Diagonalization

Volker Halbach

5th March 2023

There is something fishy about the liar paradox:

(1) (1) is not true.

Somehow the sentence ‘says’ something about itself, and when people are confronted with the paradox for the first time, they usually think that this feature is the source of the paradox.

However, there are many self-referential sentence that are completely unproblematic:

(2) (2) contains 5 occurences of the letter 'c'.

If (1) is illegitimate because of its self-referentiality, then (2) must be illegitimate as well. Moreover, the effect that is achieved via the label '(1)' can be achieved without this device. At the same time one can dispense with demonstratives like 'this' that might be used to formulate the liar sentence:

This sentence is not true.

In fact, the effect can be achieved using weak arithmetical axioms only. And the axioms employed are beyond any (serious) doubt. This was shown by Gödel.

In the following I describe a language $\mathcal{L}$. An expression of $\mathcal{L}$ is an arbitrary finite string of the following symbols. Such strings are also called expressions of $\mathcal{L}$.

Definition

The symbols of $\mathcal{L}$ are:

1. infinitely many variable symbols $\mathrm{v}, \mathrm{v}_1, \mathrm{v}_2, \mathrm{v}_3, \ldots$
2. predicate symbols $=$ and T,
3. function symbols q, $\frown$ and sub,
4. the connectives $\neg$, $\rightarrow$ and the quantifier symbol $\forall$,
5. auxiliary symbols (and),
6. possibly finitely many further function and predicate symbols, and
7. If e is a string of symbols then $\overline{e}$ is also a symbol. $\overline{e}$ is called a quotation constant.

In the following I shall use x, y and z as (meta-)variables for variables. Thus x may stand for any symbol v, v_1, v_2, ... It is also assumed that x, y etc stand for different variables. Moreover, it is always presupposed variable clashes are avoided by renaming variables in a suitable way.

In the following I shall use x, y and z as (meta-)variables for variables. Thus x may stand for any symbol v, v_1, v_2, … It is also assumed that x, y etc stand for different variables. Moreover, it is always presupposed variable clashes are avoided by renaming variables in a suitable way.

It is important that $\overline{a}$ is a single symbol and not a string of more than one symbols even if a itself is a string built from several symbols.

In the following I shall use x, y and z as (meta-)variables for variables. Thus x may stand for any symbol v, v_1, v_2, ... It is also assumed that x, y etc stand for different variables. Moreover, it is always presupposed variable clashes are avoided by renaming variables in a suitable way.

It is important that $\overline{a}$ is a single symbol and not a string of more than one symbols even if a itself is a string built from several symbols.

A string of symbols of $\mathcal{L}$ is any string of the above symbols. Usually I suppress mention of $\mathcal{L}$. The empty string is also a string.

We shall now define the notions of a term and of a formula of $\mathcal{L}$.

Definition

The $\mathcal{L}$-terms are defined as follows:

1. All variables are terms.

We shall now define the notions of a term and of a formula of $\mathcal{L}$.

Definition

The $\mathcal{L}$-terms are defined as follows:

1. All variables are terms.
2. If e is a string of symbols, then $\overline{e}$ is a term.

We shall now define the notions of a term and of a formula of $\mathcal{L}$.

Definition

The $\mathcal{L}$-terms are defined as follows:

1. All variables are terms.
2. If e is a string of symbols, then $\overline{e}$ is a term.
3. If t, r and s are terms, then $\mathrm{q}(t)$, $(s^\frown t)$, $\mathrm{sub}(r, s, t)$ are terms, and similarly for all further function symbols

Since the empty string is a string of symbols $\overline{\ }$ is a term. Since $\overline{\ }$ looks so odd, I shall write $\underline{0}$ for $\overline{\ }$.

What the empty string is for the expressions is the number zero for the natural numbers. It is not hard to see that 0 is useful in number theory.

Formulæ, sentences, free and bound occurrences of variables are defined in the usual way.

Example

1. $\forall v_3(v_3 = \overline{\wedge\forall} \wedge T\overline{v_3})$ is a sentence.
2. $\overline{v_{12}} = \overline{\neg T\overline{\neg}}$ is a sentence, i.e., the formula does not feature a free variable.

The theory $\mathcal{A}$ which will be described in this section is designed in order to obtain smooth proofs. I have not aimed at a particularly elegant axiomatization.

The theory $\mathcal{A}$ which will be described in this section is designed in order to obtain smooth proofs. I have not aimed at a particularly elegant axiomatization.

A simple intended model of the theory has all expressions of $\mathcal{L}$ as its domain. The intended interpretation of the function symbols will become clear from the axioms A1–A4 except for the interpretation of sub. I shall return to sub below.

All instances of the following schemata and rules are axioms of the theory $\mathcal{A}$:

Definition

A1 all axioms and rules of first-order predicate logic including the identity axioms.

A2 $(\overline{a}^\frown\overline{b}) = \overline{ab}$, where a and b are arbitrary strings of symbols.

A3 $\mathrm{q}(\overline{a}) = \overline{\overline{a}}$

A4 $\mathrm{sub}(\overline{a}, \overline{b}, \overline{c}) = \overline{d}$, where a and c are arbitrary strings of symbols, b is a single symbol (or, equivalently, a string of symbols of length 1), and d is the string of symbols obtained from a by replacing all occurrences of the symbol b by the strings c.

A2

$(\overline{a}^\frown\overline{b}) = \overline{ab}$, where a and b are arbitrary strings of symbols.

The concatenation of two expressions e_1 and e_2 is simply the expression e_1 followed by e_2. For instance, ¬¬v is the concatenation of ¬ and ¬v.

A2

$(\overline{a}{}^\frown\overline{b}) = \overline{ab}$, where a and b are arbitrary strings of symbols.

The concatenation of two expressions e_1 and e_2 is simply the expression e_1 followed by e_2. For instance, ¬¬v is the concatenation of ¬ and ¬v.

Therefore $(\overline{\neg}{}^\frown\overline{\neg \mathrm{v}}) = \overline{\neg\neg \mathrm{v}}$ is an instance of A2 as well as $(\overline{\neg\neg}{}^\frown\overline{\mathrm{v}}) = \overline{\neg\neg \mathrm{v}}$.

A2

$(\overline{a}^\frown\overline{b}) = \overline{ab}$, where a and b are arbitrary strings of symbols.

The concatenation of two expressions e_1 and e_2 is simply the expression e_1 followed by e_2. For instance, ¬¬v is the concatenation of ¬ and ¬v.

Therefore $(\overline{\neg}^\frown\overline{\neg\mathrm{v}}) = \overline{\neg\neg\mathrm{v}}$ is an instance of A2 as well as $(\overline{\neg\neg}^\frown\overline{\mathrm{v}}) = \overline{\neg\neg\mathrm{v}}$.

Concatenating the empty string with any expression e gives again the same expression e. Therefore we have, for instance, $(\overline{\forall}^\frown\underline{0}) = \overline{\forall}$ as an instance of A2.

A2

$(\overline{a}^\frown\overline{b}) = \overline{ab}$, where a and b are arbitrary strings of symbols.

The concatenation of two expressions e_1 and e_2 is simply the expression e_1 followed by e_2. For instance, ¬¬v is the concatenation of ¬ and ¬v.

Therefore $(\overline{\neg}^\frown\overline{\neg\mathrm{v}}) = \overline{\neg\neg\mathrm{v}}$ is an instance of A2 as well as $(\overline{\neg\neg}^\frown\overline{\mathrm{v}}) = \overline{\neg\neg\mathrm{v}}$.

Concatenating the empty string with any expression e gives again the same expression e. Therefore we have, for instance, $(\overline{\forall}^\frown\underline{0}) = \overline{\forall}$ as an instance of A2.

Sometimes I forget the brackets.

A3
$q(\overline{a}) = \overline{\overline{a}}$

Instances of A3 are

$$q(\overline{\neg}) = \overline{\overline{\neg}} \qquad \text{and} \qquad q(\overline{\overline{\mathrm{v}}\neg}) = \overline{\overline{\overline{\mathrm{v}}\neg}}$$

Thus q describes the function that takes an expression and returns its quotation constant.

A4

sub($\overline{a}$, $\overline{b}$, $\overline{c}$) = $\overline{d}$, where a and c are arbitrary strings of symbols, b is a single symbol (or, equivalently, a string of symbols of length 1), and d is the string of symbols obtained from a by replacing all occurrences of the symbol b by the strings c.

I have imposed the restriction that b must be a single symbol. This does not imply that the substitution function cannot be applied to complex expressions; just A4 does not say anything about the result of substituting a complex expression.

A4

$\text{sub}(\overline{a}, \overline{b}, \overline{c}) = \overline{d}$, where a and c are arbitrary strings of symbols, b is a single symbol (or, equivalently, a string of symbols of length 1), and d is the string of symbols obtained from a by replacing all occurrences of the symbol b by the strings c.

I have imposed the restriction that b must be a single symbol. This does not imply that the substitution function cannot be applied to complex expressions; just A4 does not say anything about the result of substituting a complex expression.

The reason for this restriction is that the result of substitution of a complex strings may be not unique. For instance, the result of substituting $\neg$ for $\wedge\wedge$ in $\wedge\wedge\wedge$ might be either $\wedge\neg$ or $\neg\wedge$. The problem can be fixed in several ways, but I do not need to substitute complex expressions in the following.

A4

$\text{sub}(\overline{a}, \overline{b}, \overline{c}) = \overline{d}$, where a and c are arbitrary strings of symbols, b is a single symbol (or, equivalently, a string of symbols of length 1), and d is the string of symbols obtained from a by replacing all occurrences of the symbol b by the strings c.

Here are instances of A4:

$$\text{sub}(\overline{\text{v}\neg}, \overline{\text{v}}, \overline{\forall}) = \overline{\forall\neg}$$

$$\text{sub}(\overline{\neg\neg}, \overline{\neg}, \overline{\neg\neg\neg}) = \overline{\neg\neg\neg\neg\neg\neg}$$

$$\text{sub}(\overline{\text{q}(\text{v})}, \overline{\text{v}}, \overline{\text{q}(\text{v})}) = \overline{\text{q}(\text{q}(\text{v}))}$$

Of course, there is no such cheap way to Gödel's theorems. Gödel showed that the functions sub and q (and further operations) can be defined in an arithmetical theory for numerical codes of expressions. To this end he proved that all recursive functions can be represented in a fixed arithmetical system. And then he proved that the operation of substitution etc. are recursive. This requires some work and ideas.

The diagonalization function dia is defined in the following way:

Definition
$\mathrm{dia}(x) = \mathrm{sub}(x, \overline{\mathrm{v}}, \mathrm{q}(x))$

dia is merely an abbreviation, not a new symbol.

$\mathrm{dia}(x)$ expresses the function that substitutes v with the quotation of x in x itself.

The diagonalization function dia is defined in the following way:

Definition

$\mathrm{dia}(x) = \mathrm{sub}(x, \overline{\mathrm{v}}, \mathrm{q}(x))$

dia is merely an abbreviation, not a new symbol.

$\mathrm{dia}(x)$ expresses the function that substitutes v with the quotation of x in x itself.

Example

$\mathcal{A} \vdash \mathrm{dia}(\overline{\neg T\mathrm{v}}) = \overline{\neg T \overline{\neg T\mathrm{v}}}$

Example

$\mathcal{A} \vdash \mathrm{dia}(\overline{\mathrm{v} = \mathrm{v}}) = \overline{\overline{\mathrm{v} = \mathrm{v}} = \overline{\mathrm{v} = \mathrm{v}}}$

Lemma

Assume $\varphi(\mathrm{v})$ is a formula not containing bound occurrences of v. Then the following holds:

$$\mathcal{A} \vdash \mathrm{dia}(\overline{\varphi(\mathrm{dia}(\mathrm{v}))}) = \overline{\varphi(\mathrm{dia}(\overline{\varphi(\mathrm{dia}(\mathrm{v}))}))}$$

Lemma

Assume $\varphi(\mathrm{v})$ is a formula not containing bound occurrences of v. Then the following holds:

$$\mathcal{A} \vdash \mathrm{dia}(\overline{\varphi(\mathrm{dia}(\mathrm{v}))}) = \overline{\varphi(\mathrm{dia}(\overline{\varphi(\mathrm{dia}(\mathrm{v}))}))}$$

PROOF.

In $\mathcal{A}$ the following equations can be proved:

$$\mathrm{dia}(\overline{\varphi(\mathrm{dia}(\mathrm{v}))}) = \mathrm{sub}(\overline{\varphi(\mathrm{dia}(\mathrm{v}))}, \overline{\mathrm{v}}, \mathrm{q}(\overline{\varphi(\mathrm{dia}(\mathrm{v}))}))$$

$\dashv$

$\mathrm{dia}(x)$ is $\mathrm{sub}(x, \overline{\mathrm{v}}, \mathrm{q}(x))$

Lemma

Assume $\varphi(\mathrm{v})$ is a formula not containing bound occurrences of v. Then the following holds:

$$\mathcal{A} \vdash \mathrm{dia}(\overline{\varphi(\mathrm{dia}(\mathrm{v}))}) = \overline{\varphi(\mathrm{dia}(\overline{\varphi(\mathrm{dia}(\mathrm{v}))}))}$$

PROOF.

In $\mathcal{A}$ the following equations can be proved:

$$\begin{aligned}\mathrm{dia}(\overline{\varphi(\mathrm{dia}(\mathrm{v}))}) &= \mathrm{sub}(\overline{\varphi(\mathrm{dia}(\mathrm{v}))}, \overline{\mathrm{v}}, \mathrm{q}(\overline{\varphi(\mathrm{dia}(\mathrm{v}))})) \\ &= \mathrm{sub}(\overline{\varphi(\mathrm{dia}(\mathrm{v}))}, \overline{\mathrm{v}}, \overline{\overline{\varphi(\mathrm{dia}(\mathrm{v}))}})\end{aligned}$$

⊣

A3: $\mathrm{q}(\overline{\varphi(\mathrm{dia}(\mathrm{v}))}) = \overline{\overline{\varphi(\mathrm{dia}(\mathrm{v}))}}$

Lemma

Assume $\varphi(v)$ is a formula not containing bound occurrences of v. Then the following holds:

$$\mathcal{A} \vdash \mathrm{dia}(\overline{\varphi(\mathrm{dia}(v))}) = \overline{\varphi(\mathrm{dia}(\overline{\varphi(\mathrm{dia}(v))}))}$$

PROOF.

In $\mathcal{A}$ the following equations can be proved:

$$\begin{aligned}\mathrm{dia}(\overline{\varphi(\mathrm{dia}(v))}) &= \mathrm{sub}(\overline{\varphi(\mathrm{dia}(v))}, \bar{v}, \mathrm{q}(\overline{\varphi(\mathrm{dia}(v))}))\\ &= \mathrm{sub}(\overline{\varphi(\mathrm{dia}(v))}, \bar{v}, \overline{\overline{\varphi(\mathrm{dia}(v))}})\\ &= \overline{\varphi(\mathrm{dia}(\overline{\varphi(\mathrm{dia}(v))}))}\end{aligned} \qquad \dashv$$

This is an instance of A4.

Lemma (repeated)

Assume $\varphi(\mathrm{v})$ is a formula not containing bound occurrences of v. Then the following holds:

$$\mathcal{A} \vdash \underbrace{\mathrm{dia}(\overline{\varphi(\mathrm{dia}(\mathrm{v}))})}_{t} = \overline{\varphi(\underbrace{\mathrm{dia}(\overline{\varphi(\mathrm{dia}(\mathrm{v}))})}_{t})}$$

Lemma (repeated)

Assume $\varphi(\mathrm{v})$ is a formula not containing bound occurrences of v. Then the following holds:

$$\mathcal{A} \vdash \underbrace{\mathrm{dia}(\overline{\varphi(\mathrm{dia}(\mathrm{v}))})}_{t} = \overline{\varphi(\underbrace{\mathrm{dia}(\overline{\varphi(\mathrm{dia}(\mathrm{v}))})}_{t})}$$

Theorem (strong diagonalization)

For every formula $\varphi(\mathrm{v})$ not containing bound occurrences of v there is a closed term t such that

$$\mathcal{A} \vdash t = \overline{\varphi(t)}$$

$\varphi(t)$ is now a diagonal sentence γ of $\varphi(\mathrm{v})$. We obviously have

$$\mathcal{A} \vdash \underbrace{\varphi(t)}_{\gamma} \leftrightarrow \varphi(\underbrace{\overline{\varphi(t)}}_{\gamma})$$

Theorem (diagonalization)

If $\varphi(\mathrm{v})$ is a formula of $\mathcal{L}$ with no bound occurrences of v, then one can find a formula γ such that the following holds:

$$\mathcal{A} \vdash \gamma \leftrightarrow \varphi(\overline{\gamma})$$

Theorem (diagonalization)

If $\varphi(\mathrm{v})$ is a formula of $\mathcal{L}$ with no bound occurrences of v, then one can find a formula γ such that the following holds:

$$\mathcal{A} \vdash \gamma \leftrightarrow \varphi(\overline{\gamma})$$

PROOF.

Choose as γ the formula $\varphi(\mathrm{dia}(\overline{\varphi(\mathrm{dia}(\mathrm{v}))}))$. Then one has by the previous Lemma:

$$\mathcal{A} \vdash \underbrace{\varphi(\mathrm{dia}(\overline{\varphi(\mathrm{dia}(\mathrm{v}))}))}_{\gamma} \leftrightarrow \varphi(\overline{\underbrace{\varphi(\mathrm{dia}(\overline{\varphi(\mathrm{dia}(\mathrm{v}))}))}_{\gamma}})$$

$\dashv$

Theorem (liar paradox)

The T-scheme $T\overline{\psi} \leftrightarrow \psi$ for all sentences ψ of $\mathcal{L}$ is inconsistent.

Theorem (liar paradox)

The T-scheme $T\overline{\psi} \leftrightarrow \psi$ for all sentences ψ of $\mathcal{L}$ is inconsistent.

PROOF.

Apply the diagonalization theorem 11 to the formula $\neg T\mathrm{v}$. Then theorem 11 implies the existence of a sentence γ such that the following holds: $\mathcal{A} \vdash \gamma \leftrightarrow \neg T\overline{\gamma}$. Together with the instance $T\overline{\gamma} \leftrightarrow \gamma$ of the T-scheme this yields an inconsistency. γ is called the 'liar sentence'. $\dashv$

Since the scheme is inconsistent such a truth predicate cannot be defined in $\mathcal{A}$, unless $\mathcal{A}$ itself is inconsistent.

Corollary (Tarski's theorem on the undefinability of truth)

There is no formula $\tau(\mathrm{v})$ such that $\tau(\overline{\psi}) \leftrightarrow \psi$ can be derived in $\mathcal{A}$ for all sentences ψ of $\mathcal{L}$, if $\mathcal{A}$ is consistent.

Since the scheme is inconsistent such a truth predicate cannot be defined in $\mathcal{A}$, unless $\mathcal{A}$ itself is inconsistent.

Corollary (Tarski's theorem on the undefinability of truth)

There is no formula $\tau(\mathrm{v})$ such that $\tau(\overline{\psi}) \leftrightarrow \psi$ can be derived in $\mathcal{A}$ for all sentences ψ of $\mathcal{L}$, if $\mathcal{A}$ is consistent.

PROOF.

Apply the diagonalization theorem 11 to $\tau(\mathrm{v})$ as above. If $\tau(\mathrm{v})$ contains bound occurrences of v they can be renamed such that there are no bound occurrences of v. $\dashv$

Theorem (Montague's 1963 paradox)

The schema $T\overline{\psi} \to \psi$ is inconsistent with the rule $\frac{\psi}{T\overline{\psi}}$.

The rule $\frac{\psi}{T\overline{\psi}}$ is called NEC in the following.

Theorem (Montague's 1963 paradox)

The schema $T\overline{\psi} \to \psi$ is inconsistent with the rule $\frac{\psi}{T\overline{\psi}}$.

The rule $\frac{\psi}{T\overline{\psi}}$ is called NEC in the following.

PROOF.

$\gamma \leftrightarrow \neg T\overline{\gamma}$	diagonalization
$T\overline{\gamma} \leftrightarrow \neg\gamma$	

$\dashv$

Theorem (Montague's 1963 paradox)

The schema $T\overline{\psi} \to \psi$ is inconsistent with the rule $\frac{\psi}{T\overline{\psi}}$.

The rule $\frac{\psi}{T\overline{\psi}}$ is called NEC in the following.

PROOF.

$\gamma \leftrightarrow \neg T\overline{\gamma}$	diagonalization
$T\overline{\gamma} \leftrightarrow \neg\gamma$	
$T\overline{\gamma} \to \gamma$	inst. of schema

$\dashv$

Theorem (Montague's 1963 paradox)

The schema $T\overline{\psi} \to \psi$ is inconsistent with the rule $\frac{\psi}{T\overline{\psi}}$.

The rule $\frac{\psi}{T\overline{\psi}}$ is called NEC in the following.

PROOF.

$\gamma \leftrightarrow \neg T\overline{\gamma}$	diagonalization
$T\overline{\gamma} \leftrightarrow \neg\gamma$	
$T\overline{\gamma} \to \gamma$	inst. of schema
$\neg T\overline{\gamma}$	logic

$\dashv$

Theorem (Montague's 1963 paradox)

The schema $T\overline{\psi} \to \psi$ is inconsistent with the rule $\frac{\psi}{T\overline{\psi}}$.

The rule $\frac{\psi}{T\overline{\psi}}$ is called NEC in the following.

PROOF.

$\gamma \leftrightarrow \neg T\overline{\gamma}$	diagonalization
$T\overline{\gamma} \leftrightarrow \neg\gamma$	
$T\overline{\gamma} \to \gamma$	inst. of schema
$\neg T\overline{\gamma}$	logic
γ	first line

$\dashv$

Theorem (Montague's 1963 paradox)

The schema $T\overline{\psi} \to \psi$ is inconsistent with the rule $\frac{\psi}{T\overline{\psi}}$.

The rule $\frac{\psi}{T\overline{\psi}}$ is called NEC in the following.

PROOF.

$\gamma \leftrightarrow \neg T\overline{\gamma}$	diagonalization
$T\overline{\gamma} \leftrightarrow \neg\gamma$	
$T\overline{\gamma} \to \gamma$	inst. of schema
$\neg T\overline{\gamma}$	logic
γ	first line
$T\overline{\gamma}$	NEC

$\dashv$

For the paradoxes of interaction the simple diagonal lemma is often insufficient. I need also the following version:

Theorem

Let T and N be two unary predicate symbols in $\mathcal{L}$. Then there is a sentence γ such that $\mathcal{A} \vdash \gamma \leftrightarrow \neg N\overline{T\overline{\gamma}}$.

PROOF.

Apply the simple diagonal lemma to the formula $\neg N(\overline{T}{}^\frown \mathrm{q}(\mathrm{v}))$, which gives a formula γ such that

$$\mathcal{A} \vdash \gamma \leftrightarrow \neg N(\overline{T}{}^\frown \mathrm{q}(\overline{\gamma}))$$

We also have

$$\begin{aligned}\mathcal{A} \vdash \overline{T}{}^\frown \mathrm{q}(\overline{\gamma}) &= \overline{T}{}^\frown \overline{\overline{\gamma}} \\ &= \overline{T\overline{\gamma}}\end{aligned}$$

⊣