.[1]

5. $Code(x) \longleftrightarrow (\forall y \leq x)\,((y \neq 1 \wedge Pr(Sy) \mid x) \rightarrow Pr(y) \mid x)$

   x is a code number.

6. $l(x) = (\mu y \leq x)\big(Code(x) \wedge (Pr(y) \mid x) \wedge \neg(\mathrm{Pr}(Sy) \mid x)\big)$

   If x is a code number, then its length is $l(x)$.

7. $Dec(n, x) = (\mu y \leq x)\big(Code(x) \wedge Pr(n)^{Sy} \mid x \wedge \neg(Pr(n)^{SSy} \mid x)\big)$

   If x is the code number of a sequence, $Dec(n, x)$ is the n^{th} term in that sequence.

8. $x \star y = \big[\mu z \leq [Pr(l(x) + l(y))]^{x+y}\big]\big\{Code(x) \wedge Code(y) \wedge (\forall n \leq l(x))$
   $[Dec(n, z) = Dec(n, x)] \wedge (\forall n \leq l(y))\,[Dec(n + l(x), z) = Dec(n, y)]\,\big\}$

   For code numbers x and y, $x \star y$ is a code number with length $l(x) + l(y)$ and the first $l(x)$ indices are the same as x, the rest are the same as y (in the same order). This is used to represent the concatenation of two expressions.

9. $Part(y, x) \longleftrightarrow Code(x) \wedge Code(y) \wedge (\exists u \leq x)(\exists v \leq x)(x = u \star y \star v)$

   y represent a sequence which is a part of the sequence represented by x.

[1] On the bound of y: By Bertrand's Postulate (or called Chebyshev's theorem), for every number $n > 1$, there is a prime number p such that $n < p < 2p$, hence we have the inequality $p_n < p_{n+1} < p_n + p_n$.

10. $R(x) = 2^{Sx}$

    $R(x)$ is the code number of a sequence with only one number x.

11. $E(x) = (R(11) \star x) \star R(13)$

    Given that x is the code number of a finite sequence, $E(x)$ corresponds to the code number of a sequence with 11 at the beginning, 13 at the end, and the sequence encoded by x in between. If x is a Gödel number of an expression, then $E(x)$ is the Gödel number of the expression added a pair of brackets.

12. $Var(x) \longleftrightarrow (\forall y \leq x)\,(y \mid x \to 19 \mid y)$

    x is the Gödel number of a variable.

13. $Neg(x) = R(5) \star x$

    If x is the Gödel number of a formula φ, $Neg(x)$ is the Gödel number of the formula $\neg\varphi$.

14. $Imp(x, y) = E(x \star R(7) \star y)$

    If $x = \ulcorner\varphi\urcorner, y = \ulcorner\psi\urcorner$, then $Imp(x, y) = \ulcorner(\varphi \to \psi)\urcorner$.

15. $Dis(x, y) = Imp(Neg(x), y)$

    If $x = \ulcorner\varphi\urcorner, y = \ulcorner\psi\urcorner$, then $Dis(x, y) = \ulcorner\varphi \vee \psi\urcorner$.

16. $Con(x, y) = Neg(Dis(Neg(x), Neg(y)))$

    If $x = \ulcorner\varphi\urcorner, y = \ulcorner\psi\urcorner$, then $Con(x, y) = \ulcorner\varphi \wedge \psi\urcorner$.

17. $Gen(x, y) = (R(9) \star R(x)) \star E(y)$

    If $x = \ulcorner v_i \urcorner$ for some variable v_i and $y = \ulcorner\varphi\urcorner$, then $Gen(x, y) = \ulcorner\forall v_i(\varphi)\urcorner$.

18. $N(0) = 1; N(Sn) = R(3) \star E(N(n))$

    $N(n) = \ulcorner n \urcorner$ for every number n.

19. $transt(x, y, z) \longleftrightarrow \big(x = R(3) \star E(y)\big) \vee \big(x = E(y \star R(15) \star z)\big) \vee \big(x = E(y \star R(17) \star z)\big)$

    If y and z are Gödel numbers of terms and $transt(x, y, z)$ is provable, then x is the Gödel number of a term formed according to the terms transformation rule (T3).

20. $Seqt(x) \longleftrightarrow \big(\forall y \leq l(x)\big)\Big[\big(Dec(y,x)=1\big) \vee Var\big(Dec(y,x)\big) \vee$
$(\exists z \leq y)(\exists w \leq y)\big[transt\big(Dec(y,x), Dec(z,x), Dec(w,x)\big)\big]\Big]$
x is the Gödel number of a sequence of expressions which represents the formation of a term.

21. $Term(x) \longleftrightarrow \Big[\exists y \leq \big[Pr\big(l(x)\big)\big]^{x\times l(x)}\Big]\big(Seqt(y) \wedge Dec(l(y),y)=x\big)$
x is the Gödel number of a term. [2]

22. $Atf(x) \longleftrightarrow (\exists y \leq x)\,(\exists z \leq x)\,[x = (y \star R(21)) \star z]$
x is the Gödel number of an atomic formula.

23. $transf(x,y,z) \longleftrightarrow \big(x = Neg(y)\big) \vee \big(x = Imp(y,z)\big) \vee (\exists v \leq x)\big(Var(v) \wedge x = Gen(v,y)\big)$
If y and z are Gödel numbers of formulas and $transt(x,y,z)$ is provable, then x is the Gödel number of a formula formed according to the formula transformation rules (F2) and (F3).

24. $Seqf(x) \longleftrightarrow \big(\forall y \leq l(x)\big)\Big[Atf\big(Dec(y,x)\big) \vee (\exists z \leq y)(\exists w \leq y)$
$\big[transf\big(Dec(y,x), Dec(z,x), Dec(w,x)\big)\big]\Big]$
x is the Gödel number of a sequence of expressions which represents the formation of a formula.

25. $Form(x) \longleftrightarrow \Big[\exists y \leq \big[Pr\big(l(x)\big)\big]^{x\times l(x)}\Big]\big(Seqf(y) \wedge Dec(l(y),y)=x\big)$
x is the Gödel number of a formula.[3]

26. $Sbf(y,x) \longleftrightarrow Form(x) \wedge Form(y) \wedge Part(y,x)$
y represents a subformula of the formula represented by x.

27. $Bdd(x,n,v) \longleftrightarrow Var(v) \wedge Form(x) \wedge (\exists y \leq x)(\exists z \leq x)(\exists w \leq x)$
$\big[(x = y \star Gen(v,z) \star w) \wedge Form(z) \wedge \big(l(y)+1\big) \leq n \leq \big(l(y)+l(Gen(v,z))\big)\big]$
v is the Gödel number of a variable which is bounded at the n^{th} place in the formula having the Gödel number x.

[2] On the upper bound of y: Let x be the Gödel number of a term t, $l(x)=n$ means there are n symbols in the expression t. By induction on the formation of a term, we can prove that we need at most n steps to form the term t. Hence the corresponding sequence has length at most n, let it be $t_1, t_2, \ldots, t_n$, where $t_n = t$. Every $t_i (0 < i \leq n)$ is a part of t, therefore for every $0 < i \leq n, \ulcorner t_i \urcorner \leq \ulcorner t \urcorner = x$. This leads to the result that $\ulcorner t_0, t_1, \ldots, t_n \urcorner \leq \ulcorner t, t, \ldots t \urcorner \leq \big(p_n^{\ulcorner t \urcorner}\big)^n$, which is the upper bound of y in the above formula.

[3] On the upper bound of y: similar to the definition of $Term(x)$, cf. footnote 2.

28. $Fr(x,n,v) \longleftrightarrow Var(v) \wedge Form(x) \wedge \big(Dec(n,x)=v\big) \wedge \big(n \leq l(x)\big) \wedge \neg Bdd(x,n,v)$

v is the Gödel number of a variable which is free at the n^{th} place in the formula having the Gödel number x.

29. $Free(v,x) \longleftrightarrow \big(\exists y \leq l(x)\big)Fr(x,y,v)$

v is the Gödel number of a variable which is free in the formula having the Gödel number x.

30. $Su(x,n,y) = \big[\mu z \leq \big[Pr\big(l(x)+l(y)\big)\big]^{x+y}\big]\Big\{(\exists u \leq x)(\exists v \leq x) \big[\big(x = u \star R\big(Dec(n,x)\big) \star v\big) \wedge \big(z = u \star y \star v\big) \wedge \big(n = l(u)+1\big)\big]\Big\}$

If x is the code number of a sequence X, $Su(x,n,y)$ is the code number of the sequence that substituting the n^{th} term of X by the sequence represented by y (assuming that y is a code number).

31. $Pl(x,0,v) = \big(\mu n \leq l(x)\big)\big[Fr(x,n,v) \wedge \big(\forall y \leq l(x)\big)(Fr(v,y,x) \rightarrow n \leq y)\big]$

$Pl(x,Sk,v) = \big(\mu n \leq l(x)\big)\big[Fr(x,n,v) \wedge \big(\forall y \leq l(x)\big)\big((Fr(x,y,v) \wedge Pl(x,k,v) < y) \rightarrow n \leq y\big)\big]$

If v is the Gödel number of a variable, x is the Gödel number of a formula, then $Pl(x,k,v)$ is the place of the $(k+1)^{st}$ free occurrence of that variable in that formula, counting from the beginning. If there is no such place, then $Pl(x,k,v)=0$.

32. $A(x,v) = \big(\mu n \leq l(x)\big)\big(Pl(x,n,v)=0\big)$

$A(x,v)$ is the number of places that the variable represented by v is free in the formula represented by x.

33. $Sbst(x,0,v,y) = x$

$Sbst(x,Sk,v,y) = Su\big(Sbst(x,k,v,y), Pl(x,k,v), y\big)$

Let x represents a formula φ, v a variable v_i and y a term t. Then $Sbst(x,n,v,y)$ is the Gödel number of the formula after substitution of v_i by t in the first n free occurrences of v_i in φ.

34. $Subs(x,v,y) = Sbst\big(x,A(x,v),v,y\big)$

Let x represents a formula φ, v a variable v_i and y a term t. $Subs(x,v,y)$ is the Gödel number of the formula after substitution of v_i by t in all free occurrences of v_i in φ.

35. $FirVar(x) = (\mu y \leq x)\{Free(y,x) \wedge \big(\exists z \leq l(x)\big)\big[(Dec(z,x) = y) \wedge (\forall v \leq x)\big(\forall n \leq l(x)\big)\big((Free(v,x) \wedge (Dec(n,x) = v)) \rightarrow (z \leq n)\big)\big]\}$

    $FirVar(x)$ represents the first free variable (counting from the beginning of the expression) of a formula.

36. $Sub(x,y) = Subs(x, FirVar(x), y)$

    $Sub(x,y)$ represents the expression obtained by substituting the expression represented by y to the first free variable to the formula represented by x.

37. $Prop(x) \longleftrightarrow Form(x) \wedge (\exists y \leq x)\big[Var(y) \wedge Free(y,x) \wedge (\forall z \leq x)\big(Var(x) \wedge Free(z,x) \rightarrow z = y\big)\big]$

    x is a formula with exactly one free variable.

38. $Frfor(x,v,y) \longleftrightarrow Form(x) \wedge Term(y) \wedge Var(v) \wedge (\forall z \leq y)$
    $\big[\big(Part(R(z),y) \wedge Var(z)\big) \rightarrow (\forall w \leq x)\big(Sbf(Gen(z,w),x) \rightarrow \neg Free(v,w)\big)\big]$

    If x represents a formula φ, v a variable v_i and y a term t, then $Frfor(x,v,y)$ is provable if and only if t is free for v_i in φ.

39. $Ax_1(x) \longleftrightarrow Form(x) \wedge (\exists y \leq x)(\exists z \leq x)\{Form(y) \wedge Form(z) \wedge \big[x = Imp\big(y, Imp(y,z)\big)\big]\}$

    x represents an instance of axiom schema 1.

40. $Ax_2(x) \longleftrightarrow Form(x) \wedge (\exists y \leq x)(\exists z \leq x)(\exists w \leq x)\Big\{Form(y) \wedge Form(z) \wedge Form(w) \wedge \Big[x = Imp\big[Imp\big(y, Imp(z,w)\big), Imp\big(Imp(y,z), Imp(y,w)\big)\big]\Big]\Big\}$

    x represents an instance of axiom schema 2.

41. $Ax_3(x) \longleftrightarrow Form(x) \wedge (\exists y \leq x)(\exists z \leq x)\{Form(y) \wedge Form(z) \wedge \big[x = Imp\big(Imp\big(Neg(z), Neg(y)\big), Imp(y,z)\big)\big]\}$

    x represents an instance of axiom schema 3.

42. $Ax_4(x) \longleftrightarrow Form(x) \wedge (\exists y \leq x)(\exists z \leq x)(\exists w \leq x)\Big\{Form(y) \wedge Var(z) \wedge Term(w) \wedge Frfor(y,z,w) \wedge \Big[\big[x = Imp\big(Gen(z,y), Subs(y,z,w)\big)\big]\Big]\Big\}$

    x represents an instance of axiom schema 4.

43. $Ax_5(x) \longleftrightarrow Form(x) \wedge (\exists y \leq x)\{Var(y) \wedge \big[x = Gen\big(y, R(y) \star R(21) \star R(y)\big)\big]\}$

    x represents an instance of axiom schema 5.

44. $Ax_6(x) \longleftrightarrow Form(x) \wedge (\exists y \leq x)(\exists z \leq x)(\exists u \leq x)(\exists v \leq x)\Big\{Form(y) \wedge Form(z) \wedge Term(u) \wedge Term(v) \wedge \Big[x = Imp\Big(R(u) \star R(21) \star R(v), Imp\big(y, z\big)\Big)\Big] \wedge (\exists w_1 \leq y)(\exists w_2 \leq y)\big[(y = w_1 \star R(u) \star w_2) \wedge (z = w_1 \star R(v) \star w_2) \wedge (\forall n \leq l(u))(\forall w \leq v)[Var(w) \wedge Part(w, v) \wedge \neg Part(w, u) \rightarrow \neg Bdd(y, l(w) + n, w)]\big]\Big\}$

    x represents an instance of axiom schema 6.

45. $Ax_{7-12}(x) \longleftrightarrow (x = a_7) \vee (x = a_8) \vee (x = a_9) \vee (x = a_{10}) \vee (x = a_{11}) \vee (x = a_{12})$ where $a_7, a_8, a_9, a_{10}, a_{11}, a_{12}$ are Gödel numbers of the axioms 7, 8, 9, 10, 11, 12 respectively.

    x represents one of the axioms 7 to 12.

46. $Ax_{13}(x) \longleftrightarrow (\exists y \leq x)(\exists z \leq x)\Big\{Prop(y) \wedge Var(z) \wedge Free(z, y) \wedge x = Imp\Big[Con\Big[Subs(y, z, 1), Gen\big[z, Imp\big(y, Subs(y, z, R(3) \star z)\big)\big]\Big), Gen(z, y)\Big]\Big\}$

    x represents an instance of axiom schema 13.

47. $Ax(x) \longleftrightarrow Ax_1(x) \vee Ax_2(x) \vee Ax_3(x) \vee Ax_4(x) \vee Ax_5(x) \vee Ax_6(x) \vee Ax_{7-12}(x) \vee Ax_{13}(x)$

    x represents an axiom.

48. $Consq(x, y, z) \longleftrightarrow Form(y) \wedge Form(z) \wedge \big(y = Imp(z, x)\big) \vee (\exists w \leq x)(\exists u \leq y)(\exists v \leq y)$
    $\big[Var(w) \wedge \big(y = Imp(u, v)\big) \wedge \neg Free(w, u) \wedge \big(x = Imp(u, Gen(w, v)\big)\big]$

    x represented a formula obtained by applying a inference rule on the formulas represented by y and z.

49. $Proof(x) \longleftrightarrow \big(\forall y \leq l(x)\big)\big\{(y = 0) \vee Ax(Dec(y, x)) \vee (\exists z \leq y)(\exists w \leq y)$
    $\big[(0 < z) \wedge (0 < w) \wedge Consq\big(Dec(y, x), Dec(z, x), Dec(w, x)\big)\big]\big\}$

    x represents a proof.

50. $Prf(x, y) \longleftrightarrow Proof(x) \wedge Dec(l(x), x) = y$

    x represents a proof of the formula y.

Note that the similarity between the definitions of $Seqt(x), Seqf(x)$ and $Proof(x)$, all three of them are defining a sequence of expressions that is defined recursively.

However the transformation rules for terms and formulas are all lengthening the expressions, where the transformation rules for theorems, i.e. the inference rules, can both lengthen and shorten the expressions. Therefore we can give an upper bound of the length for a shortest sequence[4] that forms a specific term or formula, but we cannot give one for a shortest proof of a specific theorem.

But still we can define a provability predicate, $Prov(x)$, that if x represents a theorem then $Prov(x)$ is provable, by the following formula:

$$Prov(x) \longleftrightarrow \exists y Prf(y, x)$$

2.9 Diagonal Lemma

We now prove the Diagonal Lemma. Though it is called a lemma, it is actually a quite important and useful theorem.

Lemma 1 (Diagonal Lemma)**.** *Let $\varphi(x)$ be a formula with exactly one free variable. Then there is a sentence ψ such that $\psi \leftrightarrow \varphi(\ulcorner \psi \urcorner)$ is provable.*

Proof. Let $\delta(y)$ be the formula $\varphi\big(Sub(y, y)\big)$, and ψ be $\delta(\ulcorner \delta \urcorner)$. Then:

$$\begin{aligned} \vdash \psi &\longleftrightarrow \delta(\ulcorner \delta \urcorner) \\ &\longleftrightarrow \varphi\big(Sub(\ulcorner \delta \urcorner, \ulcorner \delta \urcorner)\big) \\ &\longleftrightarrow \varphi(\ulcorner \delta(\ulcorner \delta \urcorner) \urcorner) \\ &\longleftrightarrow \varphi(\ulcorner \psi \urcorner) \end{aligned}$$

$\square$

If ψ is a sentence such that $\vdash \psi \longleftrightarrow \varphi(\ulcorner \psi \urcorner)$, then we will call ψ a *fixed point* of $\varphi(x)$.

2.10 Consistency, completeness, ω-consistency

The following are three properties of a formal system that will be used later.

A formal system is *consistent* if there is no sentence φ such that both φ and $\neg\varphi$ are provable. In other words, no sentence is both provable and refutable.

A formal system is *complete* if for every sentence φ either φ or $\neg\varphi$ is provable. In other words, every sentence is either provable or refutable.

A formal system is *ω-consistent* if it has representations for all natural numbers, and there is no formula $\varphi(x)$ with one free variable x such that for every

[4] There may be more than one.

natural number n, the formula $\varphi(\overline{n})$ is provable (where $\overline{n}$ is a representation of n in the system), but $\neg\forall x\varphi(x)$ is also provable. Or equivalently, there is no formula $\varphi(x)$ with one free variable x such that for every natural number n the formula $\varphi(\overline{n})$ is not provable but $\exists x\varphi(x)$ is provable.

For our system, ω-consistency implies consistency, since it is using classical logic, if it is inconsistent then every formula is provable, particularly $\neg\forall x\varphi(x)$ for every $\varphi(x)$.

Also, if **T** is ω-consistent and $\vdash \exists x\varphi(x)$ for some φ, then there is a number n such that $\vdash \varphi(\overline{n})$. Otherwise φ would be a counter-example that makes **T** not ω-consistent.

2.11 Gödel's First Incompleteness Theorem

Here we prove two lemmas concerning consistency, ω-consistency and our provability predicate $Prov(x)$:

Lemma 2. *If* $\vdash \varphi$ *then* $\vdash Prov(\ulcorner\varphi\urcorner)$.

Proof. Without loss of generality, suppose **T** is consistent[5] and $\vdash \varphi$. Then there is a proof of φ, let it be $\varphi_1, \varphi_2, \ldots \varphi_n$, where φ_n is the formula φ. Let $\ulcorner\varphi_1, \varphi_2, \ldots \varphi_n\urcorner = k$, the Gödel number of the proof.

By the definitions of $Proof(x)$ and $Prf(x, y)$, we have $\vdash Proof(k)$ and $\vdash Prf(k, \ulcorner\varphi\urcorner)$, hence we have $\vdash Prov(\ulcorner\varphi\urcorner)$. □

Lemma 3. *If* ***T*** *is* ω*-consistent and* $\vdash Prov(\ulcorner\varphi\urcorner)$ *for some formula* φ*, then* $\vdash \varphi$.

Proof. Suppose **T** is ω-consistent. Suppose on the contrary that there is a formula φ such that $\vdash Prov(\ulcorner\varphi\urcorner)$ but φ is not provable.

Then $\vdash \exists x Prf(x, \ulcorner\varphi\urcorner)$, there is a number n such that $\vdash Prf(n, \ulcorner\varphi\urcorner)$.

However this means that there is a number n which is the Gödel number of a proof of φ, contradicting our assumption that φ is not provable. □

Now we can prove the First Incompleteness Theorem:

[5] Otherwise, every formula is provable, including $Prov(\ulcorner\varphi\urcorner)$,

Theorem 4 (Gödel's First Incompleteness Theorem)**.** *If $\boldsymbol{T}$ is ω-consistent, then there is a sentence* $\mathbf{G}$ *such that neither* $\vdash \mathbf{G}$ *nor* $\vdash \neg\mathbf{G}$*. In other words, if $\boldsymbol{T}$ is ω-consistent then $\boldsymbol{T}$ is incomplete.*

Proof. Suppose $\mathbf{T}$ is ω-consistent. By the Diagonal Lemma (Lemma 1), there is a sentence G such that

$$\mathbf{G} \longleftrightarrow \neg Prov(\ulcorner \mathbf{G} \urcorner) \tag{1}$$

is provable.

Suppose on the contrary that $\vdash \mathbf{G}$, applying the inference rule (MP) to G and (1), we have $\vdash \neg Prov(\ulcorner \mathbf{G} \urcorner)$. On the other hand, by Lemma 2 we have if $\vdash \mathbf{G}$ then $\vdash Prov(\ulcorner \mathbf{G} \urcorner)$. But this means that $\mathbf{T}$ is inconsistent, which is impossible since $\mathbf{T}$ is ω-consistent.

Suppose on the contrary that $\vdash \neg\mathbf{G}$. (1) implies that $\vdash \neg\mathbf{G} \longleftrightarrow Prov(\ulcorner \mathbf{G} \urcorner)$. Applying (MP) we have $\vdash Prov(\ulcorner \mathbf{G} \urcorner)$. But by Lemma 3 we get $\vdash \mathbf{G}$, again this means that $\mathbf{T}$ is inconsistent, which is impossible.

Therefore, both $\mathbf{G}$ and $\neg\mathbf{G}$ are not provable. $\square$

2.12 Two more proofs

Proof using a refutability predicate

We have already define $Prov(x)$, now we define another predicate by:

$$Ref(x) \longleftrightarrow \exists y Prf\big(y, Neg(x)\big)$$

Note that $Ref(x)$ is actually $Prov(Neg(x))$.

Using $Ref(x)$ we can have another proof of Gödel's first incompleteness theorem.

Proof. Suppose $\mathbf{T}$ is ω-consistent.

Let $\mathbf{G}^*$ be the sentence such that $\vdash G^* \longleftrightarrow Ref(\ulcorner G^* \urcorner)$. The existence of this sentence is guaranteed by the Diagonal Lemma. Now we will argue that ψ is neither provable nor refutable.

Suppose on the contrary that $\vdash \mathbf{G}^*$. By the choice of $\mathbf{G}^*$ we have $\vdash Ref(\ulcorner \mathbf{G}^* \urcorner)$, i.e. $\vdash Prov(\ulcorner \neg\mathbf{G}^* \urcorner)$. And by Lemma 3 this implies $\vdash \neg\mathbf{G}^*$, hence $\mathbf{T}$ is not consistent, contradicting our assumption.

Suppose on the contrary that $\vdash \neg\mathbf{G}^*$, by Lemma 2 we have $\vdash Prov(\ulcorner\neg\mathbf{G}^*\urcorner)$. On the other hand, by the choice of $\mathbf{G}^*$ we have $\vdash \neg Ref(\ulcorner\mathbf{G}^*\urcorner)$, i.e. $\vdash \neg Prov(\ulcorner\neg\mathbf{G}^*\urcorner)$. Then $Prov(\ulcorner\neg\mathbf{G}^*\urcorner)$ is both provable and refutable, hence **T** is not consistent, contradicting our assumption.

Therefore, both $\mathbf{G}^*$ and $\neg\mathbf{G}^*$ are not provable. $\square$

Rosser's Theorem

In fact, we can get rid of the assumption that **T** is ω-consistent. This is a result first proven in Rosser (1936). To prove this result, we need to define another provability predicate $Prov_R(x)$:

$$Prov_R(x) \longleftrightarrow \exists y\big(Prf(y,x) \wedge (\forall z \leq y)\neg Prf(z, Neg(x))\big)$$

For $Prov_R(x)$, we have a result similar to Lemma 2:

Lemma 5. *If* ***T*** *is consistent and* $\vdash \varphi$*, then* $Prov_R(\ulcorner\varphi\urcorner)$*.*

Proof. Suppose **T** is consistent and $\vdash \varphi$. Then there is a proof of φ, let p be the number of this proof. By consistency there is no proof of $\neg\varphi$, particularly no proof of $\neg\varphi$ which has a Gödel number less than p.

Therefore $\vdash Prf(p,\ulcorner\varphi\urcorner) \wedge (\forall z \leq p)\neg Prf(z,\ulcorner\neg\varphi\urcorner)$, which implies $\exists y\big(Prf(y,\ulcorner\varphi\urcorner) \wedge (\forall z \leq y)\neg Prf(z, Neg(\ulcorner\varphi\urcorner))\big)$ i.e. $\vdash Prov_R(\ulcorner\varphi\urcorner)$. $\square$

We also need the following lemma which will be stated without a proof:

Lemma 6. $\vdash \forall x \forall y(x \leq y \vee y \leq x)$

Now we can prove Rosser's result:

Theorem 7. *Rosser's Theorem If* ***T*** *is consistent, then there is a sentence* **R** *such that neither* **R** *nor* $\neg$**R** *is provable.*

Proof. Suppose **T** is consistent. By the Diagonal Lemma, there is a sentence **R** such that $\vdash \mathbf{R} \longleftrightarrow \neg Prov_R(\ulcorner\mathbf{R}\urcorner)$.

Suppose $\vdash \mathbf{R}$. By Lemma 5 we have $\vdash Prov_R(\ulcorner\mathbf{R}\urcorner)$. On the other hand, by the choice of R we also have $\vdash \neg Prov_R(\ulcorner\mathbf{R}\urcorner)$, which means $Prov_R(\ulcorner\mathbf{R}\urcorner)$ is both provable and refutable, contradicting to our assumption.

Suppose $\vdash \neg\mathbf{R}$. By Lemma 5 we have $\vdash Prov_R(\ulcorner\neg\mathbf{R}\urcorner)$. Let p be the Gödel number of a proof of $\neg\mathbf{R}$, hence we have $\vdash Prf(p,\ulcorner\neg\mathbf{R}\urcorner) \wedge (\forall z \leq p)\neg Prf(z,\ulcorner\neg\neg\mathbf{R}\urcorner)$.

By the choice of $\mathbf{R}$ we also have $\vdash Prov_R(\ulcorner\mathbf{R}\urcorner)$, that is, $\vdash \exists y\big(Prf(y,\ulcorner\mathbf{R}\urcorner) \wedge (\forall z \leq y)\neg Prf(z, Neg(\ulcorner\mathbf{R}\urcorner))\big)$. Therefore, for some q we have $\vdash Prf(q,\ulcorner\mathbf{R}\urcorner) \wedge (\forall z \leq q)\neg Prf(z, Neg(\ulcorner\mathbf{R}\urcorner))\big)$.

Applying Lemma 6, we have $p \leq q \vee q \leq p$. If $p \leq q$, then $\vdash Prf(q,\ulcorner\mathbf{R}\urcorner) \wedge \neg Prf(p,\ulcorner\neg\mathbf{R}\urcorner)$, resulting the sentence $Prf(p,\ulcorner\neg\mathbf{R}\urcorner)$ be both provable and refutable, contradicting our assumption. Similarly, if $q \leq p$ we will get another contradiction.

Therefore both $\mathbf{R}$ and $\neg\mathbf{R}$ are not provable. $\square$

3 The Second Incompleteness Theorem

In the following we will prove the Gödel's Second Incompleteness Theorem, which is often said to be "proving that mathematics cannot prove itself consistent". Putting aside what it means by "prove itself consistent", it is well known that in an inconsistent theory anything can be proven, so even if "mathematics" can "proves itself consistent", it does not follows that mathematics is in fact consistent.

A more rigorous formulation of the theorem is the following: If $\mathbf{T}$ is consistent, then the sentence $\neg Prov(\ulcorner 0 = 1\urcorner)$ is not provable.

It should be noted that the above theorem depends on the definition of $Prov(x)$. For example, let us define $Prov^\star(x) \longleftrightarrow Prov(x) \wedge (x \neq \ulcorner 0 = 1\urcorner)$, then if $\mathbf{T}$ is consistent, $Prov(x)$ is provable if and only if $Prov^\star(x)$ is provable. This is because if $Prov(x)$ is provable while $Prov^\star(x)$ is not, then the only possibility is that $Prov(x) \wedge x = \ulcorner 0 = 1\urcorner$, which is impossible by consistency. (The other direction is trivial.) However $\neg Prov^\star(\ulcorner 0 = 1\urcorner)$ is provable, since it is equivalent to $\neg Prov(\ulcorner 0 = 1\urcorner) \vee (\ulcorner 0 = 1\urcorner = \ulcorner 0 = 1\urcorner)$ which is obviously true.[6]

[6] Ferenc Csaba suggested this point to me, and the idea can be found in Lindström (1997).

3.1 About $Prov(x)$

Before presenting the proof, a little discussion on the meaning of the sentence $\neg Prov(\ulcorner 0 = 1 \urcorner)$. It is regarded by some people as "saying that $\mathbf{T}$ is consistent", because they regard $Prov(x)$ is "saying that x is the Gödel number of a provable formula". Hence $\neg Prov(\ulcorner 0 = 1 \urcorner)$ is "saying that $0 = 1$ is not provable" for them.

However, we have to be careful that though the intended meaning of $Prov(\ulcorner \varphi \urcorner)$ is to imitate the predicate "φ is provable" in the metatheory, it does not fully represent or express that meaning. For, the "not" in our metatheory is represented by the symbol "$\neg$" in $\mathbf{T}$, but "φ not provable" is different from $\vdash \neg Prov(\ulcorner \varphi \urcorner)$. For example, we know that $\mathbf{G}$, the Gödel sentence in the proof of first incompleteness theorem, is not provable, at the same time we do not have $\vdash \neg Prov(\ulcorner \mathbf{G} \urcorner)$.

Moreover, it is not just the problem of our choice of $Prov(x)$, we have the following result:

Proposition 8. *If $\boldsymbol{T}$ is consistent, then it is impossible to have a predicate $P(x)$ satisfying the following two conditions:*

1. *If $\vdash \varphi$ then $\vdash P(\ulcorner \varphi \urcorner)$.*

2. *If $\nvdash \varphi$ then $\vdash \neg P(\ulcorner \varphi \urcorner)$.*

Proof. Otherwise consider a sentence ψ such that $\vdash \psi \leftrightarrow \neg P(\ulcorner \psi \urcorner)$, whose existence is guaranteed by the Diagonal Lemma.

Suppose $\vdash \psi$, then by the choice of ψ we have $\vdash \neg P(\ulcorner \varphi \urcorner)$. But by our assumption on $P(x)$ this implies $\nvdash \psi$.

On the other hand, suppose $\nvdash \psi$, then $\vdash \neg P(\ulcorner \varphi \urcorner)$. But by the choice of ψ we have $\vdash \psi$.

In both cases there is a contradiction, hence such a $P(x)$ cannot exist. $\square$

Remember now we have:

1. If $\mathbf{T}$ is consistent and $\vdash \varphi$, then $\vdash P(\ulcorner \varphi \urcorner)$.

2. If $\mathbf{T}$ is ω-consistent and $\vdash P(\ulcorner \varphi \urcorner)$, then $\vdash \varphi$.

In other words, if $\mathbf{T}$ is ω-consistent, then $\vdash \varphi$ if and only if $\vdash P(\ulcorner \varphi \urcorner)$.

It is natural to ask whether it is possible to weaken the assumption from ω-consistent to consistency. After this section we will see that it is impossible under certain assumption.

3.2 Derivability conditions

A predicate $P(x)$ is a provability predicate if it satisfies the following three conditions:

1. If $\vdash \varphi$ then $\vdash P(\ulcorner \varphi \urcorner)$
2. $\vdash P(\ulcorner \varphi \to \psi \urcorner) \to \big(P(\ulcorner \varphi \urcorner) \to P(\ulcorner \psi \urcorner)\big)$
3. $\vdash P(\ulcorner \varphi \urcorner) \to P\big(\ulcorner P(\ulcorner \varphi \urcorner) \urcorner\big)$

These three conditions are proposed by Martin Löb in Löb (1955). We called $Prov(x)$ a provability predicate before, so it is not difficult to guess $Prov(x)$ satisfies the above conditions. Indeed, we have proven (1) in our Lemma 2.

Intuitively, (2) and (3) are "pushing down" modus ponens and (1) into the object level by using $Prov(x)$. Unfortunately the proofs of them – as said in Boolos et al. (2007) – "would take up too much time and patience". In the following we will simply assume them to be true. For more details, see Ch.26 of Smith (2007) for a "sketch of a proof sketch", and Ch.2 of Boolos (1993) for a sketch of a proof.

For readability, instead of writing $Prov(\ulcorner \varphi \urcorner)$, which will be used iteratively a lot of times, we will use a new notation $\Box\varphi$ in the rest of this section. For example, for the Gödel sentence $\mathbf{G}$ in the above proof of first incompleteness theorem, we know that $\vdash \mathbf{G} \longleftrightarrow \neg\Box\mathbf{G}$. Also, the sentence "$0 = 1$" will be abbreviated as "$\bot$".

So we have the following:

D1 If $\vdash \varphi$ then $\vdash \Box\varphi$.

D2 $\vdash \Box(\varphi \to \psi) \to \big(\Box\varphi \to \Box\psi\big)$

D3 $\vdash \Box\varphi \to \Box\Box\varphi$

3.3 Löb's Theorem and Second Incompleteness Theorem

In Löb (1955), Martin Löb solved a problem proposed by Leon Henkin, concerning whether a fixed point of $Prov(x)$ is provable or not. That is, given that $\vdash \varphi \leftrightarrow \Box\varphi$, is φ itself provable?

Löb's answer is positive, and we have the following theorem:

Theorem 9 (Löb's Theorem)**.** *If $\boldsymbol{T}$ is consistent, and $\vdash \Box\varphi \to \varphi$, then $\vdash \varphi$.*

Proof. Suppose $\vdash \Box\varphi \to \varphi$. By applying the Diagonal Lemma to the formula $Prov(x) \to \varphi$, we will get a sentence ψ such that $\vdash \psi \leftrightarrow (\Box\psi \to \varphi)$. Then:

1. $\vdash \psi \leftrightarrow (\Box\psi \to \varphi)$	
2. $\vdash \psi \to (\Box\psi \to \varphi)$	
3. $\vdash \Box(\psi \to (\Box\psi \to \varphi))$	(By D1)
4. $\vdash \Box\psi \to \Box(\Box\psi \to \varphi)$	(By D2)
5. $\vdash \Box\psi \to (\Box\Box\psi \to \Box\varphi)$	(By D2)
6. $\vdash \Box\psi \to \Box\Box\psi$	(By D3)
7. $\vdash \Box\psi \to \Box\varphi$	(By 6,7)
8. $\vdash \Box\varphi \to \varphi$	(By assumption)
9. $\vdash \Box\psi \to \varphi$	(By 7,8)
10. $\vdash \psi$	(By 9,1)
11. $\vdash \Box\psi$	(By D1)
12. $\vdash \varphi$	(By 9,11)

□

Now we can rephrase (in our new notation) and prove the Second Incompleteness Theorem from Löb's theorem:

Theorem 10 (Gödel's Second Incompleteness Theorem)**.** *If $\boldsymbol{T}$ is consistent, then $\nvdash \neg\Box\bot$.*

Proof. Suppose $\vdash \neg\Box\bot$. Then:

$$\begin{aligned} &\vdash \Box\bot \to \neg\neg\Box\bot && \text{(By a basic fact in classical logic)} \\ &\vdash \neg\neg\Box\bot \to \bot && \text{(By consistency of } \mathbf{T}\text{)} \\ \Rightarrow &\vdash \Box\bot \to \bot && \text{(By MP)} \\ \Rightarrow &\vdash \bot && \text{(By Löb's theorem)} \end{aligned}$$

Which implies that **T** is inconsistent. Therefore if **T** is consistent, then $\neg\Box\bot$ is not provable. □

From the Second Incompleteness Theorem we have the following corollary:

Lemma 11. *If **T** is consistent, then for any formula φ, $\nvdash \neg\Box\varphi$.*

Proof. Let φ be any formula. Then:

$$\begin{aligned} &\vdash \bot \to \varphi && \text{(By a basic fact in classical logic)} \\ \Rightarrow &\vdash \Box(\bot \to \varphi) && \text{(By D1)} \\ \Rightarrow &\vdash \Box\bot \to \Box\varphi && \text{(By D2)} \\ \Rightarrow &\vdash \neg\Box\varphi \to \neg\Box\bot && \text{(By a basic fact in classical logic)} \end{aligned}$$

Hence if $\neg\Box\varphi$ is provable, by modus ponens we will have $\vdash \neg\Box\bot$, contradicting the Second Incompleteness Theorem. Therefore for any formula φ, $\neg\Box\varphi$ is not provable. □

From the above corollary, we can have one more result:

Corollary 12. *If **T** is ω-consistent and φ is refutable, then $\Box\varphi$ is undecidable.*

Proof. φ is refutable, therefore $\vdash \neg\varphi$. By the consistency of **T**, we know that $\nvdash \varphi$. Using the contrapositive of Lemma 3, $\nvdash \varphi$ implies $\nvdash \Box\varphi$, that is, $\Box\varphi$ is not provable.

By Lemma 11, $\neg\Box\varphi$ is not provable. Therefore $\Box\varphi$ is undecidable. □

3.4 Provability predicate again

In the end of Section 3.1, we asked this question: given that **T** is consistent, is it possible to have a predicate $P(x)$ such that for any formula φ, $\vdash \varphi$ if and only if $\vdash P(\ulcorner\varphi\urcorner)$? A partial answer is the following:

Proposition 13. *If* $\boldsymbol{T}$ *is consistent, there is no* $P(x)$ *satisfying the following conditions:*

1. *For any formula* φ*,* φ *is provable if and only if* $P(\ulcorner\varphi\urcorner)$ *is provable;*
2. *The derivability conditions D1, D2, and D3;*
3. *For any formula* φ*,* $P(\ulcorner P(\ulcorner\varphi\urcorner)\urcorner) \to P(\ulcorner\varphi\urcorner)$ *is provable.*

Proof. Suppose such a $P(x)$ exists. Let $\dot{\Box}\varphi$ abbreviates $P(\ulcorner\varphi\urcorner)$. Let φ be a formula, then:

$$\begin{aligned} & \vdash \dot{\Box}\dot{\Box}\varphi \to \dot{\Box}\varphi && \text{(By the third condition)} \\ \Rightarrow & \vdash \dot{\Box}\varphi && \text{(By Löb's theorem)} \\ \Rightarrow & \vdash \varphi && \text{(By the first condition)} \end{aligned}$$

But φ is arbitrary, this would lead to the inconsistency of **T**. Hence by contradiction there is no such a $P(x)$. □

4 Grelling's Paradox

It is usually said that the proof of first incompleteness theorem is related to the liar paradox, in a way that the Gödel sentence G is "saying" that "I am not provable" or "This sentence is not provable". However it is quite inaccurate, even misleading, as there is no indexical words like "I" or "this" used in the proof and in the language.

In this section, we will see that the formalization of Grelling's paradox yields two undecidable sentences, and one of them is the Gödel sentence in the previous section.[7] [8]

Grelling's paradox is a semantical paradox, a popular version of the paradox is like this:

[7]In Boolos et al. (2007), an undecidable sentence obtained from formalizing Grelling's paradox is briefly sketched. But it seems that the authors do not notice the relation between this sentence and the original Gödel sentence, as they say that it is an "undecidable sentence without the Diagonal Lemma", which is arguably not the case.

[8]After finishing the first draft of this section, I find a similar idea which turns the Grelling's paradox into a proof of the Diagonal Lemma in Serény (2006).

> Let us call an adjective *autological* if it can be applied to itself, and *heterological* if it cannot. For example "polysyllabic" is polysyllabic, so it is an autological adjective; "monosyllabic" is not monosyllabic, so it is heterological. The problem is: is "heterological" heterological?
>
> If "heterological" is heterological, then it is autological and cannot be heterological. If it is not, then it cannot be applied to itself, which means it is heterological. In short, "heterological" is heterological if and only if it is not.

Since we are trying to talk about it in formal languages, it seems to be better if we talks about predicates. Hence we have:

> A predicate P is *autological* if $P(P)$ is true, and *heterological* if $P(P)$ is false.

But there is already a problem, it is meaningless to put a predicate as its own argument. In mathematical logic, the usual interpretation of an $(n+1)^{st}$-order predicate is a set of n^{th}-order objects, and in any well-founded set theory, a set cannot be an element of itself. In fact, this leads us to the Russell's paradox in naive set theory if we are talking about sets instead of predicates.

Take a closer look to the adjective version of Grelling's paradox, we find that we are saying "polysyllabic' is polysyllabic", not "polysyllabic is polysyllabic". In other words, we are talking about whether the word "heterological" is satisfying the property of being heterological. The *word* "heterological" is the name of the *property* of being heterological. So we can write in this way:

> A predicate P with a name 'P' is *autological* if P('P') is true, and *heterological* if P('P') is false. For the predicate H which expresses heterologicality, H('H') is true if and only if it is false.

4.1 $Aut(x)$, $Het_1(x)$ and $Het_2(x)$

Now we have to ensure the names of predicates can be an argument of the predicate, otherwise it is meaningless again. By the arithmetization of syntax, for any predicate $F(x)$ there is a Gödel number $\ulcorner F(x) \urcorner$, and we can talk about $F(\ulcorner F(x) \urcorner)$.

Hence we can define F to be *autological* if it is provable that $F(\ulcorner F(x) \urcorner)$, i.e. $\vdash F(\ulcorner F(x) \urcorner)$.

Since we have defined a provability predicate $Prov(x)$ and a substitution relation $Sub(x, y)$ such that $Sub(\ulcorner F(x) \urcorner, k) = \ulcorner F(k) \urcorner$ for any predicate $F(x)$ and number k, we can define the autological predicate $Aut(x)$ by the following:

$$Aut(x) \overset{def}{\Longleftrightarrow} Prov(Sub(x, x))$$

And we have two choices for the definition of the heterological predicate, depending on how we interpret "false". So here we use $Het_1(x)$ and $Het_2(x)$ to denote them:

$$\begin{aligned} Het_1(x) &\overset{def}{\Longleftrightarrow} \neg(Prov(Sub(x, x))) \\ Het_2(x) &\overset{def}{\Longleftrightarrow} Prov(Neg(Sub(x, x))) \end{aligned}$$

Let $h_1 = \ulcorner Het_1(x) \urcorner$, $h_2 = \ulcorner Het_2(x) \urcorner$. Then we have the following results:

Theorem 14. *If* $\boldsymbol{T}$ *is* ω*-consistent, then* $Het_1(h_1)$ *and* $Het_2(h_2)$ *are both undecidable.*

Proof. Suppose $\vdash Het_1(h_1)$.

$$\begin{aligned} \vdash Het_1(h_1) &\Rightarrow \vdash \neg(Prov(Sub(h_1, h_1))) \\ &\Rightarrow \vdash \neg(Prov(\ulcorner Het_1(h_1) \urcorner)) \end{aligned}$$

At the same time we have:

$$\vdash Het_1(h_1) \Rightarrow \vdash Prov(\ulcorner Het_1(h_1) \urcorner)$$

By consistency this is impossible, hence $Het_1(h_1)$ is not provable. On the other hand:

$$\begin{aligned} \vdash \neg Het_1(h_1) &\Rightarrow \vdash \neg\neg(Prov(Sub(h_1, h_1))) \\ &\Rightarrow \vdash Prov(Sub(h_1, h_1)) \\ &\Rightarrow \vdash Het_1(h_1) && \text{(by } \omega\text{-consistency)} \end{aligned}$$

By consistency this is impossible. Therefore $Het_1(h_1)$ is neither provable nor refutable.

Similarly, suppose $\vdash Het_2(h_2)$

$$\begin{aligned}
\vdash Het_2(h_2) &\Rightarrow \vdash Prov(Neg(Sub(h_2, h_2))) \\
&\Rightarrow \vdash Neg(Sub(h_2, h_2)) && \text{(by ω-consistency)} \\
&\Rightarrow \vdash Neg(\ulcorner Het_2(h_2)\urcorner) \\
&\Rightarrow \vdash \neg Het_2(h_2)
\end{aligned}$$

By consistency $Het_2(h_2)$ is not provable. On the other hand:

$$\begin{aligned}
\vdash \neg Het_2(h_2) &\Rightarrow \vdash Prov(\ulcorner \neg Het_2(h_2)\urcorner) \\
&\Rightarrow \vdash Prov(Neg(Sub(h_2, h_2))) \\
&\Rightarrow \vdash Het_2(h_2) && \text{(by the definition of Het_2.)}
\end{aligned}$$

By consistency this is impossible, therefore $Het_2(h_2)$ is neither provable nor refutable. $\square$

We now have two undecidable sentences, $Het_1(h_1)$ and $Het_2(h_2)$. Looking into the details of the proof of the Diagonal Lemma, we will find that $Het_1(h_1)$ is actually what we get when we apply the lemma to the predicate $\neg Prov(x)$, i.e. $\neg Prov(Sub(x, x))$, so $Het_1(h_1)$ is the usual Gödel sentence; and similarly $Het_2(h_2)$ is the result of applying the lemma to $Prov(Neg(x))$, which is the refutability predicate $Ref(x)$.

5 Curry's Paradox

In Curry (1942), Haskell Curry presented a paradox which is later called Curry paradox. The original paper deals with a certain formal systems, but the paradox can also be presented in an informal way.

Consider the following sentence X: "If X is true, then the Earth is flat". Here the "if ..., then ..." clause is understood as the material conditional, so X says that either X is not true, or the Earth is flat.

To prove a conditional sentence, we can first suppose the antecedent, then try to derive the consequent. If we can do that, then the conditional sentence is proven.

So let us suppose the antecedent, i.e. X is true. Then "If X is true, then the Earth is flat." is true, and by modus ponens we derive that the Earth is flat, given that X is true. Therefore we have just “proven” that “If X is true, then the Earth is flat”.

But “If X is true, then the Earth is flat” is the very sentence X, and by modus ponens again we can derive that the Earth is flat, which violates modern science yet proven by logic.

Furthermore, we can substitute any sentence to “the Earth is flat”, and using nearly the same “proof” we can derive any sentence we want. Obviously something goes wrong, and this is an informal version of the Curry’s paradox. We will call any sentence of the form “If this sentence is true, then A” or “If X is true, then A”, where X is the same sentence, a Curry sentence.

Though first discovered by Curry, this paradox is also called Löb’s paradox, since it is closely related to Löb’s theorem and mentioned in his Löb (1955). In the following we can see how to formalize this paradox to obtain an undecidable sentence.

5.1 Fomalized Curry sentence

As in the case of the Liar paradox, we need the Diagonal Lemma to imitate the Curry paradox in our system **T**. Therefore consider the following formula with one free variable:

$$Prov(x) \to (0 = 1)$$

and let **C** be a fixed point of this formula. Then we have:

$$\vdash \mathbf{C} \longleftrightarrow \big(Prov(\ulcorner \mathbf{C} \urcorner) \to (0 = 1)\big)$$

Now we can prove the following theorem:

Theorem 15. *If* ***T*** *is ω-consistent, then neither* **C** *nor* $\neg$**C** *is provable.*

Proof. Suppose $\vdash \mathbf{C}$. By Lemma 2 we have $\vdash Prov(\ulcorner \mathbf{C} \urcorner)$. By the choice of **C** we have $\vdash Prov(\ulcorner \mathbf{C} \urcorner) \to (0 = 1)$. By modus ponens we get $\vdash 0 = 1$, contradicting the assumption that **T** is consistent.

Suppose $\vdash \neg\mathbf{C}$. By the choice of **C** we have $\vdash \neg\big(Prov(\ulcorner \mathbf{C} \urcorner) \to (0 = 1)\big)$, that is $\vdash Prov(\ulcorner \mathbf{C} \urcorner) \wedge \neg(0 = 1)$. Therefore we get $\vdash Prov(\ulcorner \mathbf{C} \urcorner)$. Since we assume that **T** is ω-consistent, by Lemma 3 we get $\vdash \mathbf{C}$, again contradicting the assumption that **T** is consistent.

Therefore **C** is undecidable. $\square$

We also have another similar undecidable sentence $\mathbf{C}'$, which is a fixed point of $Prov(Imp(x, \ulcorner(0=1)\urcorner))$. That is,

$$\vdash \mathbf{C}' \longleftrightarrow Prov(\ulcorner\mathbf{C}' \to (0=1)\urcorner))$$

Theorem 16. *If* $\boldsymbol{T}$ *is ω-consistent, then neither* $\mathbf{C}'$ *nor* $\neg\mathbf{C}'$ *is provable.*

Proof. Suppose $\vdash \mathbf{C}'$, by the choice of $\mathbf{C}'$ we have $\vdash Prov(\ulcorner\mathbf{C}' \to (0=1)\urcorner))$. Since we assume $\mathbf{T}$ is ω-consistent, by Lemma 3 we get $\vdash \mathbf{C}' \to (0=1)$. By our hypothesis and modus ponens we get $\vdash 0 = 1$, contradicting our assumption that $\mathbf{T}$ is consistent.

Suppose $\vdash \neg\mathbf{C}'$, by the choice of $\mathbf{C}'$ we have $\vdash \neg Prov(\ulcorner\mathbf{C}' \to (0=1)\urcorner))$. By Corollary 11, this is impossible.

Therefore neither $\mathbf{C}'$ nor $\neg\mathbf{C}'$ is provable. $\square$

In fact, $\mathbf{C}$ is a fixed point of $\neg Prov(x)$, that is:

Theorem 17. $\vdash \mathbf{C} \longleftrightarrow \neg Prov(\ulcorner\mathbf{C}\urcorner)$

To prove the above theorem, we need the following lemma:

Lemma 18. *Let φ be a formula, then* $\vdash \big(\varphi \to (0=1)\big) \longleftrightarrow \neg\varphi$

Proof. Let φ be a formula. Suppose $\neg\varphi$ is provable then by classical logic so is $\neg(0=1) \to \neg\varphi$, and the latter is equivalent to $\varphi \to (0=1)$.

Suppose $\varphi \to (0=1)$ is provable, then we have $\neg(0=1) \to \neg\varphi$. And since $\neg(0=1)$ is provable, by modus ponens we know that $\neg\varphi$ is provable.

Therefore, for any formula φ, $\vdash \varphi \to (0=1) \longleftrightarrow \neg\varphi$. $\square$

Now the proof of Theorem 17 becomes trivial:

Proof. By the choice of $\mathbf{C}$, we have $\vdash \mathbf{C} \longleftrightarrow \big(Prov(\ulcorner\mathbf{C}\urcorner) \to (0=1)\big)$. By Lemma 18, we have $\vdash \big(Prov(\mathbf{C}) \to (0=1)\big) \longleftrightarrow \neg Prov(\mathbf{C})$.

Therefore $\vdash \mathbf{C} \longleftrightarrow \neg Prov(\ulcorner\mathbf{C}\urcorner)$. $\square$

Hence we can see that $\mathbf{C}$ is a fixed point of $\neg Prov(x)$. Similarly we can prove that $\mathbf{C}'$ is a fixed point of $Prov(Neg(x))$.

6 Berry Paradox

The Berry paradox, first published in Russell (1967), is a paradox about the names and descriptions of natural numbers.

Let us consider the definite descriptions of any natural number, which are those descriptions in which exactly one number satisfies. For example, "the number which is not a successor of another number" is a definite description of the number 0, "the smallest prime number" is a definite description of the number 2.

For convenience let us just consider those descriptions formed by English words. There are 26 letters in the alphabet, plus a space, so there are 27^k possible combinations for expressions of length k (space included), and there are $1 + 27 + 27^2 + \cdots + 27^{k-1}$ possible combinations for expressions of length less than k.

These numbers are, of course, finite. Which means for every natural number k, there must be infinitely many numbers that cannot be definitely described in less than k letters. As the set of natural numbers is well-ordered, there is a *least* number that cannot be definitely described in less than k letters, for any number k.

Now consider the definite description "the least number that cannot be definitely described in less than 80 letters", which must be a definite description of some number n. Nevertheless the above description is less than 80 letters long[9], therefore it cannot be a definite description of this n.

This is a version of Berry paradox, other versions involves syllables or words instead of letters, definitions or names instead of definite descriptions, but the key idea is the same.

In this section we will see two informal proofs of Gödel first incompleteness theorems involving Berry paradox. One is from Boolos (1998), another is from Chaitin (1971). Both proofs below are presented in an informal way, and then there will be a more formalized proof.[10]

In the following, we need to talk about the formal expression of a number

[9] Count it if you want.

[10] It is a matter of degree of what is considered "formalized", as most proofs are not totally formalized.

n, to avoid confusion we will denote them with an overline. That is, the formal expression of n (which is a number, not a variable) is $\overline{n}$, i.e.$\underbrace{SS\ldots S}_{n}0$.

6.1 Boolos's proof

In this proof, there is a slight difference between the language used here and the rest of the article. For example, Boolos's variables are $x, x', x'', x''', \ldots$ instead of $v_0, v_1, v_2, \ldots$, hence all his variables are built from the symbols x and $'$. Again we will abbreviate the variables by single letters.

And the main difference between what we have done so far and Boolos's proof is that Boolos uses a semantic argument. It is assumed that we are in the standard model of first-order Peano arithmetic. That means we have the concept of truth and falsity (in the model with ω, the set of natural numbers).

Now suppose our formal system **T** is sound, that is, every theorem of **T** is true. We say that a formula $\varphi(x)$ *names* a number n if $\forall x(\varphi(x) \leftrightarrow x = \overline{n})$ is provable.

Note that every formula can name at most one number. Otherwise, suppose $m \neq n$ and $F(x)$ names both m and n, then $\forall x(\varphi(x) \leftrightarrow x = \overline{n})$ and $\forall x(\varphi(x) \leftrightarrow x = \overline{m})$ are both provable. By substitution we will get both $\varphi(\overline{m}) \leftrightarrow \overline{m} = \overline{n}$ and $\varphi(\overline{m}) \leftrightarrow \overline{m} = \overline{m}$, then $(\overline{m} = \overline{m}) \leftrightarrow (\overline{m} = \overline{n})$ is provable, contradicting our assumption that **T** is sound.

Also note that for every number k, there are finitely many numbers that can be named by a formula with less than k symbols. This is because we have finitely many symbols (including the variables, as we use x and $'$ to denote them) in our language, so there are finitely many formulas with less than k symbols. And as we have seen, every formula can name at most one number, even if every formula with less than k symbols names a unique number, the total number of numbers named is still finite.

Since natural numbers are well-ordered[11], we have the following:

Lemma 19. *For every number k there is a least number cannot be named by formulas with less than k symbols.*

Let $C(x, z)$ be a formula saying that x is named by a formula containing z

[11]That means in every subset of natural number there is a least element in it.

symbols, and $B(x,y)$ be the formula $\exists z(z < y \wedge C(x,z))$, which says that x is named by a formula containing less than y symbols. Let $A(x,y)$ be the formula $\neg B(x,y) \wedge \forall a\big(a < x \to B(a,y)\big)$, which says that x is the least number not named by any formula containing less than y symbols.

Then let k be the number of symbols in $A(x,x')$, obviously $k > 3$, and $F(x)$ be the formula $\exists x'\big((x' = \overline{10} \times \overline{k}) \wedge A(x,x')\big)$, which says that x is the least number that cannot be named by any formula containing less than $10k$ symbols.

After that, let us count the number of symbols in $F(x)$: the formal expression of k contains $k+1$ symbols and that of 10 contains 11 symbols, so it will be $8 + 11 + 1 + (k+1) + 2 + k + 1 = 2k + 24$ (do not miss out the $'$), which is less than $10k$ as $k > 3$.

By Theorem 19, there is a number that cannot be named by any formula containing less than $10k$ symbols, let it be n. So $A(n, 10k)$ is true.

$F(x)$ contains less than $10k$ symbols, hence n cannot be named by $F(x)$, and $\forall x\big(F(x) \leftrightarrow x = \overline{n}\big)$ cannot be provable (by the definition of naming).

However $F(\overline{n})$ is true as it says "n is the least number that cannot be named by any formula containing less than $10k$ symbols. Which means $\forall x\big(F(x) \leftrightarrow x = \overline{n}\big)$ is true. Then the negation of this sentence is false and cannot be provable by our soundness assumption.

Therefore $\forall x\big(F(x) \leftrightarrow x = \overline{n}\big)$ is undecidable.

6.2 Chaitin's proof

For Chaitin's proof we need to introduce a concept called *Kolmogorov complexity*. Let us first fixed a programming language, and consider the programs written by that programming language.

We further assume that numbers can be printed out on the screen in decimal expressions, and for every number there is a program that just runs and prints that number (only). For example, there may be a program that with code (translated to English) saying "print 50" and the output is "50", or "calculate the factorial of 5 and print the result" and the output is "120".

Then we can define Kolmogorov complexity (with respect to this language) of a number n, denoted by $K(n)$, as the length of the shortest program that the output is n. It is well defined for every n, by our assumption that there is a program with output n.

Similar to the case in Boolos's proof, for every number k there are finitely many numbers that is an output of a program of length less than k. Therefore for every k there is a number u_k such that $K(u_k) \geq k$.

Now suppose there is a consistent formal system like our **T** such that we can prove or disprove every sentence of the form $K(\overline{n}) \geq \overline{C}$, where n and C are numbers.[12]

There is a minor problem: how can we talk about programs (and length of them) in the formal system? But this does not bother us, we can just take a specific alphabet and a universal Turing machine of that alphabet. It is well-known that every Turing computable function is a recursive function. So for a program that just runs and prints the number n can be considered as a Turing machine that having output n (or some representation of n) when the input is empty, or a recursive function $f(x)$ such that $f(0) = n$. And we can take the length of the definition of a recursive function as the complexity of it.[13]

Without working out the details, let us just assume that we can more or less "talk about" Kolmogorov complexity in a formal system, and every sentence of the form $K(\overline{n}) \geq \overline{C}$ is decidable for every number n and C.

Fix a number C, then consider sentences of the form $K(\overline{n}) \geq \overline{C}$ which are provable. Note that if it is provable, then $K(n) \geq C$. This is because if $K(\overline{n}) \geq \overline{C}$ is provable but $K(n) < C$, then there is a program of length less than C that outputs n. It is a finite computation and $K(\overline{n}) < \overline{C}$ can be proved in our system, contradicting our assumption that the system is consistent.

Then let w be the first[14] proof of sentences of this form, and n_w be the corresponding number. That is, w a proof of the sentence $K(\overline{n_w}) \geq \overline{C}$.

And we can have a program that output n_w: enumerate all the possible proofs (in the order described in footnote 14), and for the first proof of the form $K(\overline{n}) \geq \overline{C}$, prints n and then stops.

The length of this program depends on the length of the expression of C.

[12]Caution: This is an abuse of notation, since "$K(\overline{n})$ is different from "$K(n)$", the former is a corresponding function in the system.

[13]In fact, 3-tape-symbol Turing machines were used to prove the theorem in Chaitin (1971).

[14]It is not difficult to define a well-ordering on the set of expressions of an countable (possibly finite) alphabet. For example, we can order them by lengths, and order expressions of the same length according to alphabetical order. Hence there must be a first proof of such sentences.

Let the length of the other parts of the program (without any occurrence of the expression of C) be p, and the expression of C occurs q times in the program. Then the length of this program is smaller than $p + q(\log C + 1)$, where $\log C$ is the logarithm of C, the number (not necessarily a natural number) such that $10^{\log C} = C$.

For a sufficiently large C_0, we have $C_0 > p + q(\log C_0 + 1)$. However such program outputs a number n_0, that means $K(n_0) < C_0$, and n_0 is supposed to have a Kolmogorov complexity at least C_0, from the above argument that if $K(\overline{n_0}) \geq \overline{C_0}$ is provable then $K(n_0) \geq C_0$. Thus we get a contradiction.

Therefore, $K(\overline{n_0}) \geq \overline{C_0}$ is not provable, and we get another undecidable sentence.

6.3 Formalized Boolos's proof

Boolos only gave a sketch of proof in his paper, more detailed and formalized proofs can be found in Kikuchi (1994) and Serény (2004). In the following we will present a proof based on Boolos's idea with the notations in this work.

Let us call a formula φ *standard* if all variables, free or bounded, in the formula are $v_0, v_1, \ldots, v_{k-1}$, where k is the number of variables in φ. That is, we use the smallest possible indices of variables in the formula.

We say a formula $\varphi(x)$ with one variable *describes* a number n if the sentence $\forall x\big(\varphi(x) \leftrightarrow x = \overline{n}\big)$ is provable.

Now, we have a few relations to define:

1. $Occ(x, y) \longleftrightarrow Form(x) \wedge Var(y) \wedge \big(\exists z \leq l(x)\big)\big(Dec(z, x) = y\big)$
   y represents a variable that occurs in the formula represented by x.

2. $Eqv(x, y) \longleftrightarrow Con\big(Imp(x, y), Imp(y, x)\big)$
   If $x = \ulcorner\varphi\urcorner$ and $y = \ulcorner\psi\urcorner$, then $Eqv(x, y) = \ulcorner\varphi \leftrightarrow \psi\urcorner$.

3. $StdFm(x) \longleftrightarrow Form(x) \wedge \big(\forall y \leq \overline{19}^{l(x)}\big)\big[Occ(x, y) \to (\forall z \leq y)\big(z \mid y \to Occ(z, y)\big)\big]$
   x represents a standard formula.

4. $Desc(x, n) \longleftrightarrow Prop(x) \wedge (\exists y \leq x)Free(y, x) \wedge$
   $Prov\Big[Gen\Big(y, Eqv\big(x, y \star R(\overline{21}) \star R(N(n))\big)\Big)\Big]$
   x represents a formula that describes n.

5. $sDesc(n) = \big(\mu x \leq R(\overline{19}) \star R(\overline{21}) \star R(N(n))\big)\big[Desc(x,n) \wedge StdFm(x) \wedge \big(\forall y \leq R(\overline{19}) \star R(\overline{21}) \star R(N(n))\big)\big(Desc(y,n) \rightarrow l(x) \leq l(y)\big)\big]$
   $sDesc(n)$ is the Gödel number of the shortest standard formula that describes n.[15]

6. $Undes(n) = \big(\mu x \leq [n \times (\overline{10} + n)^n]\big)\neg\big(l(sDesc(x)) \leq n\big)$
   $Undes(n)$ is the smallest number that cannot be described by any formula of length less than or equal to n.[16]

Now let k be the number of symbols of the formula $Undes(v_1)$. Note that for every number n, the length of $\overline{n}$ (the formal expression of the number n) is $n+1$. Then consider the following formula (under the formula there are some numbers showing how many symbols are used):

$$\underbrace{\exists v_1(v_1 =}_{5}\underbrace{\overline{10}}_{11}\underbrace{\times}_{1}\underbrace{\overline{k}}_{k+1}\underbrace{\wedge}_{1}\underbrace{Undes(v_1)}_{k}\underbrace{= v_0)}_{3} \tag{2}$$

Let us call this formula $B(v_0)$, which contains $5+11+1+k+1+1+k+3 = 2k+22$ symbols. $B(v_0)$ describes a number b, which is the least number cannot be described by any formula of length less than $10 \times k$. Obviously $k > 3$, therefore $2k + 22 \leq 10 \times k$, the length of $B(v_0)$ is less than $10 \times k$.

For convenience, in the following let $m = 10 \times k$. Before we construct an undecidable sentence, we need a few lemmas.

Lemma 20. *If $\boldsymbol{T}$ is ω-consistent and $\vdash Desc(\overline{p}, \overline{n_1}) \wedge Desc(\overline{p}, \overline{n_2})$ then $\vdash \overline{n_1} = \overline{n_2}$.*

[15]On the bound of x: $R(19) \star R(21) \star R(N(n))$ is the Gödel number of the formula $x = \underbrace{SS\ldots S}_{n}0$, which is, of course, a description of n.

[16]On the bound of n: we have 10 symbols and infintiely many variables, but there are at most first k variables in a standard formula of length k, so there are $(10+k)^k$ possible expressions of length k, and $11 + 12^2 + 13^3 + \ldots + (10+n)^n$ possible expressoins of length less than or equal to n, which is much less than $n \times (10+n)^n$. By the pegionhole principle, there must be a number less than or equal to $n \times (10+n)^n$ that cannot be described by any formula of length less than or equal to n.

Proof. Let $p = \ulcorner\varphi(v_0)\urcorner$, suppose $\vdash Desc(\ulcorner\varphi(v_0)\urcorner, \overline{n_1}) \wedge Desc(\ulcorner\varphi(v_0)\urcorner, \overline{n_2})$.

$$\begin{aligned}
&\vdash Desc(\ulcorner\varphi(v_0)\urcorner, \overline{n_1}) \\
\Longrightarrow &\vdash Prop(\ulcorner\varphi(v_0)\urcorner) \wedge (\exists\ulcorner v_0\urcorner \leq \ulcorner\varphi(v_0)\urcorner) Free(\ulcorner v_0\urcorner, \ulcorner\varphi(v_0)\urcorner) \\
&\quad \wedge Prov\big(\ulcorner\forall v_0(\varphi(v_0) \leftrightarrow v_0 = \overline{n_1})\urcorner\big) \\
\Longrightarrow &\vdash Prov\big(\ulcorner\forall v_0(\varphi(v_0) \leftrightarrow v_0 = \overline{n_1})\urcorner\big) \\
\Longrightarrow &\vdash \forall v_0(\varphi(v_0) \leftrightarrow v_0 = \overline{n_1})
\end{aligned}$$

Similarly, we have:

$$\begin{aligned}
&\vdash \forall v_0(\varphi(v_0) \leftrightarrow v_0 = \overline{n_2}) \\
\Longrightarrow &\vdash \forall v_0(v_0 = \overline{n_1} \leftrightarrow v_0 = \overline{n_2}) \\
\Longrightarrow &\vdash \overline{n_1} = \overline{n_1} \longleftrightarrow \overline{n_1} = \overline{n_2} \\
\Longrightarrow &\vdash \overline{n_1} = \overline{n_2}
\end{aligned}$$

□

Corollary 21. *If $\boldsymbol{T}$ is ω-consistent, $\vdash m \neq n$ and $\vdash Desc(p, m) \wedge Desc(q, m)$, then $\vdash p \neq q$.*

Lemma 22. *For every number n, there is a number $s_n > 0$ such that $\vdash sDesc(\overline{n}) = \overline{s_n}$.*

Proof. This is because the formula $v_0 = \overline{n}$ is a description of n, both its length and Gödel number are finite, hence there must be an s_n such that $\vdash sDesc(\overline{n}) = \overline{s_n}$. □

Theorem 23. *If $\boldsymbol{T}$ is ω-consistent, then the sentence $\exists v_0\big(Undes(\overline{m}) = v_0\big)$ is neither provable nor refutable.*

Proof. Suppose $\vdash \exists v_0 Undes(\overline{m}) = v_0$. By ω-consistency, there is a number u_m

such that $\vdash Undes(\overline{m}) = \overline{u_m}$

$$\vdash \exists v_1(v_1 = \overline{m} \wedge Undes(v_1) = \overline{u_m})$$

$$\Longrightarrow \vdash B(u_m)$$

$$\Longrightarrow \vdash \forall v_0(B(v_0) \leftrightarrow v_0 = \overline{u_m})$$

$$\Longrightarrow \vdash Prov\big(\ulcorner \forall v_0(B(v_0) \leftrightarrow v_0 = \overline{u_m})\urcorner\big)$$

$$\Longrightarrow \vdash Desc(\ulcorner B(v_0)\urcorner, \overline{u_m})$$

However, since $\vdash l(\ulcorner B(v_0)\urcorner) \leq \overline{m}$,

we have $\vdash \neg Undes(\overline{m}) = \overline{u_m}$ which leads to a contradiction.

On the other hand, suppose $\vdash \neg\exists v_0 Undes(\overline{m}) = v_0$, which is equivalent to:

$$\vdash \forall v_0 \neg Undes(\overline{m}) = v_0$$

$$\Longrightarrow \vdash \neg Undes(\overline{m}) = 0$$

By the definition of $Undes(\overline{m})$, we have:

$$\vdash (\exists x \leq \overline{(m \times (10+m)^m})\neg(l(sDesc(x) \leq \overline{m})$$

$$\Longrightarrow \vdash \exists x(x = 0 \vee x = 1 \vee \ldots \vee x = \overline{m \times (10+m)^m}) \wedge \neg(l(sDesc(x) \leq \overline{m})$$

Then there is a least number $a_0 \leq m \times (10+m)^m$ such that

$$\vdash \neg l(sDesc(\overline{a_0}) \leq \overline{m})$$

But then $\vdash Undes(\overline{m}) = \overline{a_0}$, contradicting

$$\vdash \forall v_0 \neg Undes(\overline{m}) = v_0$$

Therefore $\exists v_0\big(Undes(\overline{m}) = v_0\big)$ is neither provable nor refutable. $\square$

6.4 Relation between Boolos's and Chaitin's proof

Using the functions and relations we have defined so far, it is straightforward to define a function $K(x)$ as an analogue of Kolmogorov complexity in Chaitin's proof: $K(x) = l(sDesc(x))$, which is the length of the shortest description of x.

However, Chaitin's idea does not work here, that is, we cannot find a number C_0 such that the length of the formula $K(x) \geq \overline{C_0}$ is less than C_0. This is because in **T** n is represented by an expression of length $n+1$, while in Chaitin's proof it is assumed that numbers are represented in positional systems with base larger than 1. The latter assumption ensures the existence of C_0 such that the program required in the proof, which contains a string representing C_0, is of length less than C_0.

This is where Boolos's tricks works, that is, instead of directly putting a long numeral into the formula, we can break it down by multiplication. In the previous proof, we use the formula $\exists v_1(v_1 = \overline{10} \times \overline{k} \wedge Undes(v_1) = v_0)$, where k is the length of the formula $Undes(v_1)$, which is provably equivalent to the formula $Undes(\overline{10 \times k}) = v_0$.

Similarly we can let h be the length of the full expression of the formula $K(v_0) > v_1$, and $C(v_0)$ be the formula $\exists v_1(v_1 = 10 \times \overline{h} \wedge K(v_0) > v_1)$. The length of $C(v_0)$ is $2h + 20$, which is less than $10h$ since $h > 3$. And from the definition of $Undesc(x)$ it is not difficult to see that $C(v_0)$ is actually $B(v_0)$.

7 Yablo's Paradox

Stephen Yablo propose a relatively young paradox in (Yablo, 1993), which can be formulated as the following.

Consider the following infinite list of sentences:

$$\begin{aligned}
S_0 &: \text{For all } k > 0,\ S_k \text{ is not true.} \\
S_1 &: \text{For all } k > 1,\ S_k \text{ is not true.} \\
S_2 &: \text{For all } k > 2,\ S_k \text{ is not true.} \\
&\vdots \qquad\qquad \vdots
\end{aligned}$$

That is, for any i, S_i is the sentence "For all $k > i$, S_k is not true."

Is S_0 true? Suppose it is, then S_1 is not true. Since S_1 is not true, there is some $n > 1$ such that S_n is true. But by our assumption that S_0 is true, S_n is not true as $n > 0$.

On the other hand, suppose S_0 is not true, then there is some $n > 0$ such that S_n is true. Obviously $n + 1 > n$, so S_{n+1} is not true, and there is some $m > n+1$ such that S_m is true. But then $m > n$ and S_m is true, contradicting S_n.

In fact, by substituting the number 0 by any other number (including the subscripts) in the above argument, every sentence on the list is paradoxical, as we can start the sequence from any point.

7.1 Formalization using $Prov(x)$

In a footnote of Priest (1997), Priest said "one can turn Yablo's argument into a proof of Gödel's first incompleteness theorem", this idea is developed in Cook (2006) without going to the discussion about undecidable sentences, and in Cieślinski and Urbaniak (2013) of which this section is based on.

Before formalizing the paradox, we need the following version of Lemma 1:

Lemma 24 (Diagonal Lemma)**.** *Let $\varphi(x, y)$ be a formula with exactly two free variables x and y. Then there is a formula $\psi(x)$ with exactly one free variable x such that $\psi(x) \longleftrightarrow \varphi(x, \ulcorner\psi(x)\urcorner)$ is provable.*

Proof. Let $\delta(x, y)$ be the formula $\varphi\big(x, Subs(y, \ulcorner y\urcorner, y)\big)$, $d = \ulcorner\delta(x, y)\urcorner$, $\psi(x)$ be the formula $\delta(x, d)$. Then:

$$
\begin{aligned}
\vdash \psi(x) &\longleftrightarrow \delta(x, d) \\
&\longleftrightarrow \varphi\big(x, Subs(d, \ulcorner y\urcorner, d)\big) \\
&\longleftrightarrow \varphi\big(x, \ulcorner\delta(x, d)\urcorner\big) \\
&\longleftrightarrow \varphi\big(x, \ulcorner\psi(x)\urcorner\big)
\end{aligned}
$$

□

Then consider the following open formula:

$$\forall z\big(z > x \to \neg Prov(Sub(y, \ulcorner z\urcorner))\big)$$

By the above lemma, there is a formula $\mathbf{Y}(x)$ such that

$$\vdash \mathbf{Y}(x) \longleftrightarrow \forall z\big(z > x \to \neg Prov(\ulcorner\mathbf{Y}(z)\urcorner)\big)$$

Then we have the following result:

Theorem 25. *If $\boldsymbol{T}$ is ω-consistent, then for any natural number k, the sentence $\mathbf{Y}(\overline{k})$ is neither provable nor refutable.*

Proof. Let k be a natural number. Suppose $\mathbf{Y}(\overline{k})$ is provable. Then:

$$
\begin{aligned}
\vdash \mathbf{Y}(\overline{k}) &\Rightarrow \vdash \forall z\big(z > \overline{k} \to \neg Prov(\ulcorner \mathbf{Y}(z) \urcorner)\big) && \text{(By definition)}\\
&\Rightarrow \vdash \forall z\big(z > \overline{k+1} \to \neg Prov(\ulcorner \mathbf{Y}(z) \urcorner)\big) && \text{(Since } \forall z(z > \overline{k+1} \to z > \overline{k}))\\
&\Rightarrow \vdash \mathbf{Y}(\overline{k+1}) && \text{(By definition)}\\
&\Rightarrow \vdash Prov(\ulcorner \mathbf{Y}(\overline{k+1}) \urcorner) && \text{(By Lemma 2)}
\end{aligned}
$$

But since $\vdash \overline{k+1} > \overline{k}$, by $\forall z\big(z > \overline{k} \to \neg Prov(\ulcorner \mathbf{Y}(z) \urcorner)$, we have $\vdash \neg Prov(\ulcorner \mathbf{Y}(\overline{k+1}) \urcorner)$. By consistency of $\mathbf{T}$, this is impossible. Hence for any natural number k, $\mathbf{Y}(\overline{k})$ is not provable.

On the other hand, suppose $\mathbf{Y}(\overline{k})$ is refutable. Then:

$$
\begin{aligned}
\vdash \neg\mathbf{Y}(\overline{k}) &\Rightarrow \vdash \neg\forall z\big(z > \overline{k} \to \neg Prov(\ulcorner \mathbf{Y}(z) \urcorner)\big) && \text{(By definition)}\\
&\Rightarrow \vdash \exists z\big(z > \overline{k} \wedge Prov(\ulcorner \mathbf{Y}(z) \urcorner)\big)
\end{aligned}
$$

Since $\mathbf{T}$ is ω-consistent, there is a number n such that:

$$
\begin{aligned}
&\vdash \overline{n} > \overline{k} \wedge Prov(\ulcorner \mathbf{Y}(\overline{n}) \urcorner)\\
\Rightarrow &\vdash Prov(\ulcorner \mathbf{Y}(\overline{n}) \urcorner)\\
\Rightarrow &\vdash \mathbf{Y}(\overline{n}) && \text{(By Lemma 3)}
\end{aligned}
$$

But by the first half of this proof, $\mathbf{Y}(\overline{n})$ cannot be provable, so there is a contradiction. Therefore, for any number k, $\mathbf{Y}(\overline{k})$ is neither provable nor refutable. □

7.2 Existential Yablo's paradox

There is a dual form of the original Yablo's paradox, using existential quantifiers instead of universal quantifiers. Consider the following infinite list of sentences:

T_0 There is at least one $k > 0$ such that T_k is not true.

T_1 There is at least one $k > 1$ such that T_k is not true.

T_2 There is at least one $k > 2$ such that T_k is not true.

$\vdots$ $\qquad\qquad\vdots$

Let n be an arbitrary number. Suppose T_n is not true. Then for all $k > n$, T_k is true. But then T_{n+1} is true, that means there is an $m > n+1$ such that

T_m is not true, contradicting that T_k is true for every k larger than n. Hence it is impossible for T_n not to be true, no matter what n is.

On the other hand, suppose T_n is true. Then there is some $m > n$ such that T_m is not true. But by the above paragraph if T_m is not true then there is a contradiction. Therefore T_n is paradoxical for any n.

Now consider the following open formula:

$$\exists z\big(z > x \wedge \neg Prov(Sub(y, \ulcorner z \urcorner))\big)$$

Then by Lemma 24, there is a formula $\mathbf{Z}(x)$ such that:

$$\vdash \mathbf{Z}(x) \longleftrightarrow \exists z\big(z > x \wedge \neg Prov(\ulcorner \mathbf{Z}(z) \urcorner)\big)$$

And we will have another queue of undecidable sentences.

Theorem 26. *If* $\boldsymbol{T}$ *is* ω*-consistent, then for any natural number* k*, the sentence* $\mathbf{Z}(\overline{k})$ *is neither provable nor refutable.*

Proof. Let k be a natural number. Suppose $\mathbf{Z}(\overline{k})$ is refutable.

$$\begin{aligned} \vdash \neg\mathbf{Z}(\overline{k}) &\Rightarrow \vdash \neg\exists z\big(z > \overline{k} \wedge \neg Prov(\ulcorner \mathbf{Z}(z) \urcorner)\big) && \text{(By definition)} \\ &\Rightarrow \vdash \forall z\big(z > \overline{k} \rightarrow Prov(\ulcorner \mathbf{Z}(z) \urcorner)\big) \\ &\Rightarrow \vdash \forall z\big(z > \overline{k+1} \rightarrow Prov(\ulcorner \mathbf{Z}(z) \urcorner)\big) \end{aligned}$$

But by universal instantiation, the fact that $\vdash \overline{k+1} > \overline{k}$, and modus ponens, we get

$$\begin{aligned} &\vdash Prov(\mathbf{Z}(\ulcorner \overline{k+1)} \urcorner) \\ \Rightarrow &\vdash \mathbf{Z}(\overline{k+1)} && \text{(By Lemma 3)} \\ \Rightarrow &\vdash \exists z\big(z > \overline{k+1} \wedge \neg Prov(\ulcorner \mathbf{Z}(z) \urcorner)\big) \end{aligned}$$

Hence there is a contradiction.

On the other hand, suppose $\mathbf{Z}(\overline{k})$ is refutable. Then by the choice of $\mathbf{Z}(x)$, $\exists z\big(z > \overline{k} \wedge \neg Prov(\ulcorner \mathbf{Z}(z) \urcorner)\big)$ is provable. By existential elimination, there is an a such that $\neg Prov(\ulcorner \mathbf{Z}(a) \urcorner)\big)$ is provable. But by Lemma 11 this implies that $\mathbf{T}$ is inconsistent.

Therefore if $\mathbf{T}$ is ω-consistent, then for any natural number k, the sentence $\mathbf{Z}(\overline{k})$ is neither provable nor refutable. $\square$

8 Variants of the Preface Paradox

The preface paradox, first introduced by Makinson (1965), is an epistemic paradox. It is about a usual practice that an author claiming, in the preface of the book, that there will inevitably some errors, mistakes in that book. It seems that an author can rationally believes all the claims in his or her book, at the same time believes that he or she is fallible. But then this author is believing something contradictory, namely, that all the claims in the book are correct, and some of them are incorrect.

A variant of this paradox is related to the liar paradox: Consider a book that the author wrote "there are inevitably some (at least one) mistakes" in the preface. Though usually quite improbable, the rest of the book is correct. Then the problem is, obviously, the sentence "there are inevitably some (at least one) mistakes" is correct if and only if it is not. This leads to a liar-like situation, depending on a contingent fact that the rest of the book is correct.

This section is about three variants of this paradox.

8.1 Variant 1: Someone is wrong

To proceed further, let us change the metaphor. Suppose there are some people in a room, everyone writes a sentence on a paper and there are no other things in that room. If one of them writes that "At least one of the sentences written in this room is not true", while the others are writing, say, "$1 + 1 = 2$", then it becomes the same situation as the previous paradox.

Now what if everyone writes "At least one of the sentences written in this room is not true"? If there is only one people, it becomes a version of liar paradox, as the only sentence written is referring to itself and saying that it is not true.

If there are more than one people, say n people, and let one of these sentence be true, then there is another sentence that is not true. By the falsity[17] of the latter sentence, all sentences written in that room are true, which is a plain contradiction to the very same sentence. Thus we can see that every sentence in such a situation is true if and only if it is not.

[17] Let us assume the principle of bivalence here.

It may sound like a variant of circular liars, but unlike circular liars, this paradox can be extended to the infinite case, which resembles the Yablo's paradox [18]. So if there are infinitely many people in the room (assuming that it is possible) and every one of them is writing that sentence, there will be a paradox again.

Again we can formalize this paradox to obtain some undecidable sentences, for both finite and infinite case.

8.2 Formalization of variant 1

Finite case

First we need another more Generalized Diagonal Lemma:

Lemma 27 ((Really) Generalized Diagonal Lemma)**.** *Let $x_0, x_1, \ldots, x_n, y_1, \ldots, y_m$ be distinct variables, $\varphi_0(x_0, x_1, \ldots, x_n, \mathbf{y}), \varphi_1(x_0, x_1, \ldots, x_n, \mathbf{y}), \ldots, \varphi_n(x_0, x_1, \ldots, x_n, \mathbf{y})$ be formulas in which all free variables are among $x_0, x_1, \ldots, x_n, \mathbf{y}$ (where $\mathbf{y}$ is the abbreviation of $y_1, \ldots, y_m$). Then there exists formulas $\psi_0(\mathbf{y}), \psi_1(\mathbf{y}), \ldots, \psi_n(\mathbf{y})$ with free variables $\mathbf{y}$ such that:*

$$
\begin{aligned}
&\vdash \psi_0(\mathbf{y}) \longleftrightarrow \varphi_0(\ulcorner \psi_0(\mathbf{y}) \urcorner, \ulcorner \psi_1(\mathbf{y}) \urcorner, \ldots, \ulcorner \psi_n(\mathbf{y}) \urcorner, \mathbf{y}) \\
&\vdash \psi_1(\mathbf{y}) \longleftrightarrow \varphi_1(\ulcorner \psi_0(\mathbf{y}) \urcorner, \ulcorner \psi_1(\mathbf{y}) \urcorner, \ldots, \ulcorner \psi_n(\mathbf{y}) \urcorner, \mathbf{y}) \\
&\qquad\vdots \qquad\qquad\qquad\qquad \vdots \\
&\vdash \psi_n(\mathbf{y}) \longleftrightarrow \varphi_n(\ulcorner \psi_0(\mathbf{y}) \urcorner, \ulcorner \psi_1(\mathbf{y}) \urcorner, \ldots, \ulcorner \psi_n(\mathbf{y}) \urcorner, \mathbf{y})
\end{aligned}
$$

A proof can be found in (Boolos, 1993). The basic idea of the proof is again using substitution, but a more complicated one. So we skip the proof.

In the following, we write $Con_n(x_0, x_1, \ldots, x_n)$ to abbreviate the expression $\underbrace{Con(\ldots Con(Con(}_{n\text{-many 'Con('s}} x_0, x_1), x_2), \ldots), x_n)$.

Let n be a fixed natural number. For any $0 \leq k \leq n$, take $\varphi_k(x_0, x_1, \ldots, x_n)$ be the formula $\neg Prov(Con_n(x_0, x_1, \ldots, x_n))$. By the above Generalized Diagonal Lemma, there are sentences $\mathbf{A_0}, \mathbf{A_1}, \ldots, \mathbf{A_n}$ such that for any $0 \leq k \leq n$, the sentence $\mathbf{A_k} \longleftrightarrow \neg Prov(Con_n(\ulcorner \mathbf{A_0} \urcorner, \ulcorner \mathbf{A_1} \urcorner, \ldots, \ulcorner \mathbf{A_n} \urcorner))$ is provable. It is

[18]But they are not the same.

not difficult to see that such a sentence is provably equivalent to the sentence $\mathbf{A_k} \longleftrightarrow \neg Prov(\ulcorner \mathbf{A_0} \wedge \mathbf{A_1} \wedge \ldots \wedge \mathbf{A_n} \urcorner)$.

Then we have the following result:

Theorem 28. *If* $\boldsymbol{T}$ *is* ω*-consistent, then for any* $0 \leq k \leq n$, $\mathbf{A_k}$ *is neither provable nor refutable.*

Proof. Suppose $\mathbf{T}$ is ω-consistent. Let $0 \leq k \leq n$. Suppose $\mathbf{A_k}$ is provable. Then $\neg Prov(\ulcorner \mathbf{A_0} \wedge \mathbf{A_1} \wedge \ldots \wedge \mathbf{A_n} \urcorner)$ is also provable. By Lemma 11, it is impossible.

Suppose $\mathbf{A_k}$ is refutable. Then:

$$\begin{aligned}
\vdash \neg \mathbf{A_k} &\Rightarrow \vdash Prov(\ulcorner \mathbf{A_0} \wedge \mathbf{A_1} \wedge \ldots \wedge \mathbf{A_n} \urcorner) && \text{(By the choice of } \mathbf{A_k}\text{)} \\
&\Rightarrow \vdash \mathbf{A_0} \wedge \mathbf{A_1} \wedge \ldots \wedge \mathbf{A_n} && \text{(By Lemma 3)} \\
&\Rightarrow \vdash \mathbf{A_k} && \text{(By conjunction elimination)}
\end{aligned}$$

And we get a contradiction, as by our assumption $\mathbf{T}$ is consistent.

Therefore if $\mathbf{T}$ is ω-consistent, then all of the sentences $\mathbf{A_0}, \mathbf{A_1}, \ldots, \mathbf{A_n}$ are neither provable nor refutable. $\square$

Using this Generalized Diagonal Lemma, we can also formalize the circular liars and get another family of undecidable sentences: Fixed a positive number n, for $k < n$, let $\varphi_k(x_0, x_1, \ldots, x_n)$ be $\neg Prov(x_{k+1})$, and let $\varphi_n(x_0, x_1, \ldots, x_n)$ be $Prov(x_0)$. The fixed points of these open formulas will have a circular liar like structure, and it is easy to see that if $\mathbf{T}$ is ω-consistent, then none of them is decidable.

Infinite case

For the infinite case, we can use a similar trick as in the formalization of Yablo's paradox. Consider the open formula:

$$\exists z \big(\neg Prov(Sub(x, \ulcorner z \urcorner))\big) \wedge (0 \leq y)$$

Then by Lemma 24 there is an open formula $\mathbf{P}(y)$ with one free variable y

such that

$$\begin{aligned}
\vdash \mathbf{P}(y) &\longleftrightarrow \exists z\big(\neg Prov(Sub(\ulcorner \mathbf{P}(y)\urcorner, \ulcorner z\urcorner))\big) \wedge (0 \leq y) \\
&\longleftrightarrow \exists z\big(\neg Prov(\ulcorner \mathbf{P}(z)\urcorner)\big) \wedge (0 \leq y) \\
&\longleftrightarrow \exists z\big(\neg Prov(\ulcorner \mathbf{P}(z)\urcorner)\big)
\end{aligned}$$

We have the following result:

Theorem 29. *If $\boldsymbol{T}$ is ω-consistent, then for any natural number k, $\mathbf{P}(\overline{k})$ is neither provable nor refutable.*

Proof. Suppose $\mathbf{T}$ is ω-consistent, let k be a natural number. Suppose $\mathbf{P}(\overline{k})$ is provable. Then by the choice of $\mathbf{P}(y)$, $\exists z\big(\neg Prov(\ulcorner \mathbf{P}(z)\urcorner)\big)$ is also provable. By existential instantiation, there is an a such that $\neg Prov(\ulcorner \mathbf{P}(a)\urcorner)$ is provable. But by Lemma 11 this is impossible.

On the other hand, suppose $\mathbf{P}(\overline{k})$ is refutable. Then:

$$\begin{aligned}
\vdash \neg\mathbf{P}(\overline{k}) &\Rightarrow \vdash \neg\exists z\big(\neg Prov(\ulcorner \mathbf{P}(z)\urcorner)\big) && \text{(By the choice of } \mathbf{P}(y)) \\
&\Rightarrow \vdash \forall z Prov(\ulcorner \mathbf{P}(z)\urcorner) \\
&\Rightarrow \vdash Prov(\ulcorner \mathbf{P}(\overline{k})\urcorner) && \text{(By universal instantiation)} \\
&\Rightarrow \vdash \mathbf{P}(\overline{k}) && \text{(By Lemma 3)}
\end{aligned}$$

By our assumption on the consistency of $\mathbf{T}$, this is impossible.

Therefore if $\mathbf{T}$ is ω-consistent, then for any natural number k, $\mathbf{P}(\overline{k})$ is neither provable nor refutable. □

8.3 Variant 2: Someone else is wrong

Consider a situation similar to the previous paradox, except that everyone writes that "At least one of the sentences other then this one written in this room is not true". In other words, they are say that "someone else is wrong".[19]

[19]A note after finishing this section: Professor András Máté points out that this variant resembles a paradox by Jean Buridan, which says that "Socrates says that Plato tells a lie, Plato says that Socrates tells a lie." This paradox is also called the No-No paradox in Sorensen (2004), which further contains a finite case of the paradox in this subsection.

When there is only one person in the room, the only sentence is false since there is no other sentences. However, if there are more than one people, the situation becomes paradoxical, not in the sense that there will be sentences that are true if and only if they are false, but the truth values of them seems to be arbitrary. This can be illustrated by the four people case (any number larger than one will do the job)[20].

Let S_1, S_2, S_3, S_4 be four sentences, each of them saying that some of the other sentences is not true. To be precise, let S_1 be "It is not the case that S_2, S_3, S_4 are all true", S_2 be "It is not the case that S_1, S_3, S_4 are all true", S_3 and S_4 can be defined similarly.

Suppose S_1 is true, then at least one of S_2, S_3, S_4 is not true. Let S_2 be false[21], then S_1, S_3, S_4 are all true, and there is no contradiction. It is easy to check that there is no contradiction if and only if exactly one sentence is not true, but it does not matter which one is the false sentence.

In other words, we can assign the truth values of the sentence arbitrarily, under the condition that exactly one sentence is not true. Hence it is like the truth-teller, i.e. the sentence "This sentence is true", that the truth value assignment is kind of bootstrapping.

By Löb's theorem, the Henkin sentence, which is a formalized version of the truth teller, is provable. Nevertheless, if we formalized the above paradox by using the provability predicate, the sentences obtained are undecidable.

8.4 Formalization of variant 2

First we formalize the finite case. Let n be a natural number, $P_0(x_0, x_1, \ldots, x_n)$ be the formula $\neg Prov\big(Con_{n-1}(x_1, x_2, \ldots, x_n)\big)$. For every $0 < k < n$, let $P_k(x_0, x_1, \ldots, x_n)$ be the formula $\neg Prov\big(Con_{n-1}(x_1, x_2, \ldots, x_{k-1}, x_{k+1}, \ldots, x_n)\big)$. Finally let $P_n(x_0, x_1, \ldots, x_n)$ be the formula $\neg Prov\big(Con_{n-1}(x_0, x_1, \ldots, x_{n-1})\big)$.

By the Generalized Diagonal Lemma there are sentences $\mathbf{B_0}, \mathbf{B_1}, \ldots, \mathbf{B_n}$

[20]The same holds for the infinite case.

[21]Again we assume the principle of bivalence here.

such that

$$\begin{aligned}
&\vdash \mathbf{B_0} \longleftrightarrow \neg Prov(\ulcorner \mathbf{B_1} \wedge \mathbf{B_2} \wedge \ldots \wedge \mathbf{B_n} \urcorner) \\
&\vdash \mathbf{B_1} \longleftrightarrow \neg Prov(\ulcorner \mathbf{B_0} \wedge \mathbf{B_2} \wedge \ldots \wedge \mathbf{B_n} \urcorner) \\
&\quad \vdots \qquad\qquad\qquad\quad \vdots \\
&\vdash \mathbf{B_n} \longleftrightarrow \neg Prov(\ulcorner \mathbf{B_0} \wedge \mathbf{B_1} \wedge \ldots \wedge \mathbf{B_{n-1}} \urcorner)
\end{aligned}$$

Theorem 30. *If* $\boldsymbol{T}$ *is* ω*-consistent, then for any* $0 \leq k \leq n$*,* $\mathbf{B_k}$ *is neither provable nor refutable.*

Proof. Suppose $\mathbf{T}$ is ω-consistent. Let $0 \leq k \leq n$. Suppose $\mathbf{B_k}$ is provable. Then by the choice of $\mathbf{B_k}$, the sentence $\neg Prov(\ulcorner \mathbf{B_0} \wedge \mathbf{B_1} \wedge \ldots \wedge \mathbf{B_{k-1}} \wedge \mathbf{B_{k+1}} \wedge \ldots \wedge \mathbf{B_n} \urcorner)$ is also provable. But this is impossible by Lemma 11.

On the other hand, suppose $\mathbf{B_k}$ is refutable. Then:

$$\begin{aligned}
&\vdash \neg \mathbf{B_k} \\
\Rightarrow &\vdash Prov(\ulcorner \mathbf{B_0} \wedge \mathbf{B_1} \wedge \ldots \wedge \mathbf{B_{k-1}} \wedge \mathbf{B_{k+1}} \wedge \ldots \wedge \mathbf{B_n} \urcorner) && \text{(By the choice of } \mathbf{B_k}\text{)} \\
\Rightarrow &\vdash \mathbf{B_0} \wedge \mathbf{B_1} \wedge \ldots \wedge \mathbf{B_{k-1}} \wedge \mathbf{B_{k+1}} \wedge \ldots \wedge \mathbf{B_n} && \text{(By Lemma 3)} \\
\Rightarrow &\vdash \mathbf{B_0} && \text{(By conjunction elimination)}
\end{aligned}$$

However, by the first part of the proof, we know that it is impossible. Therefore $\mathbf{B_k}$ is neither provable nor refutable. □

For the infinite case, consider the following open formula:

$$\exists z \big(z \neq y \wedge \neg Prov(Sub(x, \ulcorner z \urcorner)) \big)$$

By the Generalized Diagonal Lemma, there is an open formula $\mathbf{Q}(y)$ with one free variable y such that $\mathbf{Q}(y) \longleftrightarrow \exists z \big(z \neq y \wedge \neg Prov(\ulcorner \mathbf{Q}(z) \urcorner) \big)$ is provable.

Theorem 31. *If* $\boldsymbol{T}$ *is* ω*-consistent, then for any natural number* k*,* $\mathbf{Q}(\overline{k})$ *is neither provable nor refutable.*

Proof. Suppose $\mathbf{T}$ is ω-consistent. Let k be a natural number. Suppose $\mathbf{Q}(\overline{k})$ is provable, then by the choice of $\mathbf{Q}(y)$, $\exists z \big(z \neq \overline{k} \wedge \neg Prov(\ulcorner \mathbf{Q}(z) \urcorner) \big)$ is also provable. By existential instantiation, there is an a such that $\neg Prov(\ulcorner \mathbf{Q}(a) \urcorner)$ is provable. But by Lemma 11 this is impossible.

On the other hand, suppose $\mathbf{Q}(\overline{k})$ is refutable. Then:

$$\begin{aligned}
\vdash \neg\mathbf{Q}(\overline{k}) &\Rightarrow \vdash \neg\exists z\big(z \neq \overline{k} \wedge \neg Prov(\ulcorner\mathbf{Q}(z)\urcorner)\big) && \text{(By the choice of } \mathbf{Q}(y)\text{)} \\
&\Rightarrow \vdash \forall z\big(z \neq \overline{k} \rightarrow Prov(\ulcorner\mathbf{Q}(z)\urcorner)\big) \\
&\Rightarrow \vdash Prov(\ulcorner\mathbf{Q}(\overline{k+1})\urcorner) && \text{(By Universal Instantiation, } \vdash \overline{k+1} \neq \overline{k}\text{)} \\
&\Rightarrow \vdash \mathbf{Q}(\overline{k+1}) && \text{(By Lemma 3)}
\end{aligned}$$

But by the first half of this proof we know that it is impossible. Therefore, for any natural number k, $\mathbf{Q}(\overline{k})$ is neither provable nor refutable. □

8.5 Variant 3: At least k people are wrong

Consider another situation with similar setting, where there are n people and they queue up. The k^{th} person writes that "There are at least k sentences which are not true"[22], in other words, he or she is saying that "at least k people are wrong"[23].

If there is only one person in the room, then it is again the liar paradox. However, if there are more than one people in the room, unlike the previous situations, it is not entirely paradoxical (but still a little bit paradoxical, in some cases).

Obviously either all of them are right, or someone is wrong, in both cases the first person is right, hence someone is wrong, in particular the last person is wrong. Furthermore, if someone is right, everyone before him or her is also right; conversely if someone is wrong, everyone after him or her is also wrong. Since someone is wrong, there must be someone who is the first person who is wrong. Let that person be the k_0^{th} person, then at most $k_0 - 1$ people are wrong. At the same time, the $(k_0 - 1)^{st}$ person is right by definition of k_0, so at least $k_0 - 1$ people are wrong. Combining these two results, we know that there are exactly $k_0 - 1$ people are wrong.

But everyone after the k_0^{th} person is wrong, by a simple calculation[24] we can conclude that $k_0 - 1$ is a half of n. Therefore, if n is even, then people in the

[22] Except for the first one, who should write "There is at least 1 sentence which is not true", for grammatical reason.

[23] For convenience, we may say "someone is right" instead of saying that the corresponding sentence is true.

[24] $n - k_0 + 1 = k_0 - 1 \Rightarrow n = 2(k_0 - 1)$

first half of the queue are right, and those in the second half are wrong. There is nothing paradoxical at all.

However if n is odd, we can deduce that the first person is right and the last one is wrong, similarly the second one is right and the second-last person is wrong, and so on. Then the person who stands in the exact middle of the queue is in a liar paradox like situation, while everyone before him or her is right and everyone after is wrong.

Things get more interesting and paradoxical when, as always, infinity comes into the picture. If there are infinitely many people in a queue, the k^{th} is saying that "At least k people are wrong", then saying any one of them is right or wrong will lead to contradiction. For, suppose the k^{th} person is wrong, then there are at most $k-1$ people are wrong. But everyone after the k^{th} person is wrong, and there are more than k people no matter how large k is, so there is a contradiction. On the other hand, if the k^{th} person is right, then someone is wrong, but we know that it leads to a contradiction.

8.6 Formalization of variant 3

As in the finite case the situations are not really new, we will concentrate on formalizing the infinite case, which is slightly more complicated than previous variants.

At the first sight, it seems that we can get a fixed point of a formula using the usual $\exists_k$ quantifier. However this is an abbreviation at the meta-level, to do things in the object level, we need to refer to a set of numbers indirectly. The idea is to refer to a sequence with length k where no two terms are the same, and then we can refer to those terms in the formula we need.

Let us introduce two definitions:

- $HetSeq(x) \longleftrightarrow Code(x) \wedge \big(\forall y \leq l(x)\big)\big(\forall z \leq l(x)\big)\big(y \neq z \rightarrow Dec(y,x) \neq Dec(z,x)\big)$

  If $HetSeq(x)$ is provable, then x is a code number of a sequence where no two terms are the same.

- $Ele(x,y) \longleftrightarrow Code(y) \wedge \big(\exists u \leq l(y)\big)\big(Dec(u,y) = x\big)$

  If $Ele(x,y)$ is provable, then x represents a number which is a term of the sequence represented by y.

- $Trun(1,x) = \mu y\big[(y \le x) \wedge Pr\big(l(x)\big)^{Dec\left(l(x),x\right)} \times y = x\big]$
  $Trun(n+1,x) = Trun\big(Trun(n,x)\big)$
  If x is a code number, then $Trun(n,x)$ is the code number of the sequence obtained by deleting the last n terms.

We will use the following facts.

Let c be a code number with length n, and $k < n$. Then:

1. If $HetSeq(\overline{c})$ is provable, then $HetSeq(Trun(\overline{k},\overline{c}))$ is also provable.
2. $\forall x\big[Ele(x, Trun(\overline{k},\overline{c})) \to Ele(x,\overline{c})\big]$ is provable.
3. $Ele\big(Dec(\overline{k},\overline{c}),\overline{c}\big)$ is provable.

Proofs of the above two facts will be skipped.

Then consider the open formula

$$\exists z\Big[HetSeq(z) \wedge l(z) = Sy \wedge (\forall t \le z)\big[Ele(t,z) \to Prov\big(Neg(Sub(x, \ulcorner t \urcorner))\big)\big]\Big]$$

By the Generalized Diagonal Lemma, there is an open formula $\mathbf{R}(y)$ such that

$$\vdash \mathbf{R}(y) \longleftrightarrow \exists z\Big(HetSeq(z) \wedge l(z) = Sy \wedge (\forall t \le z)\big(Ele(t,z) \to Prov(\ulcorner \neg \mathbf{R}(t) \urcorner)\big)\Big)$$

We have the following two lemmas:

Lemma 32. *If m, n are natural numbers and $m < n$, then* $\vdash \mathbf{R}(\overline{n}) \to \mathbf{R}(\overline{m})$.

Proof. Let $m + d = n$. Suppose $\vdash \mathbf{R}(\overline{n})$. Then by the choice of $\mathbf{R}(y)$:

$\vdash \exists z\Big(HetSeq(z) \wedge l(z) = S\overline{n} \wedge (\forall t \le z)\big(Ele(t,z) \to Prov(\ulcorner \neg \mathbf{R}(t) \urcorner)\big)\Big)$

Then by existential instantiation, there is an a such that

$\vdash \Big(HetSeq(a) \wedge l(a) = S\overline{n} \wedge (\forall t \le a)\big(Ele(t,a) \to Prov(\ulcorner \neg \mathbf{R}(t) \urcorner)\big)\Big)$

Let $b = Trun(d,a)$. By the two facts above, it is not difficult to see that

$\vdash \Big(HetSeq(b) \wedge l(b) = S\overline{m} \wedge (\forall t \le b)\big(Ele(t,b) \to Prov(\ulcorner \neg \mathbf{R}(t) \urcorner)\big)\Big)$

Therefore $\vdash \mathbf{R}(\overline{m})$. By the deduction theorem we get $\vdash \mathbf{R}(\overline{n}) \to \mathbf{R}(\overline{m})$. □

Lemma 33. *If m, n are natural numbers and $m > n$, then* $\vdash \neg\mathbf{R}(\overline{n}) \to \neg\mathbf{R}(\overline{m})$.

Proof. By Lemma 32 we have $\vdash \mathbf{R}(\overline{m}) \to \mathbf{R}(\overline{n})$, which implies the contrapositive of the formula, therefore $\vdash \neg\mathbf{R}(\overline{n}) \to \neg\mathbf{R}(\overline{m})$. □

Now we can prove the following result:

Theorem 34. *If **T** is ω-consistent, then for any natural number n, $\mathbf{R}(\overline{n})$ is neither provable nor refutable.*

Proof. Suppose **T** is ω-consistent. Let n be a natural number.

Suppose $\mathbf{R}(\overline{n})$ is refutable, then $\vdash \neg\mathbf{R}(\overline{n})$.

By Lemma 33, for every $m > n$, $\mathbf{R}(\overline{m})$ is refutable. Therefore the sentences $\neg\mathbf{R}(\overline{n}), \neg\mathbf{R}(\overline{n+1}), \ldots, \neg\mathbf{R}(\overline{n+n})$ are all provable, by Lemma 2 the sentences $Prov(\ulcorner\neg\mathbf{R}(\overline{n})\urcorner), Prov(\ulcorner\neg\mathbf{R}(\overline{n+1})\urcorner), \ldots, Prov(\ulcorner\neg\mathbf{R}(\overline{n+n})\urcorner)$ are also provable.

Let c be the code number of the sequence $(n, n+1, \ldots, n+n)$[25]. Then the sentences $HetSeq(\overline{c})$, $l(\overline{c}) = S\overline{n}$, and $(\forall t \leq \overline{c})\big(Ele(t, \overline{c}) \to Prov(\ulcorner\neg\mathbf{R}(t)\urcorner)\big)$ are all provable. This implies that $\mathbf{R}(\overline{n})$ is provable and we get another contradiction.

Suppose $\mathbf{R}(\overline{n})$ is provable. Then $\exists z\Big(HetSeq(z) \wedge l(z) = S\overline{n} \wedge (\forall t \leq z)\big(Ele(t, z) \to Prov(\ulcorner\neg\mathbf{R}(t)\urcorner)\big)\Big)$ is also provable.

By ω-consistency, there is an number c such that

$$\vdash \Big(HetSeq(\overline{c}) \wedge l(\overline{c}) = S\overline{n} \wedge (\forall t \leq \overline{c})\big(Ele(t, \overline{c}) \to Prov(\ulcorner\neg\mathbf{R}(t)\urcorner)\big)\Big)$$

$$\Rightarrow\ \vdash (\forall t \leq \overline{c})\big(Ele(t, \overline{c}) \to Prov(\ulcorner\neg\mathbf{R}(t)\urcorner))\big)\Big) \qquad \text{(By conjunction elimination)}$$

$$\Rightarrow\ \vdash Prov(\ulcorner\neg\mathbf{R}(Dec(\overline{1}, \overline{c})\urcorner) \qquad \text{(By fact 3)}$$

But it is impossible by the first half of this proof, so $\mathbf{R}(\overline{n})$ is not provable. Therefore $\mathbf{R}(\overline{n})$ is neither provable nor refutable. □

9 The Surprise Examination Paradox

The surprise examination (or test) paradox, or the unexpected hanging paradox, is a paradox about a person's belief or expectation on the time of an event, which is told to be unexpected. It has various formulation, here we give one:

> A teacher tells the students that there will be an unexpected examination in a day on the next week, they will not know which day it is until that very day the teacher enters the classrooms with the examination papers.

[25] Which is the number $2^{n+1} \times 3^{n+2} \times \ldots \times p_n^{2n+1}$.

> A student start thinking: The examination cannot be held on Friday, otherwise they can deduce the examination will be on Friday, from the fact that there is no examination on Monday to Thursday[26], on Thursday after school.
>
> Then the remaining possibilities are Monday to Thursday. But if it is on Thursday, the students will know it on Wednesday after school, by a similar reasoning. By repeating this process, the student concludes that the examination cannot be held on Wednesday, Tuesday and Monday, but that means there cannot be an unexpected examination at all. The student tells this discovery to other classmates, so they just ignore what the teacher says.
>
> Eventually, there is an examination on Wednesday, and the students are surprised.

In Sorensen (1993), there is a different but related paradox called the earliest unexpected class inspection paradox. It can be presented in the following way:

> You are a new teacher, and you are told that there will be a class inspection. There are two conditions on the date of the inspection: first, the sooner the better; second, you do not know and cannot guess the day so that you cannot prepare for it. Therefore, the inspection will be on the first day which you do not believe there will be a class inspection.
>
> Now, the next school day is the first available day for class inspection, but then the inspection cannot be on that day since you can reason it. Similarly you can rule out the possibilities of the inspection being on the second day, the third day, the fourth day, and so on. Hence the earliest unexpected class inspection is impossible.

Though the reasoning processes in the two paradoxes seems to be different, namely, one is backward in time and another is forward in time, their structures are similar: First, an event, a surprise examination or an unexpected class inspection, cannot be on the first day (in the surprise examination case, we count backward from the last day). Second, if it cannot be on the first k days,

[26]Let us assume that they go to school on Monday to Friday.

then it cannot be on the $(k+1)^{st}$ day. Therefore the event cannot happen at all.

This is obviously the strong (mathematical) induction. For the surprise examination paradox, we only need to add an extra upper bound, that is, the total number of days.

9.1 Fitch's formalization

Frederic Fitch gave a formalization of the surprise examination paradox in Fitch (1964), as presented below with some modifications[27].

First, let n be a natural number, and $\mathbf{D_1}, \mathbf{D_2}, \ldots \mathbf{D_n}$ denote sentences representing "the examination occurs on the first day", "the examination occurs on the second day", ..., "the examination occurs on the n^{th} day" respectively. Notice that these symbols should be taken as extra constants, which are not in our original language. Hence in this subsection we are working in a theory which extends **T** with these symbols, let us call it **T'**.

In the following, we use the symbol $\nleftrightarrow$ for the "exclusive or", that is, $P \nleftrightarrow Q$ is defined as $\neg P \leftrightarrow Q$. And we will use two facts:

1. $P \nleftrightarrow Q$ implies $P \vee Q$.
2. $(P \nleftrightarrow Q) \nleftrightarrow R$ and $\neg R$ implies $P \nleftrightarrow Q$.

The idea is that, we want to find a formula **E** such that we cannot deduce $\mathbf{D_1}$ from **E**, $\mathbf{D_2}$ from **E** and $\neg\mathbf{D_1}$, and so on. This is similar to the reasoning of the student, but of course here we use the Diagonal Lemma, and we use $\neg Prov(Imp(x, y))$ for "cannot deduce X from Y" (where x and y are Gödel numbers of X and Y respectively).

Then consider the open formula

$$\big[\mathbf{D_1} \wedge \neg Prov\big(Imp(x, \ulcorner\mathbf{D_1}\urcorner)\big)\big] \nleftrightarrow \big[\mathbf{D_2} \wedge \neg Prov\big(Imp(Con(x, \ulcorner\neg\mathbf{D_1}\urcorner), \ulcorner\mathbf{D_2}\urcorner)\big)\big] \nleftrightarrow \\ \ldots \nleftrightarrow \big[\mathbf{D_n} \wedge \neg Prov\big(Imp(Con_{n-1}(x, \ulcorner\neg\mathbf{D_1}\urcorner, \ulcorner\neg\mathbf{D_2}\urcorner, \ldots, \ulcorner\neg\mathbf{D_{n-1}}\urcorner), \ulcorner\mathbf{D_n}\urcorner)\big)\big]$$

By the Diagonal Lemma, there is a sentence **E** such that

$$\vdash \mathbf{E} \longleftrightarrow \Big\{\big[\mathbf{D_1} \wedge \neg Prov\big(\ulcorner\mathbf{E} \to \mathbf{D_1}\urcorner\big)\big] \nleftrightarrow \big[\mathbf{D_2} \wedge \neg Prov\big(\ulcorner(\mathbf{E} \wedge \neg\mathbf{D_1}) \to \mathbf{D_2}\urcorner\big)\big] \nleftrightarrow \\ \ldots \nleftrightarrow \big[\mathbf{D_n} \wedge \neg Prov\big(\ulcorner(\mathbf{E} \wedge \neg\mathbf{D_1} \wedge \neg\mathbf{D_2} \wedge \ldots \wedge \neg\mathbf{D_{n-1}}) \to \mathbf{D_n}\urcorner\big)\big]\Big\}$$

[27] e.g. Using different letters and notations.

Then we have the following result:

Theorem 35. *If T' is consistent, then* $\mathbf{E}$ *is not provable, it is in fact refutable.*

Proof. By the choice of $\mathbf{E}$ and the fact that $P \not\leftrightarrow Q$ implies $P \vee Q$, we have $\mathbf{E} \rightarrow (\mathbf{D_1} \vee \mathbf{D_2} \vee \ldots \vee \mathbf{D_n})$. Then:

$$\vdash \mathbf{E} \wedge \neg\mathbf{D_1} \wedge \neg\mathbf{D_2} \wedge \ldots \wedge \neg\mathbf{D_{n-1}}) \rightarrow \mathbf{D_n} \qquad \text{(By propositional logic)}$$

$$\Rightarrow \vdash Prov\big(\ulcorner(\mathbf{E} \wedge \neg\mathbf{D_1} \wedge \neg\mathbf{D_2} \wedge \ldots \wedge \neg\mathbf{D_{n-1}}) \rightarrow \mathbf{D_n}\urcorner\big) \qquad \text{(By Lemma 2)}$$

$$\Rightarrow \vdash \neg\big[\mathbf{D_n} \wedge \neg Prov\big(\ulcorner(\mathbf{E} \wedge \neg\mathbf{D_1} \wedge \neg\mathbf{D_2} \wedge \ldots \wedge \neg\mathbf{D_{n-1}}) \rightarrow \mathbf{D_n}\urcorner\big)\big] \qquad \text{(By } \vee\text{-introduction)}$$

Therefore, by fact 2 above, we get $\mathbf{E} \rightarrow \Big\{\big[\mathbf{D_1} \wedge \neg Prov\big(\ulcorner\mathbf{E} \rightarrow \mathbf{D_1}\urcorner\big)\big] \not\leftrightarrow \big[\mathbf{D_2} \wedge \neg Prov\big(\ulcorner(\mathbf{E} \wedge \neg\mathbf{D_1}) \rightarrow \mathbf{D_2}\urcorner\big)\big] \not\leftrightarrow \ldots \not\leftrightarrow \big[\mathbf{D_n} \wedge \neg Prov\big(\ulcorner(\mathbf{E} \wedge \neg\mathbf{D_1} \wedge \neg\mathbf{D_2} \wedge \ldots \wedge \neg\mathbf{D_{n-2}}) \rightarrow \mathbf{D_{n-1}}\urcorner\big)\big]\Big\}$.

Similarly we can "remove" the other parts in the formula one by one, finally we get

$$\vdash \mathbf{E} \rightarrow \big[\mathbf{D_1} \wedge \neg Prov\big(\ulcorner\mathbf{E} \rightarrow \mathbf{D_1}\urcorner\big)\big]$$

$$\Rightarrow \vdash \mathbf{E} \rightarrow \mathbf{D_1}$$

$$\Rightarrow \vdash Prov(\ulcorner\mathbf{E} \rightarrow \mathbf{D_1}\urcorner) \qquad \text{(By Lemma 2)}$$

$$\Rightarrow \vdash \neg\mathbf{E}$$

□

Fitch concluded that the paradox is only a paradox in the rather weak sense that the teacher's claim is self-contradictory. Then he showed that if we modify the above formula and obtain the following fixed point $\dot{\mathbf{E}}$ where

$$\vdash \dot{\mathbf{E}} \longleftrightarrow \Big\{Prov(\ulcorner\dot{\mathbf{E}}\urcorner) \rightarrow \Big[\big[\mathbf{D_1} \wedge \neg Prov\big(\ulcorner\dot{\mathbf{E}} \rightarrow \mathbf{D_1}\urcorner\big)\big] \not\leftrightarrow \big[\mathbf{D_2} \wedge \neg Prov\big(\ulcorner(\dot{\mathbf{E}} \wedge \neg\mathbf{D_1}) \rightarrow \mathbf{D_2}\urcorner\big)\big] \not\leftrightarrow \ldots \not\leftrightarrow \big[\mathbf{D_n} \wedge \neg Prov\big(\ulcorner(\dot{\mathbf{E}} \wedge \neg\mathbf{D_1} \wedge \neg\mathbf{D_2} \wedge \ldots \wedge \neg\mathbf{D_{n-1}}) \rightarrow \mathbf{D_n}\urcorner\big)\big]\Big]\Big\}$$

then $\dot{\mathbf{E}}$ is a fixed point of $\neg Prov(x)$, hence undecidable.

However, Fitch's conclusion maybe too strong and too quick for two reason. The first reason is that he substitute knowability in the paradox by provability in his formulation, and this substitution is not self-evident. The second reason is, there maybe other ways to formalize the paradox. Also his undecidable proposition is not entirely interesting, as we cannot get rid of using the symbols $\mathbf{D_1}, \mathbf{D_2}, \ldots, \mathbf{D_n}$, which are not in our original theory.

9.2 Another formalization

In this subsection, we will formalize Sorensen's earliest unexpected class inspection paradox. With a slight modification, this formalization can also be an alternative formulation of the surprise examination paradox.

As we have seen above, the reasoning in the paradox is actually strong induction. This leads us to the idea of using the following open formula:

$$(\forall z < y)Prov\big(Neg(Sub(x, \ulcorner z \urcorner))\big) \to Prov\big(Neg(Sub(x, \ulcorner y \urcorner))\big)$$

Apply the Generalized Diagonal Lemma, we will get a fixed point $\mathbf{F}(y)$ such that

$$\mathbf{F}(y) \longleftrightarrow \big[(\forall z < y)Prov\big(\ulcorner \neg \mathbf{F}(z) \urcorner\big) \to Prov\big(\ulcorner \neg \mathbf{F}(y) \urcorner\big)\big]$$

is provable.

Then we have the following result:

Theorem 36. *If $\boldsymbol{T}$ is ω-consistent, then $\exists x \mathbf{F}(x)$ is undecidable.*

Before proving this theorem, we need to state the following fact:

Proposition 37 (The Least Number Principle)**.** *For any open formula $P(x)$ with exactly one free variable x, it is provable that $\exists x P(x) \to \exists x\big(P(x) \wedge (\forall y < x)\neg P(y)\big)$*

The proof of this principle, which can be found in (Boolos, 1993), is skipped here. Now we can prove Theorem 36.

Proof. Assume $\mathbf{T}$ is ω-consistent. Suppose $\exists x \mathbf{F}(x)$ is provable, then by the Least Number Principle, $\exists x\big(\mathbf{F}(x) \wedge (\forall z < x)\neg \mathbf{F}(z)\big)$ is also provable.

By ω-consistency of $\mathbf{T}$, there is an n such that $(\forall z < \overline{n})\neg \mathbf{F}(z) \wedge \mathbf{F}(\overline{n})$ is provable. Then $(\forall z < \overline{n})\neg \mathbf{F}(z)$ is provable, by substitution and modus ponens we have $\neg \mathbf{F}(\overline{0}), \neg \mathbf{F}(\overline{1}), \ldots, \neg \mathbf{F}(\overline{n-1})$ are all provable. Hence by Lemma 2, $Prov(\ulcorner \neg \mathbf{F}(\overline{0}) \urcorner), Prov(\ulcorner \neg \mathbf{F}(\overline{1}) \urcorner), \ldots, Prov(\ulcorner \neg \mathbf{F}(\overline{n-1}) \urcorner)$ are all provable, so is $(\forall z < \overline{n})Prov(\ulcorner \neg \mathbf{F}(z) \urcorner)$.

Since $\mathbf{F}(\overline{n})$ is also provable, we have:

$$\begin{aligned}
\vdash \mathbf{F}(\overline{n}) &\Rightarrow \vdash \big[(\forall z < \overline{n})Prov\big(\ulcorner \neg \mathbf{F}(z) \urcorner\big) \to Prov\big(\ulcorner \neg \mathbf{F}(\overline{n}) \urcorner\big)\big] && \text{(By the choie of } \mathbf{F}(x)\text{)} \\
&\Rightarrow \vdash Prov\big(\ulcorner \neg \mathbf{F}(\overline{n})\big) && \text{(By modus ponens)} \\
&\Rightarrow \vdash \neg \mathbf{F}(\overline{n}) && \text{(By Lemma 3)}
\end{aligned}$$

So we get a contradiction, and $\exists x\mathbf{F}(x)$ is not provable.

On the other hand, suppose $\exists x\mathbf{F}(x)$ is refutable. Then $\neg\exists x\mathbf{F}(x)$, and equivalently, $\forall x\neg\mathbf{F}(x)$ are provable. For any natural number n, we have:

$$\begin{aligned}
&\vdash \neg\mathbf{F}(\overline{n}) && \text{(By universal instantiation)}\\
\Rightarrow\ &\vdash (\forall z<\overline{n})Prov\big(\ulcorner\neg\mathbf{F}(z)\urcorner\big)\wedge\neg Prov(\ulcorner\neg\mathbf{F}(\overline{n})\urcorner) && \text{(By the choice of } \mathbf{F}(x))\\
\Rightarrow\ &\vdash \neg Prov(\ulcorner\neg\mathbf{F}(\overline{n})\urcorner) && \text{(By conjunction elimination)}
\end{aligned}$$

But by Lemma 2, $\neg\mathbf{F}(\overline{n})$ is provable implies that $Prov(\ulcorner\neg\mathbf{F}(\overline{n})\urcorner)$ is also provable. Hence we get a contradiction.

Therefore, $\exists x\mathbf{F}(x)$ is neither provable nor refutable. □

Roughly speaking, $\mathbf{F}(\overline{n})$ is related to the proposition "there will be a class inspection at the $(n+1)^{st}$ day"[28]. And it satisfies the condition that if it is deducible that there is no class inspection at the first n day, then it is deducible that there is no class inspection at the $(n+1)^{st}$ day.

For the surprise examination paradox, we just need to add an upper bound for the variable, say, N. There is a fixed point $\mathbf{F_N}(y)$ such that

$$\mathbf{F_N}(y)\longleftrightarrow\Big[(y<\overline{N})\wedge\big[(\forall z<y)Prov\big(\ulcorner\neg\mathbf{F_N}(z)\urcorner\big)\rightarrow Prov\big(\ulcorner\neg\mathbf{F_N}(y)\urcorner\big)\big)\big]\Big]$$

is provable. And we have the following theorem:

Theorem 38. *If* $\boldsymbol{T}$ *is* ω*-consistent, then* $(\exists x<\overline{N})\mathbf{F_N}(x)$ *is undecidable.*

Proof. Assume $\mathbf{T}$ is ω-consistent. Suppose $(\exists x<\overline{N})\mathbf{F_N}(x)$ is provable, then by the Least Number Principle, $\exists x\big[\big((x<\overline{N})\wedge\mathbf{F_N}(x)\big)\wedge(\forall z<x)\neg\big((z<\overline{N})\wedge\mathbf{F_N}(z)\big)\big]$ is also provable.

Since it is provable that $\big((x<\overline{N})\wedge(z<x)\big)\rightarrow z<\overline{N}$, the latter sentence is the last paragraph is equivalent to $\exists x\big((x<\overline{N})\wedge\mathbf{F_N}(x)\wedge(\forall z<x)\neg\mathbf{F_N}(z)\big)$. Then there is some $n<N$ such that $(\overline{n}<\overline{N})\wedge\mathbf{F_N}(\overline{n})\wedge(\forall z<\overline{n})\neg\mathbf{F_N}(z)$ is provable.

As in the previous proof, from $(\forall z<\overline{n})\neg\mathbf{F_N}(z)$ we can derive, by substitution and modus ponens, that $\neg\mathbf{F_N}(0),\neg\mathbf{F_N}(1),\ldots,\neg\mathbf{F_N}(n-1)$. Therefore $Prov(\ulcorner\neg\mathbf{F_N}(0)\urcorner), Prov(\ulcorner\neg\mathbf{F_N}(1)\urcorner),\ldots,Prov(\ulcorner\neg\mathbf{F_N}(n-1)\urcorner)$ are all provable, so is $(\forall z<\overline{n})Prov(\ulcorner\neg\mathbf{F_N}(z)\urcorner)$.

[28]Note that we count from 0.

Since $\mathbf{F_N}(\overline{n})$ is also provable, we have:

$$
\begin{aligned}
& \vdash \mathbf{F_N}(\overline{n}) \\
\Rightarrow & \vdash (\overline{n} < \overline{N}) \wedge \big[(\forall z < \overline{n}) Prov\big(\ulcorner\neg\mathbf{F_N}(z)\urcorner\big) \rightarrow Prov\big(\ulcorner\neg\mathbf{F_N}(\overline{n})\urcorner\big)\big)\big] && \text{(By definition)} \\
\Rightarrow & \vdash (\forall z < \overline{n}) Prov\big(\ulcorner\neg\mathbf{F_N}(z)\urcorner\big) \rightarrow Prov\big(\ulcorner\neg\mathbf{F_N}(\overline{n})\urcorner\big)\big) && \text{(By $\wedge$-elimination)} \\
\Rightarrow & \vdash Prov\big(\ulcorner\neg\mathbf{F_N}(\overline{n})\big) && \text{(By modus ponens)} \\
\Rightarrow & \vdash \neg\mathbf{F_N}(\overline{n}) && \text{(By Lemma 3)}
\end{aligned}
$$

Hence there is a contradiction, and $(\exists x < \overline{N})\mathbf{F_N}(x)$ is not provable.

On the other hand, suppose $(\exists x < \overline{N})\mathbf{F_N}(x)$ is refutable. Then the sentence $\neg(\exists x < \overline{N})\mathbf{F_N}(x)$ is provable, so is $(\forall x < \overline{N})\neg\mathbf{F_N}(x)$. For any number $n < N$, we have $\overline{n} < \overline{N}$ provable, and:

$$
\begin{aligned}
& \vdash \neg\mathbf{F_N}(\overline{n}) \\
\Rightarrow & \vdash \neg(\overline{n} < \overline{N}) \vee (\forall z < \overline{n}) Prov\big(\ulcorner\neg\mathbf{F_N}(z)\urcorner\big) \wedge \neg Prov(\ulcorner\neg\mathbf{F_N}(\overline{n})\urcorner) && \text{(By definition)} \\
\Rightarrow & \vdash (\overline{n} < \overline{N}) \rightarrow (\forall z < \overline{n}) Prov\big(\ulcorner\neg\mathbf{F_N}(z)\urcorner\big) \wedge \neg Prov(\ulcorner\neg\mathbf{F_N}(\overline{n})\urcorner) && \text{(By definition)} \\
\Rightarrow & \vdash (\forall z < \overline{n}) Prov\big(\ulcorner\neg\mathbf{F_N}(z)\urcorner\big) \wedge \neg Prov(\ulcorner\neg\mathbf{F_N}(\overline{n})\urcorner) && \text{(By modus ponens)} \\
\Rightarrow & \vdash \neg Prov(\ulcorner\neg\mathbf{F_N}(\overline{n})\urcorner) && \text{(By $\wedge$-elimination)}
\end{aligned}
$$

But by Lemma 2, $\neg\mathbf{F_N}(\overline{n})$ is provable implies that $Prov(\ulcorner\neg\mathbf{F_N}(\overline{n})\urcorner)$ is also provable. Hence there is a contradiction.

Therefore $(\exists x < \overline{N})\mathbf{F_N}(x)$ is undecidable. □

Some may argue that these are not accurate formalizations of the paradoxes, since the unexpected events in both paradoxes are unique, but the above undecidable sentences involve only existential quantifiers and there is nothing about the uniqueness.

This problem can be solved easily by noting that for any open formula $P(x)$ with one free variable x, $\exists x P(x)$ is provably equivalent to $\exists x\big(P(x) \wedge (\forall y < x)\neg P(y)\big)$. One direction of the equivalence is the Least Number Principle, while the other direction is trivial.

Hence from Theorem 36 and Theorem 38, we have the following corollaries:

Corollary 39. *If $\boldsymbol{T}$ is ω-consistent, then $\exists x\big(\mathbf{F}(x) \wedge (\forall y < x)\neg\mathbf{F}(y)\big)$ is undecidable.*

Corollary 40. *If $\boldsymbol{T}$ is ω-consistent, then* $\exists x\big[(x < \overline{N}) \wedge \mathbf{F_N}(x) \wedge (\forall y < x)\neg\big((y < \overline{N}) \wedge \mathbf{F}(y)\big)\big]$ *is undecidable.*

10 Concluding Remarks

The above survey suggests the claim that for many logical paradoxes there are corresponding undecidable sentences. However it should not be taken as a rigorous proof of the claim "for every logical paradox there is a corresponding undecidable sentence." for a few reasons.

First, this survey is by no means complete. Second, it seems that it can never be complete, unless we can prove that there is no more logical paradoxes other than the current ones, which sounds quite implausible. Third, even if it is only about the logical paradoxes we know, we still need a systematic analysis on the logical structures of the paradoxes, and then "translate" them into different undecidable sentences. This project is far beyond the scope of this thesis.

Furthermore, the relationship between the paradoxes and undecidable sentences above is also questionable. For example, in the last section we have seen that Fitch's conclusion that the surprise examination paradox is not a real paradox but a self-contradictory claim, while there is another formulation of the same paradox which is not contradictory but undecidable in Peano arithmetic. In Kritchman and Raz (2010) there is also a proof of the Second Incompleteness Theorem based on Chaitin's proof and the surprise examination paradox.

I am not saying that my formulation is *the* correct one (this is not my purpose to formulate it), but what should be seen as an accurate formalization should be further discussed. Therefore we should not draw too many conclusions from the undecidable sentences except that they are undecidable in Peano arithmetic.

References

Boolos, G. (1993). *The Logic of Provability*. Number 182. Cambridge University Press.

Boolos, G. (1998). A New Proof of the Gödel Incompleteness Theorem. In *Logic, Logic, and Logic*, pages 383–388. Harvard University Press.

Boolos, G., Burgess, J. P., and Jeffrey, R. C. (2007). *Computability and Logic*. Cambridge University Press.

Chaitin, G. J. (1971). Computational Complexity and Gödel's Incompleteness Theorem. *ACM SIGACT News*.

Chaitin, G. J. (1995). The Berry Paradox. *Complexity*, 1(1):26–30.

Cieśliński, C. and Urbaniak, R. (2013). Gödelizing the Yablo Sequence. *Journal of Philosophical Logic*, 42(5):679–695.

Cook, R. T. (2006). There Are Non-Circular Paradoxes (but Yablo's isn't one of them!). *The Monist*, 89(1):118–149.

Curry, H. B. (1942). The Inconsistency of Certain Formal Logics. *Journal of Symbolic Logic*, 7(3):115–117.

Fitch, F. B. (1964). A Goedelized Formulation of the Prediction Paradox. *American Philosophical Quarterly*, 1(2):161–164.

Gödel, K. (1992). *On Formally Undecidable Propositions of Principia Mathematica and Related Systems*. Dover Books on Mathematics Series. Dover Publications.

Kikuchi, M. (1994). A Note on Boolos' Proof of the Incompleteness Theorem. *Mathematical Logic Quarterly*, 40(4):528–532.

Kritchman, S. and Raz, R. (2010). The Surprise Examination Paradox and the Second Incompleteness Theorem.

Lindström, P. (1997). *Aspects of Incompleteness*. Lecture Notes in Logic 10. Springer Berlin Heidelberg.

Löb, M. H. (1955). Solution of a Problem of Leon Henkin. *Journal of Symbolic Logic*, 20(2):115–118.

Makinson, D. C. (1965). The Paradox of the Preface. *Analysis*, 25(6):205–207.

Priest, G. (1997). Yablo's Paradox. *Analysis*, 57(4):236–242.

Rosser, B. (1936). Extensions of Some Theorems of Gödel and Church. *Journal of Symbolic Logic*, 1(3):87–91.

Russell, B. (1967). Mathematical Logic as Based on the Theory of Types. In Heijenoort, J. V., editor, *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931*, pages 150–182. Cambridge, Harvard University Press.

Serény, G. (2004). Boolos-Style Proofs of Limitative Theorems. *Mathematical Logic Quarterly*, 50(2):211.

Serény, G. (2006). The Diagonal Lemma as the Formalized Grelling Paradox. *Collegium Logicum: Annals of the Kurt-Gödel Society*, 9(1):63–66.

Smith, P. (2007). *An Introduction to Gödel's Theorems*. Cambridge Introductions to Philosophy. Cambridge University Press.

Smoryński, C. (1977). The Incompleteness Theorems. In Barwise, J., editor, *Handbook of Mathematical Logic*, pages 821–865. North-Holland.

Smoryński, C. (1991). *Logical Number Theory I: An Introduction.* Universitext. Springer Berlin Heidelberg, softcover reprint of the original 1st ed. 1991 edition.

Smullyan, R. M. (1992). *Gödel's Incompleteness Theorems.* Oxford University Press.

Sorensen, R. A. (1993). The Earliest Unexpected Class Inspection. *Analysis*, 53(4):252 –.

Sorensen, R. A. (2004). *Vagueness and Contradiction.* Oxford University Press.

Yablo, S. (1993). Paradox Without Self–Reference. *Analysis*, 53(4):251–252.