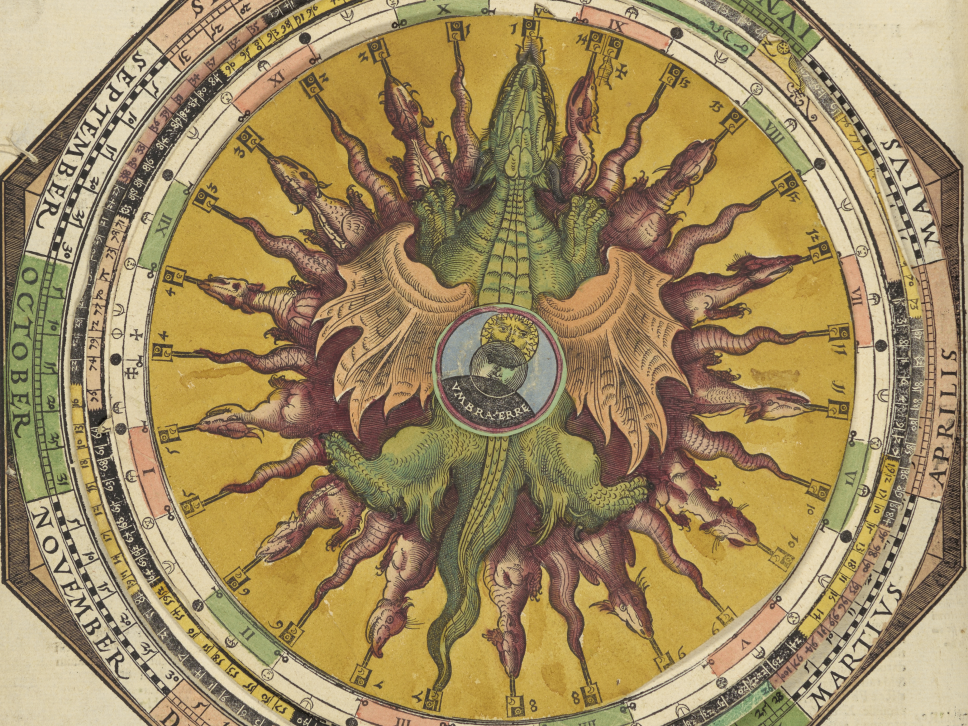
SEPTEMBER
OCTOBER
NOVEMBER
MARTIVS
APRILIS
MAIVS
VMBRA TERRE

Diagonalization and the Paradoxes

Volker Halbach

9th January 2025

Definition

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

(i) infinitely many variable symbols $\mathsf{v}_0, \mathsf{v}_1, \mathsf{v}_2, \mathsf{v}_3, \ldots$,
(ii) predicate symbols $=$ and $\square$,
(iii) function symbols q, $\frown$, and sub,
(iv) the connectives $\neg$, $\rightarrow$ and the quantifier symbol $\forall$,
(v) auxiliary symbols (and),
(vi) possibly finitely many further function and predicate symbols of arbitrary arities and finitely many further auxiliary symbols, and
(vii) for each string e of symbols exactly one constant.

The language $\mathcal{L}$

DEFINITION CONTINUED

All symbols are pairwise distinct. For instance, v_0 is distinct from $\mathsf{v}_1, \mathsf{v}_2, \ldots, =$, and so on; v_1 is distinct from $\mathsf{v}_2, \ldots, =$, and so on. In particular, if *e* is any string of symbols, the constant for *e* is distinct from *e* itself and from all symbols in (i)–(vi); and if *f* is a string of symbols distinct from *e*, then the constants for *e* and *f* are also distinct. Consequently, the constant for *e* is distinct from the constant for the *constant* of *e*, and so on. For (vii) we assume that each constant is also associated with an expression, although this will be needed only later.

There are no further symbols in $\mathcal{L}$ beyond those in (i)–(vii).

The problem mentioned above are eliminated by the following assumption:

UNIQUE READABILITY ASSUMPTION

Assume that $a_1, \ldots, a_n, b_1, \ldots, b_k$ are symbols of $\mathcal{L}$. If the string $a_1 \ldots a_n$ is identical to the string $b_1 \ldots b_k$, then $n = k$, $a_1 = b_1, \ldots,$ and $a_n = b_k$.

Example outside $\mathcal{L}$: W and W. The latter was generated by two V with reduced kerning.

The language $\mathcal{L}$ itself has different notations.

Quotation constants are strange because of this. In my notation they are complex and one can read off from them for which expression they are a constant. There are other notations.

The overlining notation has its own problems.

Later quotations will be understood as complex.

The language $\mathcal{L}$ itself has different notations.

Quotation constants are strange because of this. In my notation they are complex and one can read off from them for which expression they are a constant. There are other notations.

The overlining notation has its own problems.

Later quotations will be understood as complex.

Definition

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

(i) All variables are terms.
(ii) All quotation constants are terms.
(iii) If t, r, and s are terms, then $\mathsf{q}t$, $\frown st$, and $\mathsf{sub}\, rst$ are terms.
(iv) If $t_1, \ldots, t_n$ are terms and f is one of the additional function symbols of arity n, then $ft_1 \ldots t_n$ is a term.
(v) Nothing else is an $\mathcal{L}$-term.

The term $\frown st$ will be written as $(s \frown t)$. $(s \frown t \frown u)$ is short for $\big((s \frown t) \frown u\big)$. We will also often add brackets and commas for readability and write, for instance, $\mathsf{sub}(r, s, t)$ instead of $\mathsf{sub}\, rst$. In the following definitions we drop the analogous clauses stating that nothing else is a formula, sentence, and so on.

$^{-}$ is a term. I write $\underline{0}$ for $^{-}$.

Definition

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

(i) If s and t are terms, then $=st$ and $\Box s$ are atomic formulæ.

(ii) If $t_1, \ldots, t_n$ are terms and P is one of the additional predicate symbols of arity n, then $Pt_1 \ldots t_n$ is an atomic formula.

The atomic formula $=st$ is written as $s = t$.

Definition

If φ and ψ are formulæ and x is a variable, then $\neg\varphi$, $(\varphi \to \psi)$, and $\forall x\, \varphi$ are formulæ.

Definition

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

(i) If s and t are terms, then $=st$ and $\Box s$ are atomic formulæ.

(ii) If $t_1, \ldots, t_n$ are terms and P is one of the additional predicate symbols of arity n, then $Pt_1 \ldots t_n$ is an atomic formula.

The atomic formula $=st$ is written as $s = t$.

Definition

If φ and ψ are formulæ and x is a variable, then $\neg\varphi$, $(\varphi \to \psi)$, and $\forall x\, \varphi$ are formulæ.

Definition

(i) Every occurrence of a variable in an atomic formula is free in that formula.

(ii) All occurrences of free variables y in φ are also free in $\forall x\,\varphi$ iff y is distinct from x. All other occurrences of variables are not free.

An occurrence of a variable in a formula is bound iff it is not free.

Definition

A formula is a sentence iff it does not contain a free occurrence of a variable.

Definition

(i) Every occurrence of a variable in an atomic formula is free in that formula.

(ii) All occurrences of free variables y in φ are also free in $\forall x\, \varphi$ iff y is distinct from x. All other occurrences of variables are not free.

An occurrence of a variable in a formula is bound iff it is not free.

Definition

A formula is a sentence iff it does not contain a free occurrence of a variable.

$$\neg\Big(\forall \mathsf{v}_3\big(\mathsf{v}_3 = \overline{\to\forall} \;\to\; \neg\,\mathsf{v}_3 = \overline{\overline{\to\forall}}\,\big) \;\to\; \Box\overline{\mathsf{v}_3}\Big),$$

$$\overline{\mathsf{v}_{12}} = \overline{\neg\Box\overline{\neg}}.$$

The axioms

I will now give the axioms of the theory E.

I will say that E contains certain axioms and rules, but it may also contain more. I aim at a minimal set of assumptions that are sufficient for generating the paradoxes. The weaker the assumption, the stronger the inconsistency result.

If we try to prove more fancy result, we have to make more assumptions about E.

Definition

All instances of the following schemas are axioms of E:

A1 $\overline{a} \frown \overline{b} = \overline{ab}$, where a and b are arbitrary strings of symbols

A2 $\mathsf{q}(\overline{a}) = \overline{\overline{a}}$

A3 $\mathsf{sub}(\overline{a}, \overline{b}, \overline{c}) = \overline{d}$, where a and c are arbitrary strings of symbols, b is a symbol (or, equivalently, a length-1 string of symbols), and d is the string of symbols obtained from a by replacing all occurrences of the symbol b with c

A4 $\forall x \forall y \forall z \big((x \frown y) \frown z\big) = \big(x \frown (y \frown z)\big)$

A5 $\forall x \forall y \big(x \frown y = \underline{0} \rightarrow x = \underline{0} \land y = \underline{0}\big)$

A6 $\forall x \forall y (x \frown y = x \leftrightarrow y = \underline{0}) \land \forall x \forall y (y \frown x = x \leftrightarrow y = \underline{0})$

A7 $\forall x \forall y \, \mathsf{sub}(x \frown \overline{a}, \overline{a}, y) = \mathsf{sub}(x, \overline{a}, y) \frown y$, where a is a symbol

A8 $\forall x \forall y \forall z \forall w$
$\Big(x \frown y = z \frown w \leftrightarrow \exists v_4\big((x = z \frown v_4 \land v_4 \frown y = w) \lor (x \frown v_4 = z \land y = v_4 \frown w)\big)\Big)$

I have added brackets to A2, A3, and A7 and used infix notation.

Comments

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

Therefore $\overline{\neg\neg v} = \overline{\neg} \frown \overline{\neg v}$ is an instance of A1 as well as $\overline{\neg\neg v} = \overline{\neg\neg} \frown \overline{v}$.

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 A1.

Comments

The concatenation of two expressions e_1 and e_2 is simply the expression e_1 followed by e_2. For instance, $\neg\neg\mathsf{v}$ is the concatenation of $\neg$ and $\neg\mathsf{v}$.

Therefore $\overline{\neg\neg\mathsf{v}} = \overline{\neg} \frown \overline{\neg\mathsf{v}}$ is an instance of A1 as well as $\overline{\neg\neg\mathsf{v}} = \overline{\neg\neg} \frown \overline{\mathsf{v}}$.

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 A1.

The concatenation of two expressions e_1 and e_2 is simply the expression e_1 followed by e_2. For instance, $\neg\neg\mathsf{v}$ is the concatenation of $\neg$ and $\neg\mathsf{v}$.

Therefore $\overline{\neg\neg\mathsf{v}} = \overline{\neg} \frown \overline{\neg\mathsf{v}}$ is an instance of A1 as well as $\overline{\neg\neg\mathsf{v}} = \overline{\neg\neg} \frown \overline{\mathsf{v}}$.

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 A1.

An instance of A2 is the sentence $\mathsf{q}\ulcorner\mathsf{v}\urcorner = \ulcorner\ulcorner\mathsf{v}\urcorner\urcorner$. Thus q describes the function that takes an expression and returns its quotation constant.

In A3 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 A3 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. Therefore I do not 'solve' the problem but avoid it by the restriction of *b* to a single symbol.

In A3 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 A3 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. Therefore I do not 'solve' the problem but avoid it by the restriction of b to a single symbol.

A1-A3 are already sufficient for proving the diagonalization Theorem 10.

A4 simplifies the reasoning with strings a great deal. Since $\mathsf{E} \vdash (x \frown y) \frown z = x \frown (y \frown z)$, that is, $\frown$ is associative by A4, I shall simply write $x \frown y \frown z$. for the sake of definiteness we can stipulate that $x \frown y \frown z$ is short for $(x \frown y) \frown z$ and similarly for more applications of $\frown$.

A1-A3 are already sufficient for proving the diagonalization Theorem 10.

A4 simplifies the reasoning with strings a great deal. Since $\mathsf{E} \vdash (x \frown y) \frown z = x \frown (y \frown z)$, that is, $\frown$ is associative by A4, I shall simply write $x \frown y \frown z$. for the sake of definiteness we can stipulate that $x \frown y \frown z$ is short for $(x \frown y) \frown z$ and similarly for more applications of $\frown$.

I write $\mathsf{E} \vdash \varphi$ if and only if the formula φ is a logical consequence of the theory E.

EXAMPLE

$$\mathsf{E} \vdash \mathsf{sub}(\overline{\neg\neg}, \overline{\neg}, \overline{\neg\neg\neg}) = \overline{\neg\neg\neg\neg\neg\neg}$$

EXAMPLE

$$\mathsf{E} \vdash \mathsf{sub}(\overline{\mathsf{v} = \mathsf{v} \wedge \overline{\mathsf{v}} = \overline{\mathsf{v}}}, \overline{\mathsf{v}}, \overline{\mathsf{v}_2}) = \overline{\mathsf{v}_2 = \mathsf{v}_2 \wedge \overline{\mathsf{v}} = \overline{\mathsf{v}}}$$

I write $\mathsf{E} \vdash \varphi$ if and only if the formula φ is a logical consequence of the theory E.

EXAMPLE

$$\mathsf{E} \vdash \mathsf{sub}(\overline{\neg\neg}, \overline{\neg}, \overline{\neg\neg\neg}) = \overline{\neg\neg\neg\neg\neg\neg\neg}$$

EXAMPLE

$$\mathsf{E} \vdash \mathsf{sub}(\overline{\mathsf{v} = \mathsf{v} \wedge \overline{\mathsf{v}} = \overline{\mathsf{v}}}, \overline{\mathsf{v}}, \overline{\mathsf{v}_2}) = \overline{\mathsf{v}_2 = \mathsf{v}_2 \wedge \overline{\mathsf{v}} = \overline{\mathsf{v}}}$$

I write $\mathsf{E} \vdash \varphi$ if and only if the formula φ is a logical consequence of the theory E.

EXAMPLE

$\mathsf{E} \vdash \mathsf{sub}(\overline{\neg\neg}, \overline{\neg}, \overline{\neg\neg\neg}) = \overline{\neg\neg\neg\neg\neg\neg\neg}$

EXAMPLE

$\mathsf{E} \vdash \mathsf{sub}(\overline{\mathsf{v} = \mathsf{v} \wedge \overline{\mathsf{v}} = \overline{\mathsf{v}}}, \overline{\mathsf{v}}, \overline{\mathsf{v}_2}) = \overline{\mathsf{v}_2 = \mathsf{v}_2 \wedge \overline{\mathsf{v}} = \overline{\mathsf{v}}}$

These axioms suffice for proving Gödel's celebrated diagonalization lemma.

REMARK

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.

diagonalization

The diagonalization function dia is defined in the following way:

DEFINITION

$$\mathsf{dia}(x) = \mathsf{sub}(x, \bar{\mathsf{v}}, \mathsf{q}(x))$$

REMARK

There are at least two ways to understand the syntactical status of dia. It may be considered an additional unary functionof $\mathcal{L}$, and the above equation is then an additional axiom of E. Alternatively, one can conceive dia as a metalinguistic abbreviation, which does not form part of the language $\mathcal{L}$, but which is just short notation for a more complex expression. This situation will encountered in the following frequently.

The diagonalization function dia is defined in the following way:

Definition

$$\mathsf{dia}(x) = \mathsf{sub}(x, \bar{\mathsf{v}}, \mathsf{q}(x))$$

remark

There are at least two ways to understand the syntactical status of dia. It may be considered an additional unary functionof $\mathcal{L}$, and the above equation is then an additional axiom of E. Alternatively, one can conceive dia as a metalinguistic abbreviation, which does not form part of the language $\mathcal{L}$, but which is just short notation for a more complex expression. This situation will encountered in the following frequently.

Lemma

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

$$\mathsf{E} \vdash \mathsf{dia}(\overline{\varphi(\mathsf{dia}(\mathsf{v}))}) = \overline{\varphi(\mathsf{dia}(\overline{\varphi(\mathsf{dia}(\mathsf{v}))}))}$$

Proof.

In E the following equations can be proved::

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

□

Lemma

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

$$\mathsf{E} \vdash \mathsf{dia}(\overline{\varphi(\mathsf{dia}(\mathsf{v}))}) = \overline{\varphi(\mathsf{dia}(\overline{\varphi(\mathsf{dia}(\mathsf{v}))}))}$$

Proof.

In E the following equations can be proved::

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

□

Theorem (diagonalization)

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

$$\mathsf{E} \vdash \gamma \leftrightarrow \varphi(\overline{\gamma})$$

Proof.

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

$$\mathsf{E} \vdash \underbrace{\varphi(\mathsf{dia}(\overline{\varphi(\mathsf{dia}(\mathsf{v}))}))}_{\gamma} \leftrightarrow \varphi(\overline{\underbrace{\varphi(\mathsf{dia}(\overline{\varphi(\mathsf{dia}(\mathsf{v}))}))}_{\gamma}})$$

□

Theorem (diagonalization)

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

$$\mathsf{E} \vdash \gamma \leftrightarrow \varphi(\overline{\gamma})$$

Proof.

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

$$\mathsf{E} \vdash \underbrace{\varphi(\mathsf{dia}(\overline{\varphi(\mathsf{dia}(\mathsf{v}))}))}_{\gamma} \leftrightarrow \varphi(\overline{\underbrace{\varphi(\mathsf{dia}(\overline{\varphi(\mathsf{dia}(\mathsf{v}))}))}_{\gamma}})$$

□

The comparison with diagonalization as in Russell's paradox.

Define

$s(x, y) = \mathsf{sub}(x, \bar{\mathsf{v}}, \mathsf{q}(y))$

Now $\neg\Box s(x, y)$ is a binary predicate.

The diagonal lemma may be provable without strong diagonalization.

Tarski obtained the diagonal lemma with concatenation only.

The language of Peano arithmetic lacks function symbols for sub and q and thus a functional expression for dia.

Whether we have weak or strong diagonalization can make a difference (Heck 2015, Schindler 2014).

At this point I could say more about self-reference.

The diagonal lemma may be provable without strong diagonalization.

Tarski obtained the diagonal lemma with concatenation only.

The language of Peano arithmetic lacks function symbols for sub and q and thus a functional expression for dia.

Whether we have weak or strong diagonalization can make a difference (Heck 2015, Schindler 2014).

At this point I could say more about self-reference.

The diagonal lemma may be provable without strong diagonalization.

Tarski obtained the diagonal lemma with concatenation only.

The language of Peano arithmetic lacks function symbols for sub and q and thus a functional expression for dia.

Whether we have weak or strong diagonalization can make a difference (Heck 2015, Schindler 2014).

At this point I could say more about self-reference.

The diagonal lemma may be provable without strong diagonalization.

Tarski obtained the diagonal lemma with concatenation only.

The language of Peano arithmetic lacks function symbols for sub and q and thus a functional expression for dia.

Whether we have weak or strong diagonalization can make a difference (Heck 2015, Schindler 2014).

At this point I could say more about self-reference.

The liar and other simple paradoxes

The theory E can contain more axioms beyond those explicitly stated.

Thus the following two claims are equivalent:

(i) φ is inconsistent with E.
(ii) E is inconsistent if it contains φ.

Instead of saying 'φ is inconsistent with E' I will often only say 'φ is inconsistent'.

I do not assume that E is consistent. It is consistent with the stated axioms.

Theorem (liar paradox)

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

Proof: Apply the diagonal lemma to $\neg\Box v$.

Theorem (tarski's theorem on the undefinability of truth)

If E *is consistent, there is no formula* $\tau(\mathsf{x})$ *such that* $\tau(\overline{\psi}) \leftrightarrow \psi$ *can be derived in* E *for all sentences* ψ *of* $\mathcal{L}$.

However, we can axiomatically add a new truth predicate.

McGee (1992) proved:

Theorem (McGee's theorem on T-sentences)

Assume $\mathsf{E} \nvdash \neg\varphi$. *Then there is a sentence* γ *such that* $\mathsf{E} + (\Box\overline{\gamma} \leftrightarrow \gamma)$ *is consistent and* $\mathsf{E} + (\Box\overline{\gamma} \leftrightarrow \gamma) \vdash \varphi$.

Proof: Apply the diagonal lemma to $\Box\mathsf{v} \leftrightarrow \varphi$.

This is a variant of Curry's paradox (Curry 1942). McGee used it against a 'solution' of the truth-theoretic paradoxes in (Horwich 1998).

Theorem (Montague's paradox, Montague 1963)

The schema $\Box\overline{\varphi} \to \varphi$ *is inconsistent with the rule*

$$\text{(NEC)} \qquad \frac{\varphi}{\Box\overline{\varphi}}$$

for all sentences φ.

Which modal notions are affected?

Theorem (Montague's paradox, Montague 1963)

The schema $\Box\overline{\varphi} \rightarrow \varphi$ is inconsistent with the rule

$$\text{(NEC)} \qquad \frac{\varphi}{\Box\overline{\varphi}}$$

for all sentences φ.

Which modal notions are affected?

Theorem (Gödel's first incompleteness theorem)

Suppose for all sentences φ of $\mathcal{L}$ that $\mathsf{E} \vdash \varphi$ iff $\mathsf{E} \vdash \Box\overline{\varphi}$. Then there is a sentence γ such that neither γ itself nor its negation is derivable in E, or E is inconsistent.

This is just extracted from (Gödel 1931).

Theorem

E *is inconsistent if it contains the schema* $\Box\overline{\Box\overline{\varphi}} \leftrightarrow \varphi$.

Proof.

For any for all sentences φ of $\mathcal{L}$ and indeed any expression φ we have the following:

$$\mathsf{E} \vdash \overline{\Box}{}^\frown \mathsf{q}(\overline{\varphi}) = \overline{\Box}{}^\frown \overline{\overline{\varphi}} \qquad \text{axiom A2}$$

$$\mathsf{E} \vdash \overline{\Box}{}^\frown \mathsf{q}(\overline{\varphi}) = \overline{\Box\overline{\varphi}} \qquad \text{axiom A1}$$

The diagonal lemma is applied to the formula $\neg\Box\big(\overline{\Box}{}^\frown \mathsf{q}(\mathsf{v})\big)$:

$$\mathsf{E} \vdash \gamma \leftrightarrow \neg\Box\big(\overline{\Box}{}^\frown \mathsf{q}(\overline{\gamma})\big)$$

$$\mathsf{E} \vdash \gamma \leftrightarrow \neg\Box\overline{\Box\overline{\gamma}} \qquad \text{remark above}$$

$$\mathsf{E} \vdash \gamma \leftrightarrow \neg\gamma \qquad \text{assumption}$$

∎

(K) $$\Box\overline{\varphi \to \psi} \to (\Box\overline{\varphi} \to \Box\overline{\psi}).$$

For our abbreviation of $\wedge$ we have:

Lemma

$\mathsf{E} \vdash \Box\overline{\varphi \wedge \psi} \leftrightarrow \Box\overline{\varphi} \wedge \Box\overline{\psi}$, if E contains K and is closed under NEC.

A theory $\mathcal{T}$ is internally inconsistent (with respect to $\Box$) if and only if $\mathcal{T} \vdash \Box\overline{\varphi}$ and $\mathcal{T} \vdash \Box\overline{\neg\varphi}$ for some sentence φ.

Lemma

Assume E *is closed under* NEC*. Then every internally inconsistent theory containing* E *and* K *proves* $\Box\overline{\psi}$ *for all sentences* ψ*.*

is the schema $\Box\overline{\varphi} \rightarrow \Box\overline{\Box\overline{\varphi}}$.

This is similar to (Thomason 1980):

Theorem (thomason 1980)

Assume E *is closed under* NEC. *Then any theory* $\mathcal{T}$ *containing* E *and the schemas* K, $\mathsf{4}$, *and* $\Box\overline{\Box\overline{\varphi}} \rightarrow \varphi$ *is internally inconsistent and proves* $\Box\overline{\psi}$ *for all* $\mathcal{L}$*-sentences* ψ.

Read the modal predicate as ‘S knows *x*’.

Proof:

Roughly speaking, we run the proof of Montague's paradox in the scope of $\Box$. Assume that E and $\mathcal{T}$ have the properties mentioned in the theorem.

$$
\begin{array}{rll}
\mathsf{E} \vdash & \gamma \leftrightarrow \neg\Box\overline{\gamma} & \text{liar sentence} \\
\mathsf{E} \vdash & \Box\overline{\Box\overline{\gamma} \to \neg\gamma} & \text{logic and } \mathsf{NEC} \\
\mathsf{E} \vdash & (\Box\overline{\gamma} \to \gamma) \to \big((\Box\overline{\gamma} \to \neg\gamma) \to \neg\Box\overline{\gamma}\big) & \text{logic} \\
\mathsf{E} \vdash & (\Box\overline{\gamma} \to \gamma) \to \big((\Box\overline{\gamma} \to \neg\gamma) \to \gamma\big) & \text{first line} \\
\mathsf{E} \vdash & \Box\overline{\Box\overline{\gamma} \to \gamma} \to \big(\Box\overline{\Box\overline{\gamma} \to \neg\gamma} \to \Box\overline{\gamma}\big) & \mathsf{NEC} \text{ and } \mathsf{K} \\
\mathcal{T} \vdash & \big(\Box\overline{\Box\overline{\gamma} \to \neg\gamma} \to \Box\overline{\gamma}\big) & \Box\overline{\Box\overline{\varphi} \to \varphi} \\
\mathcal{T} \vdash & \Box\overline{\gamma} & \text{second line}
\end{array}
$$

Now we invoke 4 to conclude $\mathcal{T} \vdash \Box\overline{\Box\overline{\gamma}}$ from $\mathcal{T} \vdash \Box\overline{\gamma}$. From the first line above we also get $\mathsf{E} \vdash \Box\overline{\gamma} \to \Box\overline{\neg\Box\overline{\gamma}}$ by NEC and K. Combining this with the last line, we obtain the following internal inconsistency:

$$\mathcal{T} \vdash \Box\overline{\neg\Box\overline{\gamma}} \wedge \Box\overline{\Box\overline{\gamma}}.$$

Since E and $\mathcal{T}$ satisfy the conditions of 18, we have $\mathcal{T} \vdash \Box\overline{\psi}$ for all sentences ψ.

THEOREM (LÖB'S THEOREM)

If E *is closed under* NEC *and contains* K *and 4, then we have* $\mathsf{E} \vdash \Box\overline{\Box\overline{\varphi} \to \varphi} \to \Box\overline{\varphi}$ *for all sentences* φ *of* $\mathcal{L}$.

Then there is the mysterium of the de Jongh–Sambin fixed-point theorem.

THEOREM (LÖB'S THEOREM)

If E *is closed under* NEC *and contains* K *and* 4, *then we have* $\mathsf{E} \vdash \Box\overline{\Box\overline{\varphi} \to \varphi} \to \Box\overline{\varphi}$ *for all sentences* φ *of* $\mathcal{L}$.

Then there is the mysterium of the de Jongh–Sambin fixed-point theorem.

Corollary

Assume that E *is closed under* NEC *and contains* K *and* 4*. Then for any* φ *of* $\mathcal{L}$ *the following rule of inference holds: If* $\mathsf{E} \vdash \Box\overline{\varphi} \to \varphi$*, then* $\mathsf{E} \vdash \varphi$*.*

Proof.

Assume $\mathsf{E} \vdash \Box\overline{\varphi} \to \varphi$ and reason as follows:

$\mathsf{E} \vdash \Box\overline{\varphi} \to \varphi$	assumption
$\mathsf{E} \vdash \Box\overline{\Box\overline{\varphi} \to \varphi}$	NEC
$\mathsf{E} \vdash \Box\overline{\varphi}$	theorem above
$\mathsf{E} \vdash \varphi$	assumption in first line

□

Theorem (Gödel's second incompleteness theorem)

Assume that E *is closed under* NEC *and contains* K *and* 4*. Then* $\mathsf{E} \vdash \neg\Box\bot$ *implies that* E *is inconsistent.*

Paradoxes from Interaction

Assume Tr is also a primitive unary predicate in the language.

THEOREM

Assume E *satisfies the following three conditions:*

(i) *If* ψ *is a sentence of* $\mathcal{L}$ *not containing* Tr, *then* E *contains* $\mathsf{Tr}\,\overline{\psi} \leftrightarrow \psi$.

(ii) *If* ψ *is a sentence of* $\mathcal{L}$ *not containing* $\Box$, *then* E *contains* $\Box\overline{\psi} \to \psi$.

(iii) *If* ψ *is a sentence of* $\mathcal{L}$ *not containing* $\Box$ *with* $\mathsf{E} \vdash \psi$, *then also* $\mathsf{E} \vdash \Box\overline{\psi}$ *holds.*

Then E *is inconsistent.*

PROOF.

We apply the diagonal lemma to the formula $\neg \mathsf{Tr}\big(\overline{\Box} \frown \mathsf{q}(x)\big)$ and reason as follows:

$\mathsf{E} \vdash \gamma \leftrightarrow \neg \mathsf{Tr}\big(\overline{\Box} \frown \mathsf{q}(\overline{\gamma})\big)$	diagonal lemma
$\mathsf{E} \vdash \gamma \leftrightarrow \neg \mathsf{Tr}\,\overline{\Box\overline{\gamma}}$	axioms A1 and A2; cf. proof of 16
$\mathsf{E} \vdash \mathsf{Tr}\,\overline{\Box\overline{\gamma}} \rightarrow \neg\gamma$	logic
$\mathsf{E} \vdash \Box\overline{\gamma} \rightarrow \neg\gamma$	(i)
$\mathsf{E} \vdash \Box\overline{\gamma} \rightarrow \gamma$	(ii)
$\mathsf{E} \vdash \neg\Box\overline{\gamma}$	two preceding lines
$\mathsf{E} \vdash \neg \mathsf{Tr}\,\overline{\Box\overline{\gamma}}$	(i)
$\mathsf{E} \vdash \gamma$	second line
$\mathsf{E} \vdash \Box\overline{\gamma}$	(iii)

The last line and the fourth line from the bottom establish the claim. □

Here is another application, which is not an inconsistency. See (Halbach and Horsten 2025). We assume that we have a predicate K for knowledge and a predicate JB for justified belief.

The knower sentence is a sentence γ with $\mathsf{E} \vdash \gamma \leftrightarrow \neg\mathsf{K}\overline{\gamma}$.

JTB CONCEPTION OF KNOWLEDGE

The following assumptions are jointly inconsistent:

schematic definition of knowledge: $\mathsf{K}\overline{\varphi} \leftrightarrow (\mathsf{JB}\overline{\varphi} \wedge \varphi)$

factivity of knowledge: $\mathsf{K}\overline{\varphi} \to \varphi$

Crucial assumption: From a proof of φ conclude $\mathsf{JB}\overline{\varphi}$

	$\mathsf{K}\overline{\gamma} \to \gamma$	factivity
	$\mathsf{K}\overline{\gamma} \to \neg\gamma$	knower
(1)	$\neg\mathsf{K}\overline{\gamma}$	two preceding lines
(2)	γ	knower
(3)	$\mathsf{JB}\overline{\gamma}$	crucial assumption
	$\mathsf{JB}\overline{\gamma} \wedge \gamma$	from (2) and (3)
(4)	$\mathsf{K}\overline{\gamma}$	def. of knowledge

JTB conception of knowledge

The following assumptions are jointly inconsistent:

schematic definition of knowledge: $\mathsf{K}\overline{\varphi} \leftrightarrow (\mathsf{JB}\overline{\varphi} \wedge \varphi)$

factivity of knowledge: $\mathsf{K}\overline{\varphi} \to \varphi$

Crucial assumption: From a proof of φ conclude $\mathsf{JB}\overline{\varphi}$

	$\mathsf{K}\overline{\gamma} \to \gamma$	factivity
	$\mathsf{K}\overline{\gamma} \to \neg\gamma$	knower
(1)	$\neg\mathsf{K}\overline{\gamma}$	two preceding lines
(2)	γ	knower
(3)	$\mathsf{JB}\overline{\gamma}$	crucial assumption
	$\mathsf{JB}\overline{\gamma} \wedge \gamma$	from (2) and (3)
(4)	$\mathsf{K}\overline{\gamma}$	def. of knowledge

JTB CONCEPTION OF KNOWLEDGE

The following assumptions are jointly inconsistent:

schematic definition of knowledge: $\mathsf{K}\overline{\varphi} \leftrightarrow (\mathsf{JB}\overline{\varphi} \wedge \varphi)$

factivity of knowledge: $\mathsf{K}\overline{\varphi} \rightarrow \varphi$

Crucial assumption: From a proof of φ conclude $\mathsf{JB}\overline{\varphi}$

	$\mathsf{K}\overline{\gamma} \rightarrow \gamma$	factivity
	$\mathsf{K}\overline{\gamma} \rightarrow \neg\gamma$	knower
(1)	$\neg\mathsf{K}\overline{\gamma}$	two preceding lines
(2)	γ	knower
(3)	$\mathsf{JB}\overline{\gamma}$	crucial assumption
	$\mathsf{JB}\overline{\gamma} \wedge \gamma$	from (2) and (3)
(4)	$\mathsf{K}\overline{\gamma}$	def. of knowledge

JTB CONCEPTION OF KNOWLEDGE

The following assumptions are jointly inconsistent:

schematic definition of knowledge: $\mathsf{K}\overline{\varphi} \leftrightarrow (\mathsf{JB}\overline{\varphi} \wedge \varphi)$

factivity of knowledge: $\mathsf{K}\overline{\varphi} \to \varphi$

Crucial assumption: From a proof of φ conclude $\mathsf{JB}\overline{\varphi}$

	$\mathsf{K}\overline{\gamma} \to \gamma$	factivity
	$\mathsf{K}\overline{\gamma} \to \neg\gamma$	knower
(1)	$\neg\mathsf{K}\overline{\gamma}$	two preceding lines
(2)	γ	knower
(3)	$\mathsf{JB}\overline{\gamma}$	crucial assumption
	$\mathsf{JB}\overline{\gamma} \wedge \gamma$	from (2) and (3)
(4)	$\mathsf{K}\overline{\gamma}$	def. of knowledge

JTB CONCEPTION OF KNOWLEDGE

The following assumptions are jointly inconsistent:

schematic definition of knowledge: $\mathsf{K}\overline{\varphi} \leftrightarrow (\mathsf{JB}\overline{\varphi} \land \varphi)$

factivity of knowledge: $\mathsf{K}\overline{\varphi} \rightarrow \varphi$

Crucial assumption: From a proof of φ conclude $\mathsf{JB}\overline{\varphi}$

	$\mathsf{K}\overline{\gamma} \rightarrow \gamma$	factivity
	$\mathsf{K}\overline{\gamma} \rightarrow \neg\gamma$	knower
(1)	$\neg\mathsf{K}\overline{\gamma}$	two preceding lines
(2)	γ	knower
(3)	$\mathsf{JB}\overline{\gamma}$	crucial assumption
	$\mathsf{JB}\overline{\gamma} \land \gamma$	from (2) and (3)
(4)	$\mathsf{K}\overline{\gamma}$	def. of knowledge

JTB CONCEPTION OF KNOWLEDGE

The following assumptions are jointly inconsistent:
schematic definition of knowledge: $\mathsf{K}\overline{\varphi} \leftrightarrow (\mathsf{JB}\overline{\varphi} \wedge \varphi)$
factivity of knowledge: $\mathsf{K}\overline{\varphi} \rightarrow \varphi$
Crucial assumption: From a proof of φ conclude $\mathsf{JB}\overline{\varphi}$

	$\mathsf{K}\overline{\gamma} \rightarrow \gamma$	factivity
	$\mathsf{K}\overline{\gamma} \rightarrow \neg\gamma$	knower
(1)	$\neg\mathsf{K}\overline{\gamma}$	two preceding lines
(2)	γ	knower
(3)	$\mathsf{JB}\overline{\gamma}$	crucial assumption
	$\mathsf{JB}\overline{\gamma} \wedge \gamma$	from (2) and (3)
(4)	$\mathsf{K}\overline{\gamma}$	def. of knowledge

JTB CONCEPTION OF KNOWLEDGE

The following assumptions are jointly inconsistent:

schematic definition of knowledge: $\mathsf{K}\overline{\varphi} \leftrightarrow (\mathsf{JB}\overline{\varphi} \wedge \varphi)$

factivity of knowledge: $\mathsf{K}\overline{\varphi} \rightarrow \varphi$

Crucial assumption: From a proof of φ conclude $\mathsf{JB}\overline{\varphi}$

	$\mathsf{K}\overline{\gamma} \rightarrow \gamma$	factivity
	$\mathsf{K}\overline{\gamma} \rightarrow \neg\gamma$	knower
(1)	$\neg\mathsf{K}\overline{\gamma}$	two preceding lines
(2)	γ	knower
(3)	$\mathsf{JB}\overline{\gamma}$	crucial assumption
	$\mathsf{JB}\overline{\gamma} \wedge \gamma$	from (2) and (3)
(4)	$\mathsf{K}\overline{\gamma}$	def. of knowledge

$\mathsf{G}x$ is 'x will always be the case', while $\mathsf{H}x$ is 'x has always been the case'. Defined: $\mathsf{F}x$ is 'x will be the case at some point (in the future)', and $\mathsf{P}x$, stands for 'x has been the case at some point (in the past)'.

(G1) $\mathsf{G}\overline{\varphi \to \psi} \to (\mathsf{G}\overline{\varphi} \to \mathsf{G}\overline{\psi})$,

(H1) $\mathsf{H}\overline{\varphi \to \psi} \to (\mathsf{H}\overline{\varphi} \to \mathsf{H}\overline{\psi})$,[1]

(G2) $\varphi \to \mathsf{H}\overline{\mathsf{F}\overline{\varphi}}$,

(H2) $\varphi \to \mathsf{G}\overline{\mathsf{P}\overline{\varphi}}$,

(G3) $\mathsf{G}\overline{\varphi} \leftrightarrow \neg\mathsf{F}\overline{\neg\varphi}$,

(H3) $\mathsf{H}\overline{\varphi} \leftrightarrow \neg\mathsf{P}\overline{\neg\varphi}$,

(N) $\dfrac{\varphi}{\mathsf{G}\overline{\varphi}}$ and $\dfrac{\varphi}{\mathsf{H}\overline{\varphi}}$ for all sentences φ.

[1] In the original paper (Horsten and Leitgeb 2001, p. 260), there is a typo in the formulation of this axiom: the occurrence of G there should be an H, too.

Theorem (no future paradox, Horsten and Leitgeb 2001)

If E *contains* G1, H1, G2, H2, G3, *and* H3 *and is closed under* N, *we have* $\mathsf{E} \vdash \mathsf{H}\overline{\bot} \wedge \mathsf{G}\overline{\bot}$.

Proof.

I shall only prove that there is no future, that is, $\mathsf{E} \vdash \mathsf{G}\overline{\bot}$. The first line is obtained as in the proof of 23:

(5)	$\mathsf{E} \vdash \gamma \leftrightarrow \mathsf{G}\overline{\mathsf{P}\overline{\neg\gamma}}$	
	$\mathsf{E} \vdash \neg\gamma \leftrightarrow \neg\mathsf{G}\overline{\mathsf{P}\overline{\neg\gamma}}$	
	$\mathsf{E} \vdash \neg\gamma \rightarrow \mathsf{G}\overline{\mathsf{P}\overline{\neg\gamma}}$	H2
(6)	$\mathsf{E} \vdash \gamma$	preceding two lines
(7)	$\mathsf{E} \vdash \mathsf{G}\overline{\mathsf{P}\overline{\neg\gamma}}$	from (5) and previous line
	$\mathsf{E} \vdash \mathsf{H}\overline{\gamma}$	N and (6)
	$\mathsf{E} \vdash \neg\mathsf{P}\overline{\neg\gamma}$	H3
(8)	$\mathsf{E} \vdash \mathsf{G}\overline{\neg\mathsf{P}\overline{\neg\gamma}}$	N
(9)	$\mathsf{E} \vdash \mathsf{G}\overline{\bot}$	(7), (8), and G1

The last line follows because, by N, we have $\mathsf{G}\overline{\varphi \rightarrow (\neg\varphi \rightarrow \bot)}$ for all φ and in particular for $\mathsf{P}\overline{\neg\gamma}$. □

Yablo and Visser

Kripke (1975) metioned something about illfounded hierarchies, but said later in (2019) that this was not about the Yablo–Visser paradox. Visser (1989) and Yablo (1985, 1993) presented similar paradoxes. See also (Cook 2014).

Some authors argued that these are paradoxes without self-reference.

Kripke (1975) metioned something about illfounded hierarchies, but said later in (2019) that this was not about the Yablo–Visser paradox. Visser (1989) and Yablo (1985, 1993) presented similar paradoxes. See also (Cook 2014).

Some authors argued that these are paradoxes without self-reference.

Yablo (1993) presented his paradox as an infinitely descending list of sentences.

(S$_1$) for all $k > 1$, S_k is untrue .

(S$_2$) for all $k > 2$, S_k is untrue .

(S$_3$) for all $k > 3$, S_k is untrue .

(S$_4$) for all $k > 4$, S_k is untrue .

$\vdots$

Yablo (1993) presented his paradox as an infinitely descending list of sentences. Visser (1989) presented his paradox in a more formal way, but it is Yablo's paradox with typed truth predicates:

(S_1)	for all $k > 1$, S_k is untrue_1.
(S_2)	for all $k > 2$, S_k is untrue_2.
(S_3)	for all $k > 3$, S_k is untrue_3.
(S_4)	for all $k > 4$, S_k is untrue_4.

$$\vdots$$

The use of infinite lists is dodgy.

The paradoxes can be formulated in syntax theory.

Before we can formulate the paradoxes, I need to explain how we can quantify into quotations.

I will present the paradox with very weak assumptions.

The use of infinite lists is dodgy.

The paradoxes can be formulated in syntax theory.

Before we can formulate the paradoxes, I need to explain how we can quantify into quotations.

I will present the paradox with very weak assumptions.

The use of infinite lists is dodgy.

The paradoxes can be formulated in syntax theory.

Before we can formulate the paradoxes, I need to explain how we can quantify into quotations.

I will present the paradox with very weak assumptions.

Assume $\Box$ is read as 'necessary' and we want to say that every expression is necessarily identical with itself, that is, we want to say that for all expressions e the sentence $\overline{e} = \overline{e}$ is necessary. We cannot do this by writing $\forall \mathsf{x} \Box \overline{\mathsf{x} = \mathsf{x}}$, but we can formulate our claim in the following way:

> *For all expressions e: if we replace in the formula* $\mathsf{x} = \mathsf{x}$ *every occurrence of* x *by the quotation constant for e, then the resulting sentence is necessary.*

Assume $\Box$ is read as 'necessary' and we want to say that every expression is necessarily identical with itself, that is, we want to say that for all expressions e the sentence $\overline{e} = \overline{e}$ is necessary. We cannot do this by writing $\forall \mathsf{x} \Box \overline{\mathsf{x} = \mathsf{x}}$, but we can formulate our claim in the following way:

> *For all expressions e: if we replace in the formula* $\mathsf{x} = \mathsf{x}$ *every occurrence of* x *by the quotation constant for e, then the resulting sentence is necessary.*

This can be expressed in $\mathcal{L}$ using the following formula:

$$\forall \mathsf{x}\, \Box \mathsf{sub}\big(\overline{\mathsf{x} = \mathsf{x}}, \bar{\mathsf{x}}, \mathsf{q}(\mathsf{x})\big).$$

From this we can derive, for instance, $\Box\overline{\overline{\neg} = \overline{\neg}}$ in E in the following way:

$\forall \mathsf{x}\, \Box \mathsf{sub}\big(\overline{\mathsf{x} = \mathsf{x}}, \bar{\mathsf{x}}, \mathsf{q}(\mathsf{x})\big)$	assumption
$\Box \mathsf{sub}\big(\overline{\mathsf{x} = \mathsf{x}}, \bar{\mathsf{x}}, \mathsf{q}(\overline{\neg})\big)$	logic
$\Box \mathsf{sub}\big(\overline{\mathsf{x} = \mathsf{x}}, \bar{\mathsf{x}}, \overline{\overline{\neg}}\big)$	A2
$\Box\overline{\overline{\neg} = \overline{\neg}}$	A3

Assume $\varphi(y)$ is a formula with no bound occurrences of the variable y, then

$$\overline{\varphi(\underset{\cdot}{x})} \text{ abbreviates } \mathsf{sub}\big(\overline{\varphi(y)}, \overline{y}, \mathsf{q}(x)\big).$$

In arithmetic the dot is placed above the variable; but we have already a bar there.

Of course, we can still generalize this and stipulate that only free occurrences of y are replaced; but we don’t need this.

The proof of the diagonal lemma yields the following:

DIAGONAL LEMMA WITH FREE VARIABLES

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

$$\mathsf{E} \vdash \forall \mathsf{y}\Big(\theta(\mathsf{y}) \leftrightarrow \varphi\big(\overline{\theta(\mathsf{y})}, \mathsf{y}\big)\Big).$$

We want a dot under the overlined y.

UNIFORM DIAGONAL LEMMA, PARAMETRIZED DIAGONAL LEMMA

Let $\varphi(\mathsf{x},\mathsf{y})$ be a formula with the two free variables x and y that does not contain a bound occurrence of y. Then there is a formula $\theta(\mathsf{y})$ such that

$$\mathsf{E} \vdash \forall \mathsf{y}\Big(\theta(\mathsf{y}) \leftrightarrow \varphi\big(\overline{\theta(\dot{\mathsf{y}})}, \mathsf{y}\big)\Big).$$

PROOF.

By applying the diagonal lemma above to the formula

$$\varphi\Big(\mathsf{sub}\big(\mathsf{x}, \overline{\mathsf{y}}, \mathsf{q}(\mathsf{y})\big), \mathsf{y}\Big),$$

I obtain a formula $\theta(\mathsf{y})$ such that the following holds:

$$\mathsf{E} \vdash \forall \mathsf{y}\Big(\theta(\mathsf{y}) \leftrightarrow \varphi\Big(\mathsf{sub}\big(\overline{\theta(\mathsf{y})}, \overline{\mathsf{y}}, \mathsf{q}(\mathsf{y})\big), \mathsf{y}\Big)\Big).$$

This is the claim, since $\overline{\theta(\dot{\mathsf{y}})}$ is defined as $\mathsf{sub}\big(\overline{\theta(\mathsf{y})}, \overline{\mathsf{y}}, \mathsf{q}(\mathsf{y})\big)$. □

I prove Yablo's paradox with weak assumptions.

The informal proof is based on an infinitely descending list of sentences with a top element. But much less is needed, as Ketland (2005) has shown.

I assume that the language $\mathcal{L}$ contains a binary predicate symbol $<$ that satisfies the following conditions:

(SER) $\forall x \exists y\, x<y,$

(TRANS) $\forall x \forall y \forall z \left(x<y \rightarrow (y<z \rightarrow x<z)\right).$

That's an order of the entire universe. We don't need it; but I don't want to relativize the quantifiers. We don't need a primitive symbol for this. You can think of $<$ as a defined formula that satisfies these conditions.

Using a primitive symbol shows that nothing beyond SER and TRANS is needed in addition to …

Of course, we need also a truth-theoretic assumption:

(UTS) $$\forall y \left(\Box \overline{\varphi(\dot{y})} \leftrightarrow \varphi(y) \right).$$

Theorem (Yablo's paradox)

Assume that E *contains all the following sentences:*

(SER) $\forall x \exists y\, x < y,$

(TRANS) $\forall x \forall y \forall z \big(x < y \rightarrow (y < z \rightarrow x < z)\big),$

(UTS) $\forall y \big(\Box \overline{\varphi(\dot{y})} \leftrightarrow \varphi(y)\big).$

Then E *is inconsistent.*

I understand the disappointment: UTS is obviously inconsistent.

Bear with me…

The point is the proof.

Theorem (yablo's paradox)

Assume that E *contains all the following sentences:*

(ser) $\forall x \exists y\, x{<}y,$

(trans) $\forall x \forall y \forall z \big(x{<}y \rightarrow (y{<}z \rightarrow x{<}z) \big),$

(uts) $\forall y \big(\Box \overline{\varphi(\dot{y})} \leftrightarrow \varphi(y) \big).$

Then E *is inconsistent.*

I understand the disappointment: uts is obviously inconsistent.

Bear with me…

The point is the proof.

PROOF.

I apply the diagonal lemma with a free variable to the formula

$$\forall z \succ y\ \neg\Box \mathsf{sub}\big(v, \overline{y}, q(z)\big),$$

which has exactly x and y as free variables. This yields a formula $\theta(y)$ with the following property:

$$\mathsf{E} \vdash \forall y \Big(\theta(y) \leftrightarrow \forall z \succ y\ \neg\Box \mathsf{sub}\big(\overline{\theta(y)}, \overline{y}, q(z)\big)\Big).$$

That is

(10) $$\mathsf{E} \vdash \forall y \big(\theta(y) \leftrightarrow \forall z \succ y\ \neg\Box \overline{\theta(\dot{z})}\big).$$

The sentences $\theta(\overline{e})$ for arbitrary expressions e correspond to the Yablo sentences. Priest (1997) defined the Yablo sentences in this way in arithmetic, after Visser (1989) had used a similar method to obtain his paradox. The contradiction can now be derived in the following way in E: □

PROOF CONTINUED

$$
\begin{aligned}
\mathsf{E} \vdash \forall y \big(\Box\overline{\theta(\underset{\cdot}{y})} &\leftrightarrow \theta(y)\big) && \text{UTS} \\
&\leftrightarrow \forall z{\succ}y\ \neg\Box\overline{\theta(\underset{\cdot}{z})} && (10) \\
&\rightarrow \exists z{\succ}y\ \neg\Box\overline{\theta(\underset{\cdot}{z})} && \text{SER} \\
&\rightarrow \exists z{\succ}y\ \forall w{\succ}z\ \neg\Box\overline{\theta(\underset{\cdot}{w})} && \text{second line and TRANS} \\
&\rightarrow \exists z{\succ}y\ \theta(z) && (10) \\
&\rightarrow \exists z{\succ}y\ \Box\overline{\theta(\underset{\cdot}{z})} && \text{UTS}
\end{aligned}
$$

From the second and last lines we get $\forall y\, \neg\Box\overline{\theta(\underset{\cdot}{y})}$.

$$
\begin{aligned}
&\mathsf{E} \vdash \forall y\, \neg\theta(y) && \text{UTS} \\
&\mathsf{E} \vdash \forall y\, \exists z{\succ}y\ \Box\overline{\theta(\underset{\cdot}{z})} && (10)
\end{aligned}
$$

The last line contradicts the previously derived $\forall y\, \neg\Box\overline{\theta(\underset{\cdot}{y})}$.

UTS is inconsistent by itself because of the liar paradox, but the same proof strategy can be used to prove that the following typed schema is sufficient.

(VUTS) $$\forall \mathsf{y}\left(\Box_{\mathsf{y}}\overline{\varphi(\dot{\mathsf{y}})} \leftrightarrow \varphi(\mathsf{y})\right),$$

where all terms occurring in index position in $\varphi(\mathsf{y})$ are variables x_1, x_2, …, x_k distinct from y (but not complex terms or constants) and all quantifiers $\forall x_1$, $\forall x_2$, …, $\forall x_k$ in $\varphi(\mathsf{y})$ are restricted by $x_i \succ \mathsf{y}$, respectively.

See the book for a proof. This schema is consistent by itself.

UTS is inconsistent by itself because of the liar paradox, but the same proof strategy can be used to prove that the following typed schema is sufficient.

(VUTS) $$\forall y \big(\Box_y \overline{\varphi(\dot{y})} \leftrightarrow \varphi(y) \big),$$

where all terms occurring in index position in $\varphi(y)$ are variables x_1, x_2, …, x_k distinct from y (but not complex terms or constants) and all quantifiers $\forall x_1$, $\forall x_2$, …, $\forall x_k$ in $\varphi(y)$ are restricted by $x_i \succ y$, respectively.

See the book for a proof. This schema is consistent by itself.

I get proper inconsistencies, not just ω-inconsistencies, because I quantify over indicies.

There are other versions of these paradoxes. Some are mere ω-inconsistencies.

methodology

What do these results tell us?

(i) We cannot have a predicate with the mentioned properties. We cannot combine predicates with certain mixed axioms.

(ii) We cannot define a predicate with certain properties, but could add a typed predicate axiomatically.

(iii) There is some complexity or definability hierarchies hidden.

(iv) There are more limitative results than just plain inconsistencies: internal inconsistencies, trivialities (e.g. predicates cannot apply to anything), ω-inconsistencies, and 'unintended' consequences of various kinds.

What do these results tell us?

(i) We cannot have a predicate with the mentioned properties. We cannot combine predicates with certain mixed axioms.
(ii) We cannot define a predicate with certain properties, but could add a typed predicate axiomatically.
(iii) There is some complexity or definability hierarchies hidden.
(iv) There are more limitative results than just plain inconsistencies: internal inconsistencies, trivialities (e.g. predicates cannot apply to anything), ω-inconsistencies, and 'unintended' consequences of various kinds.

What do these results tell us?

(i) We cannot have a predicate with the mentioned properties. We cannot combine predicates with certain mixed axioms.

(ii) We cannot define a predicate with certain properties, but could add a typed predicate axiomatically.

(iii) There is some complexity or definability hierarchies hidden.

(iv) There are more limitative results than just plain inconsistencies: internal inconsistencies, trivialities (e.g. predicates cannot apply to anything), ω-inconsistencies, and 'unintended' consequences of various kinds.

What do these results tell us?

(i) We cannot have a predicate with the mentioned properties. We cannot combine predicates with certain mixed axioms.
(ii) We cannot define a predicate with certain properties, but could add a typed predicate axiomatically.
(iii) There is some complexity or definability hierarchies hidden.
(iv) There are more limitative results than just plain inconsistencies: internal inconsistencies, trivialities (e.g. predicates cannot apply to anything), ω-inconsistencies, and 'unintended' consequences of various kinds.

The paradoxes that are not just plain inconsistencies teach us a lesson about alleged solutions:

A mere consistency proof doesn't tell us that there is no paradox.

The proposed strategy may still exclude the standard model, make the notion trivial, or generate off consequences.

The paradoxes that are not just plain inconsistencies teach us a lesson about alleged solutions:

A mere consistency proof doesn't tell us that there is no paradox.

The proposed strategy may still exclude the standard model, make the notion trivial, or generate off consequences.

What can go wrong?

- outright inconsistency
- ω-inconsistency or no ω-model
- triviality: The modality applies to all (or no) sentence.
- unacceptable consequences: The theory implies a contingent truth or decides a question it should not decide.
- and there thousand ways things can go wrong.

What can go wrong?

- outright inconsistency
- ω-inconsistency or no ω-model
- triviality: The modality applies to all (or no) sentence.
- unacceptable consequences: The theory implies a contingent truth or decides a question it should not decide.
- and there thousand ways things can go wrong.

What can go wrong?

- outright inconsistency
- ω-inconsistency or no ω-model
- triviality: The modality applies to all (or no) sentence.
- unacceptable consequences: The theory implies a contingent truth or decides a question it should not decide.
- and there thousand ways things can go wrong.

What can go wrong?

- outright inconsistency
- ω-inconsistency or no ω-model
- triviality: The modality applies to all (or no) sentence.
- unacceptable consequences: The theory implies a contingent truth or decides a question it should not decide.
- and there thousand ways things can go wrong.

What can go wrong?

outright inconsistency
ω-inconsistency or no ω-model
triviality: The modality applies to all (or no) sentence.
unacceptable consequences: The theory implies a contingent truth or decides a question it should not decide.
and there thousand ways things can go wrong.

McGee's ω-inconsistency

For the paradox I require a modicum of arithmetic.

We use strings $\mathsf{xxx}\ldots$ as natural numbers and call expressions $\overline{\mathsf{xxx}\ldots}$ *numerals*; they act as constants for numbers. Moreover, $\underline{n}$, the *numeral of n*, stands for

$$\underbrace{\overline{\mathsf{x}\ldots\mathsf{x}}}_{n}.$$

For instance, $\underline{4}$, that is, $\overline{\mathsf{xxxx}}$, is the numeral for 4.

Definition

$\mathsf{Nat}(x)$ is defined as $\mathsf{sub}(x, \overline{\mathsf{x}}, \underline{0}) = \underline{0}$.

Lemma

$\mathsf{E} \vdash \mathsf{Nat}(\underline{n})$ *for all natural numbers n.*

Definition

$\mathsf{Nat}(x)$ is defined as $\mathsf{sub}(x, \overline{\mathsf{x}}, \underline{0}) = \underline{0}$.

Lemma

$\mathsf{E} \vdash \mathsf{Nat}(\underline{n})$ *for all natural numbers n.*

I write $\forall n\, \varphi(n)$ for $\forall x\big(\mathsf{Nat}(x) \to \varphi(x)\big)$, for any variable x, and similarly $\exists n\, \varphi(n)$ for $\exists x\big(\mathsf{Nat}(x) \wedge \varphi(x)\big)$. This abbreviation is used in conjunction with the underdotting convention as in $\forall k\, \overline{\varphi(\underset{\cdot}{k})}$, which is short for

$$\forall x\Big(\mathsf{Nat}(x) \to \mathsf{sub}\big(\mathsf{q}x, \overline{x}, \overline{\varphi(x)}\big)\Big),$$

where x is a variable with no bound occurrences in φ.

Lemma

Assume $\underline{n}$, $\underline{k}$, $\underline{n+k}$, and $\underline{n \cdot k}$ are numerals for n, k, $n+k$, and $n \cdot k$, respectively. Then the following holds:

(i) $\mathsf{E} \vdash \underline{n}{}^\frown\underline{k} = \underline{n+k}$,

(ii) $\mathsf{E} \vdash \mathsf{sub}(\underline{k}, \bar{\mathsf{x}}, \underline{n}) = \underline{n \cdot k}$.

In particular we have $\mathsf{E} \vdash \underline{n}{}^\frown\underline{1} = \underline{n+1}$.

We expect that every number except $\underline{0}$ has a unique predecessor. However, this relies on the linearity of expressions, which is expressed by axiom A8.

Lemma

$\mathsf{E} \vdash \forall n \forall k (n ^\frown \underline{1} = k ^\frown \underline{1} \rightarrow n = k).$

Definition

E is *ω-inconsistent* if and only if there is a formula $\varphi(\mathsf{x})$ with the following properties:

(i) $\mathsf{E} \vdash \varphi(\underline{k})$ for all natural numbers k,
(ii) $\mathsf{E} \vdash \neg\forall n\, \varphi(n)$.

Lemma

$\mathsf{E} \vdash \forall x \left(\mathsf{Nat}(x) \to \mathsf{Nat}(x^\frown \underline{1})\right)$.

Lemma

The theory E *proves* $\forall n\, \varphi(n) \to \forall n\, \varphi(n^\frown \underline{1})$ *for all formulæ* $\varphi(x)$ *of* $\mathcal{L}$.

The sentence (S) says that at least one of the following sentences is not true:

(S)

(S) is true.

'(S) is true' is true.

' "(S) is true" is true' is true.

⋮

McGee's ω-inconsistency theorem can be obtained by formalizing (S).

Here is the theorem from (McGee 1985):

Theorem (mcgee's ω-inconsistency theorem)

Assume that E *is closed under the rule in* (i) *and contains the formula* (v) *and the schemas* (ii)–(iv) *for all sentences* φ *and* ψ *and all formulæ* $\chi(\mathrm{x})$ *having at most* x *free.*

(i) NEC,

(ii) $\Box\overline{\varphi \to \psi} \to (\Box\overline{\varphi} \to \Box\overline{\psi})$,

(iii) $\Box\overline{\neg\varphi} \to \neg\Box\overline{\varphi}$,

(iv) $\forall \mathrm{x}\, \Box\overline{\chi(\dot{\mathrm{x}})} \to \Box\overline{\forall \mathrm{x}\, \chi(\mathrm{x})}$,

(v) $\forall n\, (n = \underline{0} \lor \exists k\; n = k {}^\frown \underline{1}) \land \neg\exists k\; \underline{0} = k {}^\frown \underline{1}$.

Then E *is* ω*-inconsistent.*

I only sketch a proof.

bibliography

References

Roy Cook. *The Yablo Paradox: An Essay on Circularity*. Oxford University Press, 2014.

Haskell B. Curry. The inconsistency of certain formal logics. *Journal of Symbolic Logic*, 7:115–117, 1942.

Kurt Gödel. Über formal unentscheidbare sätze der principia mathematica und verwandter systeme I. *Monatshefte für Mathematik*, 38:173–198, 1931.

Volker Halbach and Leon Horsten. Self-referential gettier sentences. In Mattia Petrolo and Giorgio Venturi, editors, *Paradoxes Between Truth and Proof*, pages ??–?? 2025. to appear.

Richard G. Heck. Consistency and the theory of truth. *Review of Symbolic Logic*, 8:424–466, 2015.

Leon Horsten and Hannes Leitgeb. No future. *Journal of Philosophical Logic*, 30:259–265, 2001.

Paul Horwich. *Truth*. Oxford University Press, second edition, 1998.

Jeffrey Ketland. Deflationism and Gödel phenomena – reply to Tennant. *Mind*, 114:75–88, 2005.

Saul Kripke. Outline of a theory of truth. *Journal of Philosophy*, 72: 690–716, 1975.

Saul Kripke. Ungroundedness in Tarskian languages. *Journal of Philosophical Logic*, 48:603–609, 2019.

Vann McGee. How truthlike can a predicate be? A negative result. *Journal of Philosophical Logic*, 14:399–410, 1985.

Vann McGee. Maximal consistent sets of instances of Tarski's schema (T). *Journal of Philosophical Logic*, 21:235–241, 1992.

Richard Montague. Syntactical treatments of modality, with corollaries on reflexion principles and finite axiomatizability. *Acta Philosophica Fennica*, 16:153–67, 1963. Reprinted in Montague 1974, 286–302.

Richard Montague. *Formal Philosophy: Selected Papers of Richard Montague*. Yale University Press, New Haven and London, 1974. Edited and with an introduction by Richmond H. Thomason.

Graham Priest. Yablo's paradox. *Analysis*, 57:236–242, 1997.

Thomas Schindler. *Type-free truth*. PhD thesis, Ludwig-Maximilians-Universität München, 2014.

Richmond H. Thomason. A note on syntactical treatments of modality. *Synthese*, 44:391–396, 1980.

Albert Visser. Semantics and the liar paradox. In Dov Gabbay and Franz Guenthner, editors, *Handbook of Philosophical Logic*, volume 4, pages 617–706. Reidel, Dordrecht, 1989.

Stephen Yablo. Truth and reflection. *Journal of Philosophical Logic*, 14: 297–349, 1985.

Stephen Yablo. Paradox without self-reference. *Analysis*, 53:251–252, 1993.