Truth & Paradox

Volker Halbach

Nordic Logic Summer School 2017

Why formal theories of truth?

- Solutions to the liar paradox and other paradoxes

Why formal theories of truth?

- Solutions to the liar paradox and other paradoxes
- Truth and consequence: What can one actually do with a truth predicate conceived, e.g., as a device of disquotation?

Why formal theories of truth?

- Solutions to the liar paradox and other paradoxes
- Truth and consequence: What can one actually do with a truth predicate conceived, e.g., as a device of disquotation?
- Truth in other areas of philosophy, e.g., in the definition of knowledge

Why formal theories of truth?

- Solutions to the liar paradox and other paradoxes
- Truth and consequence: What can one actually do with a truth predicate conceived, e.g., as a device of disquotation?
- Truth in other areas of philosophy, e.g., in the definition of knowledge
- In particular, is truth a substantial notion that allows one new insights into non-semantic issues?

Why formal theories of truth?

- Solutions to the liar paradox and other paradoxes
- Truth and consequence: What can one actually do with a truth predicate conceived, e.g., as a device of disquotation?
- Truth in other areas of philosophy, e.g., in the definition of knowledge
- In particular, is truth a substantial notion that allows one new insights into non-semantic issues?
- Is truth always founded in non-semantic facts?

Why formal theories of truth?

- Solutions to the liar paradox and other paradoxes
- Truth and consequence: What can one actually do with a truth predicate conceived, e.g., as a device of disquotation?
- Truth in other areas of philosophy, e.g., in the definition of knowledge
- In particular, is truth a substantial notion that allows one new insights into non-semantic issues?
- Is truth always founded in non-semantic facts?
- Reduction: parts of mathematics can be reduced to a theory of truth; the proof-theoretic strength of truth theories

Why formal theories of truth?

- Solutions to the liar paradox and other paradoxes
- Truth and consequence: What can one actually do with a truth predicate conceived, e.g., as a device of disquotation?
- Truth in other areas of philosophy, e.g., in the definition of knowledge
- In particular, is truth a substantial notion that allows one new insights into non-semantic issues?
- Is truth always founded in non-semantic facts?
- Reduction: parts of mathematics can be reduced to a theory of truth; the proof-theoretic strength of truth theories
- Truth and reflection: truth as a device for stating very strong reflection principles

Philosophical decisions

Truth may be taken to be applying to any kind of objects (sentences as types or tokens, propositions…) as long as these objects have the structure of sentence types.

Philosophical decisions

Truth may be taken to be applying to any kind of objects (sentences as types or tokens, propositions…) as long as these objects have the structure of sentence types.

I will take truth to apply to sentences of a fixed formal language only, but not to ‘foreign’ sentences.

Philosophical decisions

Truth may be taken to be applying to any kind of objects (sentences as types or tokens, propositions…) as long as these objects have the structure of sentence types.

I will take truth to apply to sentences of a fixed formal language only, but not to ‘foreign’ sentences.

I’ll emphasize the axiomatic approach – at least in the beginning. This is not to say that truth cannot be defined in terms of correspondence etc.

Truth & Paradox

I · The Theory of Syntax and Diagonalisation

Volker Halbach

Nordic Logic Summer School 2017

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.

Self-reference

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.

Arithmetic

The approach via arithmetic is indirect. Arithmetic talks about numbers, not about sentences. Coding sentences and expressions by numbers allows one to talk about the numerical codes of sentences and therefore arithmetic is indirectly about sentences.

Arithmetic

The approach via arithmetic is indirect. Arithmetic talks about numbers, not about sentences. Coding sentences and expressions by numbers allows one to talk about the numerical codes of sentences and therefore arithmetic is indirectly about sentences.

My approach here avoids this detour via numbers. I present a theory of expressions that is given by some (as I hope) obvious axioms on expressions. The trick (diagonalization) that is then used for obtaining a self-referential sentence is the same as in the case of arithmetic.

The alphabet

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 v, v_1, v_2, v_3,...
2. predicate symbols = and T,
3. function symbols q, $\hat{}$ 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.

All the mentioned symbols are pairwise different.

Notational conventions

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, respectively. Moreover, it is always presupposed variable clashes are avoided by renaming variables in a suitable way.

Notational conventions

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, respectively. 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.

Notational conventions

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, respectively. 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.

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.

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.

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.
3. If t, r and s are terms, then $\mathrm{q}(t)$, $(s \hat{} t)$, $\mathrm{sub}(r, s, t)$ are terms, and similarly for all further function symbols

The empty string

Since the empty string is a string of symbols $\bar{}$ is a term. Since $\bar{}$ looks so odd, I shall write $\underline{0}$ for $\bar{}$. From an ontologically point of view the empty string is a weird thing. One might be inclined to say that it is not anything. I have only a pragmatic excuse for assuming the empty string: it is useful, though not indispensable.

The empty string

Since the empty string is a string of symbols $\bar{}$ is a term. Since $\bar{}$ looks so odd, I shall write $\underline{0}$ for $\bar{}$. From an ontologically point of view the empty string is a weird thing. One might be inclined to say that it is not anything. I have only a pragmatic excuse for assuming the empty string: it is useful, though not indispensable.

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

Formulæ

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

Example

1. $\forall \mathrm{v}_3 (\mathrm{v}_3 = \overline{\wedge \forall} \wedge T \overline{\mathrm{v}_3})$ is a sentence.
2. $\overline{\mathrm{v}_{12}} = \overline{\neg T \neg}$ is a sentence, i.e., the formula does not feature a free variable.

A theory of expressions

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

A theory of expressions

The theory $\mathcal{A}$ which will be described in this section is designed in order to obtain smooth proofs. I do not aim 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.

The axioms

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 $\overline{a}\,\hat{}\,\overline{b} = \overline{ab}$, 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.

The axioms

Probably we won't need the following axioms:

Definition (additional axioms)

A5 $\forall x \forall y \forall z((x \hat{} y) \hat{} z) = (x \hat{} (y \hat{} z))$

The axioms

Probably we won't need the following axioms:

Definition (additional axioms)

A5 $\forall x \forall y \forall z((x \hat{} y) \hat{} z) = (x \hat{} (y \hat{} z))$

A6 $\forall x \forall y(x \hat{} y = \underline{0} \rightarrow x = \underline{0} \wedge y = \underline{0})$

The axioms

Probably we won't need the following axioms:

Definition (additional axioms)

A5 $\forall x \forall y \forall z((x \hat{} y) \hat{} z) = (x \hat{} (y \hat{} z))$

A6 $\forall x \forall y(x \hat{} y = \underline{0} \to x = \underline{0} \wedge y = \underline{0})$

A7 $\forall x \forall y(x \hat{} y = x \leftrightarrow y = \underline{0}) \wedge \forall x \forall y(y \hat{} x = x \leftrightarrow y = \underline{0})$

The axioms

Probably we won't need the following axioms:

Definition (additional axioms)

A5 $\forall x \forall y \forall z((x \hat{} y) \hat{} z) = (x \hat{} (y \hat{} z))$

A6 $\forall x \forall y(x \hat{} y = \underline{0} \to x = \underline{0} \wedge y = \underline{0})$

A7 $\forall x \forall y(x \hat{} y = x \leftrightarrow y = \underline{0}) \wedge \forall x \forall y(y \hat{} x = x \leftrightarrow y = \underline{0})$

A8 $\forall x_1 \forall x_2 \forall y \forall z(\mathrm{sub}(x_1, y, z) \hat{} \mathrm{sub}(x_2, y, z)) = \mathrm{sub}(x_1 \hat{} x_2, y, z)$

The axioms

Probably we won't need the following axioms:

Definition (additional axioms)

A5 $\forall x \forall y \forall z((x \hat{} y) \hat{} z) = (x \hat{} (y \hat{} z))$

A6 $\forall x \forall y(x \hat{} y = \underline{0} \to x = \underline{0} \wedge y = \underline{0})$

A7 $\forall x \forall y(x \hat{} y = x \leftrightarrow y = \underline{0}) \wedge \forall x \forall y(y \hat{} x = x \leftrightarrow y = \underline{0})$

A8 $\forall x_1 \forall x_2 \forall y \forall z(\mathrm{sub}(x_1, y, z) \hat{} \mathrm{sub}(x_2, y, z)) = \mathrm{sub}(x_1 \hat{} x_2, y, z)$

A9 $\neg \overline{a} = \overline{b}$, if a and b are distinct expressions.

Comments

A1-A4 describe the functions of concatenation, quotation and substitution by providing function values for specific entries. From these axioms one cannot derive (non-trivial) universally quantified principles and therefore axioms like the associative law for $\hat{}$ A5 are not derivable from A1–A4.

Comments

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.

Comments

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

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

Comments

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

Therefore $\overline{\neg\neg\mathrm{v}} = \overline{\neg}\,\widehat{}\,\overline{\neg\mathrm{v}}$ is an instance of A2 as well as $\overline{\neg\neg\mathrm{v}} = \overline{\neg\neg}\,\widehat{}\,\overline{\mathrm{v}}$.

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

Comments

An instance of A3 is the sentence $q\overline{\overline{v}\neg} = \overline{\overline{\overline{v}\neg}}$. Thus q describes the function that takes an expression and returns its quotation constant.

Comments

In A4 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.

Comments

In A4 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. Therefore I do not 'solve' the problem but avoid it by the restriction of b to a single symbol.

Comments

A1-A4 are already sufficient for proving the diagonalization Theorem 12.

Comments

A1-A4 are already sufficient for proving the diagonalization Theorem 12.

A5 simplifies the reasoning with strings a great deal. Since $\mathcal{A} \vdash (x^\frown y)^\frown z = x^\frown(y^\frown z)$, that is, $^\frown$ is associative by A5, 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 $^\frown$.

Comments

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

Comments

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

Example

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

Comments

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

Example

$\mathcal{A} \vdash \mathrm{sub}(\overline{\neg\neg}, \overline{\neg}, \overline{\neg\neg\neg}) = \overline{\neg\neg\neg\neg\neg\neg}$

Example

$\mathcal{A} \vdash \mathrm{sub}(\overline{\mathrm{v} = \mathrm{v} \wedge \overline{\mathrm{v}} = \overline{\mathrm{v}}}, \overline{\mathrm{v}}, \overline{\mathrm{v}_2}) = \overline{\mathrm{v}_2 = \mathrm{v}_2 \wedge \overline{\mathrm{v}} = \overline{\mathrm{v}}}$

Comments

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 arihmetical 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

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

Diagonalization

The diagonalization function dia is defined in the following way:

Definition

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

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 $\mathcal{A}$. 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.

A lemma

Lemma

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

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

A lemma

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::

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

⊣

The diagonal lemma

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:*

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

The diagonal lemma

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:

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

$\dashv$

Truth & Paradox

II · The T-Sentences

Volker Halbach

Nordic Logic Summer School 2017

Inconsistency

I shall prove the inconsistency of some theories with the theory $\mathcal{A}$.
'inconsistent' always means 'inconsistent with $\mathcal{A}$'.

Inconsistency

I shall prove the inconsistency of some theories with the theory $\mathcal{A}$. 'inconsistent' always means 'inconsistent with $\mathcal{A}$'.

Since I did not fix the axioms of $\mathcal{A}$ and admitted further axioms in $\mathcal{A}$, inconsistency results can be formulated in two ways. One can either say '$\mathcal{A}$ is inconsistent if it contains the sentence ψ' or one says 'ψ is inconsistent with $\mathcal{A}$'.

The T-scheme

The first inconsistency result is the famous liar paradox. It is plausible to assume that a truth predicate T for the language $\mathcal{L}$ satisfies the T-scheme

(3) $$T\overline{\psi} \leftrightarrow \psi$$

for all sentences ψ of $\mathcal{L}$. This scheme corresponds to the scheme

> *'A' is true if and only if A,*

where A is any English declarative sentence.

The liar in $\mathcal{A}$

Theorem (liar paradox)

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

The liar in $\mathcal{A}$

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 12 to the formula $\neg T\mathrm{v}$. Then theorem 12 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'. ⊣

Tarski's theorem

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.

Tarski's theorem

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 12 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. $\dashv$

The scope of Tarski's theorem

It is not so much surprising that the axioms listed explicitly in Definition 4 do not allow for a definition of such truth predicate $\tau(\mathrm{v})$. According to Definition 4, however, $\mathcal{A}$ may contain arbitrary additional axioms. Thus Tarski's Theorem says that adding axioms to $\mathcal{A}$ that allow for a truth definition renders $\mathcal{A}$ inconsistent.

Extending the language

Nevertheless one can add a new predicate symbol *which is not in* $\mathcal{L}$, and add True $\overline{\psi} \leftrightarrow \psi$ as an axiom scheme for all sentences of $\mathcal{L}$. In this case φ cannot contain the symbol True and the diagonalization theorem 12 does not apply to True v because it applies only to formulæ $\varphi(\mathrm{v})$ of $\mathcal{L}$.

Extending the language

Nevertheless one can add a new predicate symbol *which is not in* $\mathcal{L}$, and add $\mathrm{True}\,\overline{\psi} \leftrightarrow \psi$ as an axiom scheme for all sentences of $\mathcal{L}$. In this case φ cannot contain the symbol True and the diagonalization theorem 12 does not apply to $\mathrm{True}\,\mathrm{v}$ because it applies only to formulæ $\varphi(\mathrm{v})$ of $\mathcal{L}$.

Theorem

Assume that the language $\mathcal{L}$ *is expanded by a new predicate symbol* True *and all sentences* $\mathrm{True}\,\overline{\psi} \leftrightarrow \psi$ *(for* ψ *a sentence of* $\mathcal{L}$*) are added to* $\mathcal{A}$*. The resulting theory is consistent if* $\mathcal{A}$ *is consistent.*

The theory of disquotation

Call the theory $\mathcal{A}$ plus all these equivalences TB. Thus TB is given by the following set of axioms:

$$\mathcal{A} \cup \{\text{True}\,\overline{\psi} \leftrightarrow \psi : \psi \text{ a sentence of } \mathcal{L}\}$$

The proof

The idea for the proof is due to Tarski.

The proof

The idea for the proof is due to Tarski.

I shall show that a given proof of a contradiction $\perp$ in the theory TB can be transformed into a proof of $\perp$ in $\mathcal{A}$. In the given proof only finitely many axioms with True can occur; let

$$\text{True}\,\overline{\psi_0} \leftrightarrow \psi_0, \text{True}\,\overline{\psi_1} \leftrightarrow \psi_1, \ldots \text{True}\,\overline{\psi_n} \leftrightarrow \psi_n$$

be these axioms. $\tau(\mathrm{v})$ is the following formula of the language $\mathcal{L}$:

$$(v = \overline{\psi_0} \wedge \psi_0) \vee (v = \overline{\psi_1} \wedge \psi_1) \vee \ldots (v = \overline{\psi_n} \wedge \psi_n)$$

The proof

The idea for the proof is due to Tarski.

I shall show that a given proof of a contradiction $\perp$ in the theory TB can be transformed into a proof of $\perp$ in $\mathcal{A}$. In the given proof only finitely many axioms with True can occur; let

$$\text{True}\,\overline{\psi_0} \leftrightarrow \psi_0, \text{True}\,\overline{\psi_1} \leftrightarrow \psi_1, \ldots \text{True}\,\overline{\psi_n} \leftrightarrow \psi_n$$

be these axioms. $\tau(\mathrm{v})$ is the following formula of the language $\mathcal{L}$:

$$(v = \overline{\psi_0} \wedge \psi_0) \vee (v = \overline{\psi_1} \wedge \psi_1) \vee \ldots (v = \overline{\psi_n} \wedge \psi_n)$$

Obviously one has

$$\tau(\overline{\psi_0}) \leftrightarrow \psi_0$$

and similarly for all ψ_k $(k \leq n)$.

The proof

The idea for the proof is due to Tarski.

I shall show that a given proof of a contradiction $\perp$ in the theory TB can be transformed into a proof of $\perp$ in $\mathcal{A}$. In the given proof only finitely many axioms with True can occur; let

$$\text{True}\,\overline{\psi_0} \leftrightarrow \psi_0, \text{True}\,\overline{\psi_1} \leftrightarrow \psi_1, \ldots \text{True}\,\overline{\psi_n} \leftrightarrow \psi_n$$

be these axioms. $\tau(\mathrm{v})$ is the following formula of the language $\mathcal{L}$:

$$(v = \overline{\psi_0} \wedge \psi_0) \vee (v = \overline{\psi_1} \wedge \psi_1) \vee \ldots (v = \overline{\psi_n} \wedge \psi_n)$$

Obviously one has

$$\tau(\overline{\psi_0}) \leftrightarrow \psi_0$$

and similarly for all ψ_k $(k \leq n)$.

Now replace everywhere in the given proof any formula True t, where t is any arbitrary term, by $\tau(t)$ and add above any former axiom True $\overline{\psi_k} \leftrightarrow \psi_k$ a proof of $\tau(\overline{\psi_k}) \leftrightarrow \psi_k$, respectively. The resulting structure is a proof in $\mathcal{A}$ of the contradiction $\perp$.

Conservativity

The proof establishes a stronger result: Adding the T-sentences

$$\text{True } \overline{\psi} \leftrightarrow \psi$$

(ψ a sentence without True) to $\mathcal{A}$ yields a conservative extension of $\mathcal{A}$:

Conservativity

The proof establishes a stronger result: Adding the T-sentences

$$\text{True } \overline{\psi} \leftrightarrow \psi$$

(ψ a sentence without True) to $\mathcal{A}$ yields a conservative extension of $\mathcal{A}$:

Theorem

TB is conservative over $\mathcal{A}$. That is, If φ is a sentence without True *that is provable in TB, then φ is already provable in $\mathcal{A}$ only.*

Conservativity

The proof establishes a stronger result: Adding the T-sentences

$$\text{True } \overline{\psi} \leftrightarrow \psi$$

(ψ a sentence without True) to $\mathcal{A}$ yields a conservative extension of $\mathcal{A}$:

Theorem

TB is conservative over $\mathcal{A}$. That is, If φ is a sentence without True *that is provable in TB, then φ is already provable in $\mathcal{A}$ only.*

Proof.

Just replace $\bot$ by φ in the proof above. $\dashv$

Conservativity

The proof establishes a stronger result: Adding the T-sentences

$$\mathrm{True}\,\overline{\psi} \leftrightarrow \psi$$

(ψ a sentence without True) to $\mathcal{A}$ yields a conservative extension of $\mathcal{A}$:

Theorem

TB is conservative over $\mathcal{A}$. That is, If φ is a sentence without True *that is provable in TB, then φ is already provable in $\mathcal{A}$ only.*

Proof.

Just replace $\bot$ by φ in the proof above. ⊣

The proof shows that these T-sentences do not allow to prove any new 'substantial' insights. Works also with full induction.

Conservativity over logic

The T-sentences are not conservative over pure *logic*. The T-sentences prove that there are at least two different objects:

$\mathrm{True}\ \overline{\overline{\forall} = \overline{\forall}} \leftrightarrow \overline{\forall} = \overline{\forall}$	T-sentence
$\overline{\forall} = \overline{\forall}$	tautology
$\mathrm{True}\ \overline{\overline{\forall} = \overline{\forall}}$	two preceding lines
$\mathrm{True}\ \overline{\neg\overline{\forall} = \overline{\forall}} \leftrightarrow \neg\overline{\forall} = \overline{\forall}$	T-sentence
$\neg\mathrm{True}\ \overline{\neg\overline{\forall} = \overline{\forall}}$	
$\overline{\overline{\forall} = \overline{\forall}} \neq \overline{\neg\overline{\forall} = \overline{\forall}}$	

The technique can also be shown to show that the following sentences are not provable in TB:

$$\forall x(\mathrm{Sent}(x) \to (\mathrm{True}\, x \vee \mathrm{True}\, \dot{\neg} x)) \text{ or}$$
$$\forall x \forall y\big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \to (\mathrm{True}\, (x \dot{\wedge} y) \leftrightarrow (\mathrm{True}\, x \wedge \mathrm{True}\, y))\big)$$

This is just a remark here, as we do not yet have the notation.

Montague's paradox

Theorem (Montague's paradox Montague 1963)

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

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

Montague's paradox

Theorem (Montague's paradox Montague 1963)

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.

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

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

⊣

Montague's paradox

Theorem (Montague's paradox Montague 1963)

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
$T\overline{\gamma} \leftrightarrow \neg\gamma$	
$T\overline{\gamma} \to \gamma$	axiom

⊣

Montague's paradox

Theorem (Montague's paradox Montague 1963)

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
$T\overline{\gamma} \leftrightarrow \neg\gamma$	
$T\overline{\gamma} \to \gamma$	axiom
$\neg T\overline{\gamma}$	logic

⊣

Montague's paradox

Theorem (Montague's paradox Montague 1963)

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
$T\overline{\gamma} \leftrightarrow \neg\gamma$	
$T\overline{\gamma} \to \gamma$	axiom
$\neg T\overline{\gamma}$	logic
γ	first line

⊣

Montague's paradox

Theorem (Montague's paradox Montague 1963)

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
$T\overline{\gamma} \leftrightarrow \neg\gamma$	
$T\overline{\gamma} \to \gamma$	axiom
$\neg T\overline{\gamma}$	logic
γ	first line
$T\overline{\gamma}$	NEC

⊣

Montague's theorem is a problem not only for truth, but also for necessity and knowledge (read $T\overline{\varphi}$ as 'It is necessary that φ' or 'Volker knows that φ').

Generally, a language with the modal operator $\Box$ (for necessity, knowledge, obligation etc.) can be translated into a language with a corresponding predicate, but no tthe other way round. In particular, how would we translate $\forall x\,(\varphi(x) \to Tx)$ or $\exists x\,(\varphi(x) \wedge Tx)$? There quantified statements, however, seem important in philosophy: 'There are synthetic a priori judgements', 'There are necessary a postieriori truths'. 'All truths are verifiable', tripartite definition of knowledge,…

Another formulation of the T-sentences

Often the T-sentences are stated in the following way:

$$T\overline{\psi} \leftrightarrow \psi$$

where ψ *must not contain* T. It's thought that this is safe. But I don't trust that formulation anymore.

How not to state the T-sentences

Theorem

Assume $\mathcal{L}$ contains a unary predicate symbol N (for necessity of some kind, let's say), and assume further:

- T $T\overline{\varphi} \leftrightarrow \varphi$ *for all sentences φ of $\mathcal{L}$ not containing T.*
- N1 $N\overline{\varphi} \rightarrow \varphi$ *for all sentences φ of $\mathcal{L}$ not containing N.*
- N2 *Whenever $\mathcal{A} \vdash \varphi$, then also $\mathcal{A} \vdash N\overline{\varphi}$ for all sentences φ of $\mathcal{L}$ not containing N.*

Then $\mathcal{A}$ is inconsistent.

Proof

We apply diagonalisation to $\neg T(\overline{N}\hat{}\mathrm{q}x)$.

$\gamma \leftrightarrow \neg T\overline{N\overline{\gamma}}$	diagonalisation
$T\overline{N\overline{\gamma}} \leftrightarrow \neg\gamma$	
$N\overline{\gamma} \rightarrow \neg\gamma$	
$N\overline{\gamma} \rightarrow \gamma$	(N1)
$\neg N\overline{\gamma}$	two previous lines
$\neg T\overline{N\overline{\gamma}}$	T
γ	first and last line
$N\overline{\gamma}$	(N2)

How not to state the T-sentences

Usually it is thought that typing is a remedy to the paradoxes. The example shows that this works only as long as typing is not applied to more than one predicate.

How not to state the T-sentences

Usually it is thought that typing is a remedy to the paradoxes. The example shows that this works only as long as typing is not applied to more than one predicate.

The result is the first of various paradoxes (*vulgo* inconsistencies) that arise from the interaction of predicates.

Summary

- Adding a new truth predicate to $\mathcal{A}$ and axiomatising it by typed T-sentences yields a conservative extension of $\mathcal{A}$.

Summary

- Adding a new truth predicate to $\mathcal{A}$ and axiomatising it by typed T-sentences yields a conservative extension of $\mathcal{A}$.
- The resulting theory TB does not prove generalisation such as

$$\forall x(\mathrm{Sent}(x) \to (\mathrm{True}\, x \vee \mathrm{True}\, \dot{\neg} x)) \text{ or}$$
$$\forall x \forall y\big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \to (\mathrm{True}\, (x \dot{\wedge} y) \leftrightarrow (\mathrm{True}\, x \wedge \mathrm{True}\, y))\big)$$

Summary

- Adding a new truth predicate to $\mathcal{A}$ and axiomatising it by typed T-sentences yields a conservative extension of $\mathcal{A}$.
- The resulting theory TB does not prove generalisation such as

$$\forall x(\mathrm{Sent}(x) \to (\mathrm{True}\, x \vee \mathrm{True}\, \dot{\neg} x)) \text{ or}$$
$$\forall x \forall y\big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \to (\mathrm{True}\, (x \dot{\wedge} y) \leftrightarrow (\mathrm{True}\, x \wedge \mathrm{True}\, y))\big)$$

- TB is not finitely axiomatisable.

Summary

- Adding a new truth predicate to $\mathcal{A}$ and axiomatising it by typed T-sentences yields a conservative extension of $\mathcal{A}$.
- The resulting theory TB does not prove generalisation such as

$$\forall x(\mathrm{Sent}(x) \to (\mathrm{True}\, x \vee \mathrm{True}\, \dot{\neg} x)) \text{ or}$$
$$\forall x \forall y\big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \to (\mathrm{True}\, (x \dot{\wedge} y) \leftrightarrow (\mathrm{True}\, x \wedge \mathrm{True}\, y))\big)$$

- TB is not finitely axiomatisable.
- According to Tarski, a decent theory of truth should not only yield the T-sentences (and satisfy Convention T), but also prove those generalisations.

Summary

- Adding a new truth predicate to $\mathcal{A}$ and axiomatising it by typed T-sentences yields a conservative extension of $\mathcal{A}$.
- The resulting theory TB does not prove generalisation such as

$$\forall x(\mathrm{Sent}(x) \to (\mathrm{True}\, x \vee \mathrm{True}\, \dot{\neg} x)) \text{ or}$$
$$\forall x \forall y\big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \to (\mathrm{True}\, (x \dot{\wedge} y) \leftrightarrow (\mathrm{True}\, x \wedge \mathrm{True}\, y))\big)$$

- TB is not finitely axiomatisable.
- According to Tarski, a decent theory of truth should not only yield the T-sentences (and satisfy Convention T), but also prove those generalisations.
- 'Mixing' the T-sentences with axiomatisations of other notions such as necessity can lead to inconsistencies. So type restrictions don't solve all problems.

Liberalising the type restriction

There have been various proposals to lift the type restrictions on the T-sentences.

Liberalising the type restriction

There have been various proposals to lift the type restrictions on the T-sentences.

Motives:

- Eg the following T-sentence looks ok:

  "'Grass is red' is not true' is true iff 'Grass is red' is not true.

- A more liberal approach might help to regain deductive power.

Liberalising the type restriction

There have been various proposals to lift the type restrictions on the T-sentences.

Motives:

- Eg the following T-sentence looks ok:

  "'Grass is red' is not true' is true iff 'Grass is red' is not true.

- A more liberal approach might help to regain deductive power.

However, one seems to be caught between Scylla and Charybdis: the typed truth predicate of TB is too weak, while the full unrestricted T-schema is too strong.

Liberalising the type restriction

There have been various proposals to lift the type restrictions on the T-sentences.

Motives:

- Eg the following T-sentence looks ok:

  "Grass is red' is not true' is true iff 'Grass is red' is not true.

- A more liberal approach might help to regain deductive power.

However, one seems to be caught between Scylla and Charybdis: the typed truth predicate of TB is too weak, while the full unrestricted T-schema is too strong.

It seems reasonable to steer between the two extremes in the middle…

Liberalising the type restriction

There have been various proposals to lift the type restrictions on the T-sentences.

Motives:

- Eg the following T-sentence looks ok:

  "Grass is red' is not true' is true iff 'Grass is red' is not true.

- A more liberal approach might help to regain deductive power.

However, one seems to be caught between Scylla and Charybdis: the typed truth predicate of TB is too weak, while the full unrestricted T-schema is too strong.

It seems reasonable to steer between the two extremes in the middle…

But there are other creatures as horrifying as deductive weakness and inconsistency, as McGee (1992) has demonstrated.

Horwich's proposal

[…] we must conclude that permissible instantiations of the equivalence schema are restricted in some way so as to avoid paradoxical results. […] Given our purposes it suffices for us to concede that certain instances of the equivalence schema are not to be included as axioms of the minimal theory, and to note that the principles governing our selection of excluded instances are, in order of priority: (a) that the minimal theory not engender 'liar-type' contradictions; (b) that the set of excluded instances be as small as possible; and—perhaps just as important as (b)—(c) that there be a constructive specification of the excluded instances that is as simple as possible. Horwich 1990 p. 41f

Maximal consistent instances of schema T

So the aim is to find a set of sentences $T\overline{\varphi} \leftrightarrow \varphi$ such that

- The set is consistent.

Maximal consistent instances of schema T

So the aim is to find a set of sentences $T\overline{\varphi} \leftrightarrow \varphi$ such that

- The set is consistent.
- The set is maximal, ie no further sentences of the form $T\overline{\varphi} \leftrightarrow \varphi$ can be consistently added over $\mathcal{A}$.

Maximal consistent instances of schema T

So the aim is to find a set of sentences $T\overline{\varphi} \leftrightarrow \varphi$ such that

- The set is consistent.
- The set is maximal, ie no further sentences of the form $T\overline{\varphi} \leftrightarrow \varphi$ can be consistently added over $\mathcal{A}$.
- The set is recursively enumerable (?).

Maximal consistent instances of schema T

Theorem (McGee)

Let φ be some sentence, then there is a sentence γ such that

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

Maximal consistent instances of schema T

Theorem (McGee)

Let φ be some sentence, then there is a sentence γ such that

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

Proof.

$$\mathcal{A} \vdash \gamma \leftrightarrow (T\overline{\gamma} \leftrightarrow \varphi) \qquad \text{diagonalisation}$$

$$\mathcal{A} \vdash \varphi \leftrightarrow (T\overline{\gamma} \leftrightarrow \gamma) \qquad \text{propositional logic}$$

$\dashv$

Maximal consistent instances of schema T

McGee's observation spells disaster for Horwich's proposal.

Theorem (McGee)

- *If a consistent set of T-sentences is recursive, it's not maximal: by Gödel's first incompleteness theorem there will be an undecidable sentence φ, which is equivalent to a T-sentence.*

Maximal consistent instances of schema T

McGee's observation spells disaster for Horwich's proposal.

Theorem (McGee)

- *If a consistent set of T-sentences is recursive, it's not maximal: by Gödel's first incompleteness theorem there will be an undecidable sentence φ, which is equivalent to a T-sentence.*
- *Maximal sets are too complicated. They can't be Π^0_1 or Σ^0_1.*

Maximal consistent instances of schema T

McGee's observation spells disaster for Horwich's proposal.

Theorem (McGee)

- *If a consistent set of T-sentences is recursive, it's not maximal: by Gödel's first incompleteness theorem there will be an undecidable sentence φ, which is equivalent to a T-sentence.*
- *Maximal sets are too complicated. They can't be Π^0_1 or Σ^0_1.*
- *There are many, in fact uncountably many different maximal consistent sets of T-sentences (if $\mathcal{A}$ is consistent).*

Maximal consistent instances of schema T

McGee's observation spells disaster for Horwich's proposal.

Theorem (McGee)

- *If a consistent set of T-sentences is recursive, it's not maximal: by Gödel's first incompleteness theorem there will be an undecidable sentence φ, which is equivalent to a T-sentence.*
- *Maximal sets are too complicated. They can't be Π^0_1 or Σ^0_1.*
- *There are many, in fact uncountably many different maximal consistent sets of T-sentences (if $\mathcal{A}$ is consistent).*
- *Consistent sets of T-sentences can prove horrible results worse than any inconsistency.*

Strong instances of schema T

McGee's observation has its destructive uses, but it also has a neglected constructive side.

Strong instances of schema T

McGee's observation has its destructive uses, but it also has a neglected constructive side.

It's often assumed that an axiomatisation of truth by T-sentences is either inacceptably weak or inconsistent. McGee's theorem shows that this view is incorrect.

Strong instances of schema T

McGee's observation has its destructive uses, but it also has a neglected constructive side.

It's often assumed that an axiomatisation of truth by T-sentences is either inacceptably weak or inconsistent. McGee's theorem shows that this view is incorrect.

Assume you have a favourite axiomatisation of truth (say the KF axioms or the like). Let χ the conjunction of these axioms. Then McGee's theorem implies the existence of a T-sentence such that

$$\mathcal{A} \vdash \chi \leftrightarrow (\mathrm{True}\,\overline{\gamma} \leftrightarrow \gamma)$$

Thus Davidson's theory, KF and so on can be finitely axiomatised by a single T-sentence.

Strong instances of schema T

McGee's observation has its destructive uses, but it also has a neglected constructive side.

It's often assumed that an axiomatisation of truth by T-sentences is either inacceptably weak or inconsistent. McGee's theorem shows that this view is incorrect.

Assume you have a favourite axiomatisation of truth (say the KF axioms or the like). Let χ the conjunction of these axioms. Then McGee's theorem implies the existence of a T-sentence such that

$$\mathcal{A} \vdash \chi \leftrightarrow (\mathrm{True}\,\overline{\gamma} \leftrightarrow \gamma)$$

Thus Davidson's theory, KF and so on can be finitely axiomatised by a single T-sentence.

The problem remains to tell a story why one should accept $T\overline{\gamma} \leftrightarrow \gamma$. If one justifies the acceptance of that T-sentence by appeal to your favourite theory, we have given up disquotationalism.

Strong instances of schema T

Gut feeling

- *Tarski's way of blocking the paradoxes is less damaging to the 'inductive' definition of truth than to the T-sentences as axioms.*

Strong instances of schema T

Gut feeling

- *Tarski's way of blocking the paradoxes is less damaging to the 'inductive' definition of truth than to the T-sentences as axioms.*
- *The T-sentences are as good as any axiomatic theory of truth, if the paradoxes are blocked in an appropriate way.*

Strong instances of schema T

Gut feeling

- *Tarski's way of blocking the paradoxes is less damaging to the 'inductive' definition of truth than to the T-sentences as axioms.*
- *The T-sentences are as good as any axiomatic theory of truth, if the paradoxes are blocked in an appropriate way.*
- *We need to come up with a better method for sorting the good instances from the bad instances of schema T.*

Strong instances of schema T

Gut feeling

- *Tarski's way of blocking the paradoxes is less damaging to the 'inductive' definition of truth than to the T-sentences as axioms.*
- *The T-sentences are as good as any axiomatic theory of truth, if the paradoxes are blocked in an appropriate way.*
- *We need to come up with a better method for sorting the good instances from the bad instances of schema T.*

To me it's still unclear whether it might be possible to defend a theory based on T-sentence which is not deductively weak.

Strong instances of schema T

Gut feeling

- *Tarski's way of blocking the paradoxes is less damaging to the 'inductive' definition of truth than to the T-sentences as axioms.*
- *The T-sentences are as good as any axiomatic theory of truth, if the paradoxes are blocked in an appropriate way.*
- *We need to come up with a better method for sorting the good instances from the bad instances of schema T.*

To me it's still unclear whether it might be possible to defend a theory based on T-sentence which is not deductively weak.

Missing: maximal conservative sets of instances of schema T Cieśliński (2007).

Strong instances of schema T

Gut feeling

- *Tarski's way of blocking the paradoxes is less damaging to the 'inductive' definition of truth than to the T-sentences as axioms.*
- *The T-sentences are as good as any axiomatic theory of truth, if the paradoxes are blocked in an appropriate way.*
- *We need to come up with a better method for sorting the good instances from the bad instances of schema T.*

To me it's still unclear whether it might be possible to defend a theory based on T-sentence which is not deductively weak.

Missing: maximal conservative sets of instances of schema T Cieśliński (2007).'Uniform' T-sentences and positive T-sentences. I don't have the tools available for treating them now. But there are well motivated and stron theories based on T-sentences.

Truth & Paradox

III · More Beasts and Dragons

Volker Halbach

Nordic Logic Summer School 2017

How things can go wrong

Paradox is not the same as mere inconsistency: there are many ways things can go wrong:

- The theory is inconsistent.

How things can go wrong

Paradox is not the same as mere inconsistency: there are many ways things can go wrong:

- The theory is inconsistent.
- The theory cannot be combined with another plausible theory. If a theory of future cannot be combined with the analogous theory of past truth, something is wrong.

How things can go wrong

Paradox is not the same as mere inconsistency: there are many ways things can go wrong:

- The theory is inconsistent.
- The theory cannot be combined with another plausible theory. If a theory of future cannot be combined with the analogous theory of past truth, something is wrong.
- The theory is internally inconsistent: the theory proves that everything is true.

How things can go wrong

Paradox is not the same as mere inconsistency: there are many ways things can go wrong:

- The theory is inconsistent.
- The theory cannot be combined with another plausible theory. If a theory of future cannot be combined with the analogous theory of past truth, something is wrong.
- The theory is internally inconsistent: the theory proves that everything is true.
- The theory proves a false claim in the base language (ie in the language without the truth predicate).

How things can go wrong

Paradox is not the same as mere inconsistency: there are many ways things can go wrong:

- The theory is inconsistent.
- The theory cannot be combined with another plausible theory. If a theory of future cannot be combined with the analogous theory of past truth, something is wrong.
- The theory is internally inconsistent: the theory proves that everything is true.
- The theory proves a false claim in the base language (ie in the language without the truth predicate).
- The theory has trivial models, eg, truth can be interpreted by the empty set.

How things can go wrong

Paradox is not the same as mere inconsistency: there are many ways things can go wrong:

- The theory is inconsistent.
- The theory cannot be combined with another plausible theory. If a theory of future cannot be combined with the analogous theory of past truth, something is wrong.
- The theory is internally inconsistent: the theory proves that everything is true.
- The theory proves a false claim in the base language (ie in the language without the truth predicate).
- The theory has trivial models, eg, truth can be interpreted by the empty set.
- The theory is ω-inconsistent.

How things can go wrong

Paradox is not the same as mere inconsistency: there are many ways things can go wrong:

- The theory is inconsistent.
- The theory cannot be combined with another plausible theory. If a theory of future cannot be combined with the analogous theory of past truth, something is wrong.
- The theory is internally inconsistent: the theory proves that everything is true.
- The theory proves a false claim in the base language (ie in the language without the truth predicate).
- The theory has trivial models, eg, truth can be interpreted by the empty set.
- The theory is ω-inconsistent.

Generally, consistency proofs are good, but a full proof-theoretic analysis is better. Only such an analysis can prove that the theory doesn't contain any hidden paradoxes.

The constructive applications

On the next couple of slides I sketch some classical applications of diagonalisation.

The constructive applications

On the next couple of slides I sketch some classical applications of diagonalisation.

Many of them can be turned into ‘paradoxes’.

Gödel's first theorem

Ok, it isn't Gödel's incompleteness theorem, but it's very similar in structure:

Theorem (Gödel's first theorem)

Assume $\mathcal{A} \vdash \varphi$ if and only if $\mathcal{A} \vdash T\overline{\varphi}$ holds for all sentences. Then there is a sentence γ, such that neither γ itself nor its negation is derivable in $\mathcal{A}$ except that $\mathcal{A}$ itself is already inconsistent.

Proof

(4)	$\mathcal{A} \vdash \gamma \leftrightarrow \neg T\overline{\gamma}$	diagonalisation
(5)	$\mathcal{A} \vdash \gamma$	assumption
(6)	$\mathcal{A} \vdash T\overline{\gamma}$	NEC
(7)	$\mathcal{A} \vdash \neg T\overline{\gamma}$	(4)
(8)	$\mathcal{A} \vdash \neg\gamma$	assumption
(9)	$\mathcal{A} \vdash T\overline{\gamma}$	(4)
(10)	$\mathcal{A} \vdash \gamma$	CONEC

The liar again

Theorem

Assume $\mathcal{A} \vdash \varphi$ if and only if $\mathcal{A} \vdash T\overline{\varphi}$ holds for all sentences. Then the liar sentence is undecidable in $\mathcal{A}$, if $\mathcal{A}$ is consistent.

Thus if the T-schema is weakened to a rule, the liar sentence must be undecidable. Thus theories (such as KF) containing NEC and deciding the liar sentence, cannot have CONEC.

The real incompleteness theorem

Gödel showed that a provability predicate Bew(v) can be defined in a certain system of arithmetic corresponding to our theory $\mathcal{A}$. More precisely, he defined a formula Bew(v)

$$\mathcal{A} \vdash \psi \text{ if and only if } \mathcal{A} \vdash \text{Bew}(\overline{\psi})$$

holds for all formulæ ψ of $\mathcal{L}$ if $\mathcal{A}$ is ω-consistent. ω-consistency is a stronger condition than pure consistency.

A look at the second incompleteness theorem

The ‘modal’ reasoning leading to the second incompleteness theorem can be paraphrased in $\mathcal{A}$.

A look at the second incompleteness theorem

The ‘modal’ reasoning leading to the second incompleteness theorem can be paraphrased in $\mathcal{A}$.

The second incompleteness theorem and Löb’s theorem have been used to derive further paradoxes. I believe that most paradoxes involving self-reference can be reduced to Löb’s theorem.

A look at the second incompleteness theorem

The ‘modal’ reasoning leading to the second incompleteness theorem can be paraphrased in $\mathcal{A}$.

The second incompleteness theorem and Löb’s theorem have been used to derive further paradoxes. I believe that most paradoxes involving self-reference can be reduced to Löb’s theorem.

In particular, the incompleteness theorems yield more information on weakenings of the T-scheme and ways to block Montague’s paradox.

Weaker reflection

Theorem

The scheme $T\overline{T\overline{\varphi}} \to \varphi$ *is inconsistent with NEC. The same holds for* $T\overline{T\overline{T\overline{\varphi}}} \to \varphi$ *etc.*

Weaker reflection

Theorem

The scheme $T\overline{T\overline{\varphi}} \to \varphi$ is inconsistent with NEC. The same holds for $T\overline{T\overline{T\overline{\varphi}}} \to \varphi$ etc.

Proof.

We diagonalize the formula $T(\overline{T}\ulcorner \mathrm{qv} \urcorner)$ to obtain the following γ:

$\mathcal{A} \vdash \gamma \leftrightarrow \neg T\overline{T\overline{\gamma}}$	
$\mathcal{A} \vdash T\overline{T\overline{\gamma}} \to \gamma$	assumption
$\mathcal{A} \vdash \neg T\overline{T\overline{\gamma}}$	two preceding lines
$\mathcal{A} \vdash \gamma$	first line
$\mathcal{A} \vdash T\overline{\gamma}$	NEC
$\mathcal{A} \vdash T\overline{T\overline{\gamma}}$	4

$\dashv$

Internal inconsistency

Plain inconsistency is not the only way a system can fail to acceptable. "Internal" inconsistency is almost as startling. Let $\bot$ be some fixed logical contradiction, e.g., $\neg \neq \neg$. A theory is said to be internally inconsistent (with respect to T) if and only if $\mathcal{A} \vdash T\overline{\bot}$.

Internal inconsistency

Plain inconsistency is not the only way a system can fail to acceptable. "Internal" inconsistency is almost as startling. Let $\bot$ be some fixed logical contradiction, e.g., $\overline{\neg} \neq \overline{\neg}$. A theory is said to be internally inconsistent (with respect to T) if and only if $\mathcal{A} \vdash T\overline{\bot}$.

Theorem (Thomason 1980)

The schemata $T\overline{T\overline{\varphi} \to \varphi}$ *and* $T\overline{\varphi \to \psi} \to (T\overline{\varphi} \to T\overline{\psi})$ *is internally inconsistent with NEC.*

Internal inconsistency

Plain inconsistency is not the only way a system can fail to acceptable. "Internal" inconsistency is almost as startling. Let $\bot$ be some fixed logical contradiction, e.g., $\overline{\neg} \neq \overline{\neg}$. A theory is said to be internally inconsistent (with respect to T) if and only if $\mathcal{A} \vdash T\overline{\bot}$.

Theorem (Thomason 1980)

The schemata $T\overline{T\overline{\varphi} \to \varphi}$ *and* $T\overline{\varphi \to \psi} \to (T\overline{\varphi} \to T\overline{\psi})$ *is internally inconsistent with NEC.*

Proof.

One runs the proof of Montague's theorem in the scope of T. ⊣

The Löb derivability conditions

Let K be the following scheme:

(K) $$T\overline{\varphi \to \psi} \to (T\overline{\varphi} \to T\overline{\psi})$$

The Löb derivability conditions

Let K be the following scheme:

(K) $$T\overline{\varphi \to \psi} \to (T\overline{\varphi} \to T\overline{\psi})$$

4 is the following scheme:

(4) $$T\overline{\varphi} \to T\overline{T\overline{\varphi}}$$

K4 contains NEC, K, 4 and all axioms of $\mathcal{A}$.

The Löb derivability conditions

Let K be the following scheme:

(K) $$T\overline{\varphi \to \psi} \to (T\overline{\varphi} \to T\overline{\psi})$$

4 is the following scheme:

(4) $$T\overline{\varphi} \to T\overline{T\overline{\varphi}}$$

K4 contains NEC, K, 4 and all axioms of $\mathcal{A}$. K4 has been thought to be adequate for necessity and, in some cases, for truth.

Remark

One can show that Gödel's provability predicate satisfies K4. NEC, K, 4 formulated for the provability predicate are known as Löb's derivability conditions. See (Boolos 1993) for more information.

T can also be read as 'Nigel is justified in believing …', knowledge, necessity etc.

Löb's theorem

Now I want to generalise the question: for which sentences can we have $T\overline{\varphi} \to \varphi$?

Theorem (Löb's theorem)

$\mathrm{K4} \vdash T\overline{T\overline{\varphi} \to \varphi} \to T\overline{\varphi}$

Löb's theorem

Now I want to generalise the question: for which sentences can we have $T\overline{\varphi} \to \varphi$?

Theorem (Löb's theorem)

$\mathrm{K4} \vdash T\overline{T\overline{\varphi} \to \varphi} \to T\overline{\varphi}$

The corresponding rule follows as well:

Theorem

If $\mathrm{K4} \vdash T\overline{\varphi} \to \varphi$, *then* $\mathrm{K4} \vdash \varphi$

Löb's theorem

Now I want to generalise the question: for which sentences can we have $T\overline{\varphi} \to \varphi$?

Theorem (Löb's theorem)

$\mathrm{K4} \vdash T\overline{T\overline{\varphi} \to \varphi} \to T\overline{\varphi}$

The corresponding rule follows as well:

Theorem

If $\mathrm{K4} \vdash T\overline{\varphi} \to \varphi$, *then* $\mathrm{K4} \vdash \varphi$

Thus in the context of K4 adding $T\overline{\varphi} \to \varphi$ makes φ itself provable.

Löb's theorem: the proof

Proof.

$\gamma \leftrightarrow (T\overline{\gamma} \to \varphi)$	diagonalization
$T\overline{\gamma} \to T\overline{T\overline{\gamma} \to \varphi}$	K
$T\overline{\gamma} \to (T\overline{T\overline{\gamma}} \to T\overline{\varphi})$	K and NEC
$T\overline{\gamma} \to T\overline{\varphi}$	4
$(T\overline{\varphi} \to \varphi) \to (T\overline{\gamma} \to \varphi)$	
$(T\overline{\varphi} \to \varphi) \to \gamma$	first line
$T\overline{(T\overline{\varphi} \to \varphi) \to \gamma}$	NEC
$T\overline{(T\overline{\varphi} \to \varphi)} \to T\overline{\gamma}$	K
$T\overline{(T\overline{\varphi} \to \varphi)} \to T\overline{\varphi}$	line 4

⊣

Gödel's second theorem

Now fix a contradiction, for instance $\underline{0} \neq \underline{0}$ and call it $\bot$.

Gödel's second theorem

Now fix a contradiction, for instance $\underline{0} \neq \underline{0}$ and call it $\bot$.

Theorem (Gödel's second theorem)

$\mathrm{K4}$ is inconsistent with $\neg T\ulcorner\bot\urcorner$. Thus $\mathrm{K4} \nvdash \neg T\ulcorner\bot\urcorner$ if $\mathrm{K4}$ is consistent.

Gödel's second theorem

Now fix a contradiction, for instance $\underline{0} \neq \underline{0}$ and call it $\bot$.

Theorem (Gödel's second theorem)

K4 *is inconsistent with* $\neg T\ulcorner\bot\urcorner$. *Thus* K4 $\nvdash \neg T\ulcorner\bot\urcorner$ *if* K4 *is consistent.*

There is also a formalized version of Gödel's second incompleteness theorem, which can easily be derived from Löb's theorem.

Gödel's second theorem

Now fix a contradiction, for instance $\underline{0} \neq \underline{0}$ and call it $\bot$.

Theorem (Gödel's second theorem)

K4 *is inconsistent with* $\neg T\overline{\bot}$. *Thus* K4 $\nvdash \neg T\overline{\bot}$ *if* K4 *is consistent.*

There is also a formalized version of Gödel's second incompleteness theorem, which can easily be derived from Löb's theorem.

Theorem (Gödel's second theorem formalized)

K4 $\vdash T\overline{\bot} \vee \neg T\overline{\neg T\overline{\bot}}$.

The dark side again

Now here is another paradox. I didn't know where else it should be put.

The dark side again

Now here is another paradox. I didn't know where else it should be put.

Very much like the paradox on how not to formalise the T-sentences is arises from the interaction of two predicates, viz two truth predicates: future and past truth.

The dark side again

Now here is another paradox. I didn't know where else it should be put.

Very much like the paradox on how not to formalise the T-sentences is arises from the interaction of two predicates, viz two truth predicates: future and past truth.

Horsten and Leitgeb call it the 'no future' paradox.

No future: the language

Assume $\mathcal{L}$ contains four predicates G, H, F and P. The intended reading of Gx is "x always will be the case", while Hx should be read as "x always has been the case". Similarly Fx is to be read as "x will be the case at some point (in the future)"; finally Px stands for "x has been the case at some point (in the past)". G and H can easily be defined from F and P, respectively (or also vice versa). The four predicates correspond to the well known operators from temporal logic, the difference being that G and H are here predicates rather than operators.

No future: the axioms

The system K_t^* is given by the following axiom schemes for all sentences φ and ψ of the language $\mathcal{L}$. This means in particular that φ and ψ may contain the predicates G and H.

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

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

G2 $\varphi \to H\overline{F\overline{\varphi}}$

H2 $\varphi \to G\overline{P\overline{\varphi}}$

G3 $G\overline{\varphi} \leftrightarrow \neg F\overline{\neg\varphi}$

H3 $H\overline{\varphi} \leftrightarrow \neg P\overline{\neg\varphi}$

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

[1] In (Horsten and Leitgeb 2001) there is a typo in the formulation of this axiom: the last occurrence of H is a G in the original paper.

No future: the inconcistency

These axioms are analogues of axioms from temporal logic.

No future: the inconcistency

These axioms are analogues of axioms from temporal logic.

K_t^* is consistent (see Horsten and Leitgeb 2001), but "internally" inconsistent, i.e., K_t^* proves that there is no future and no past.

No future: the inconcistency

These axioms are analogues of axioms from temporal logic.

K_t^* is consistent (see Horsten and Leitgeb 2001), but "internally" inconsistent, i.e., K_t^* proves that there is no future and no past.

Theorem (no future paradox, Horsten and Leitgeb 2001)

$\mathrm{K}_t^* \vdash \mathrm{H}\overline{\bot} \wedge \mathrm{G}\overline{\bot}$.

No future: the inconcistency

These axioms are analogues of axioms from temporal logic.

K_t^* is consistent (see Horsten and Leitgeb 2001), but "internally" inconsistent, i.e., K_t^* proves that there is no future and no past.

Theorem (no future paradox, Horsten and Leitgeb 2001)

$K_t^* \vdash H\overline{\bot} \wedge G\overline{\bot}$.

Thus K_t^* claims that at all moments in the future $\bot$ will hold. Since $\bot$ is a contradiction, there cannot be any moment in the future. Therefore there is no future. Analogously, but less dramatically, there also has never been a moment in the past.

No future: the proof

I shall only prove that there is no future, i.e., $\mathrm{K}_t^* \vdash \mathrm{G}\overline{\bot}$.

(11)	$\mathrm{K}_t^* \vdash \gamma \leftrightarrow \mathrm{G}\overline{\mathrm{P}\overline{\neg\gamma}}$	diagonalisation
	$\mathrm{K}_t^* \vdash \neg\gamma \leftrightarrow \neg\mathrm{G}\overline{\mathrm{P}\overline{\neg\gamma}}$	
	$\mathrm{K}_t^* \vdash \neg\gamma \rightarrow \gamma$	H2
(12)	$\mathrm{K}_t^* \vdash \gamma$	
(13)	$\mathrm{K}_t^* \vdash \mathrm{G}\overline{\mathrm{P}\overline{\neg\gamma}}$	from (**??**) and previous line
	$\mathrm{K}_t^* \vdash \mathrm{H}\overline{\gamma}$	N and (12)
	$\mathrm{K}_t^* \vdash \neg\mathrm{P}\overline{\neg\gamma}$	H3
(14)	$\mathrm{K}_t^* \vdash \mathrm{G}\overline{\neg\mathrm{P}\overline{\neg\gamma}}$	N
(15)	$\mathrm{K}_t^* \vdash \mathrm{G}\overline{\bot}$	(13), (14) and G1

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

No future: an inconsistency

In this framework one can assert that there is a future by saying that if φ will always be the case then φ will be the case at some time:

No future: an inconsistency

(FUT) $\quad G\overline{\varphi} \rightarrow F\overline{\varphi}$

Corollary (Horsten and Leitgeb 2001)

H2, G3, H3, N and FUT together are inconsistent.

Proof.

One proves (13) and (14) as in he preceding Theorem and applies FUT to the latter in order to obtain $F\overline{\neg P\overline{\neg\gamma}}$, which implies in turn $\neg G\overline{P\overline{\neg\gamma}}$ by G3 and is therefore inconsistent with (13). $\dashv$

Actually (Horsten and Leitgeb 2001) proved the dual of this corollary.

Truth & Paradox

IV · Tarskian Truth Axiomatised

Volker Halbach

Nordic Logic Summer School 2017

Schemata and universally quantified axioms

The axiomatic theories I have considered so far, are based on schemata as axioms, eg, on $T\overline{\varphi} \leftrightarrow \varphi$.

Schemata and universally quantified axioms

The axiomatic theories I have considered so far, are based on schemata as axioms, eg, on $T\overline{\varphi} \leftrightarrow \varphi$.

However, I have already mentioned that one would like to have universally quantified theorems as consequences of our theory such as

$$\forall x(\mathrm{Sent}(x) \rightarrow (\mathrm{True}\, x \vee \mathrm{True}\, \dot{\neg} x))$$

Schemata and universally quantified axioms

The axiomatic theories I have considered so far, are based on schemata as axioms, eg, on $T\overline{\varphi} \leftrightarrow \varphi$.

However, I have already mentioned that one would like to have universally quantified theorems as consequences of our theory such as

$$\forall x(\mathrm{Sent}(x) \rightarrow (\mathrm{True}\, x \vee \mathrm{True}\, \dot{\neg} x))$$

Thus it's only natural to look at theories containing such sentences as axioms. Examples are Tarski's definition of truth turned into an axiomatic theory, or the Kripke-Feferman theory of truth.

Schemata and universally quantified axioms

The axiomatic theories I have considered so far, are based on schemata as axioms, eg, on $T\overline{\varphi} \leftrightarrow \varphi$.

However, I have already mentioned that one would like to have universally quantified theorems as consequences of our theory such as

$$\forall x(\mathrm{Sent}(x) \rightarrow (\mathrm{True}\, x \vee \mathrm{True}\, \dot{\neg} x))$$

Thus it's only natural to look at theories containing such sentences as axioms. Examples are Tarski's definition of truth turned into an axiomatic theory, or the Kripke-Feferman theory of truth.

To formulate such axioms I look back again at our theory $\mathcal{A}$ of syntax.

First, I should add Sent. Let's say Sent is a unary predicate symbol of our language. Moreover we have as axioms for all expressions (strings of symbols) e:

Additional Axiom

1. $Sent(\overline{e})$ *iff e is a sentence*
2. $\neg Sent(\overline{e})$ *iff e is not a sentence*

It would be better to define Sent(), but this brute-force method will do.

Quantifying-in

Quantifying into quotational context is notoriously problematic. Occurrences of variables within quotation marks cannot be bound from 'outside'. For instance, the quantifier $\forall \mathrm{v}$ is idling in the sentence $\forall \mathrm{v} T\overline{\mathrm{v} = \mathrm{v}}$.

Quantifying-in

Quantifying into quotational context is notoriously problematic. Occurrences of variables within quotation marks cannot be bound from 'outside'. For instance, the quantifier $\forall \mathrm{v}$ is idling in the sentence $\forall \mathrm{v} T\overline{\mathrm{v} = \mathrm{v}}$. In some cases, however, it is possible to bind quoted variables in a sense to be explained. As long as our expressions are assumed to range over expressions there is a way to bind variables in a quoted expression.

Quantifying-in

Assume we want to say that it is true that every expression is identical with itself, we cannot do this by $\forall \mathrm{v} T \overline{\mathrm{v} = \mathrm{v}}$, but by saying the following:

> *For all expressions e: If we replace in the formula* v = v *every occurrence of* v *by the quotational constant for e, then the resulting sentence is true.*

Quantifying-in

Assume we want to say that it is true that every expression is identical with itself, we cannot do this by $\forall \mathrm{v} T\overline{\mathrm{v} = \mathrm{v}}$, but by saying the following:

> *For all expressions e: If we replace in the formula* v = v *every occurrence of* v *by the quotational constant for e, then the resulting sentence is true.*

This can be formalized by the following expression:

$$\forall \mathrm{v}\, T\mathrm{sub}(\overline{\mathrm{v} = \mathrm{v}}, \overline{\mathrm{v}}, \mathrm{qv})$$

From this we can derive, for instance, $T\overline{\overline{\neg} = \overline{\neg}}$ in $\mathcal{A}$.

Quantifying-in

The trick can be generalized. Assume $\varphi(x)$ is a formula with no bound occurrences of the variable x, then we abbreviate by $\overline{\varphi(\overset{\bullet}{x})}$ the complex term

$$\mathrm{sub}(\overline{\varphi(x)}, \overline{x}, \mathrm{q}x)$$

Quantifying-in

The trick can be generalized. Assume $\varphi(x)$ is a formula with no bound occurrences of the variable x, then we abbreviate by $\overline{\varphi(\dot{x})}$ the complex term

$$\mathrm{sub}(\overline{\varphi(x)}, \overline{x}, \mathrm{q}x)$$

Perhaps, with your permission, can we say that sub replaces only *free* occurrences of variables? I am pretty confident that I can define a new sub' function that does exactly this. Or I define sub in this way from the beginning.

Quantifying-in

The trick can be generalized. Assume $\varphi(x)$ is a formula with no bound occurrences of the variable x, then we abbreviate by $\varphi(\overline{\bullet})$ the complex term

$$\mathrm{sub}(\overline{\varphi(x)}, \overline{x}, \mathrm{q}x)$$

Perhaps, with your permission, can we say that sub replaces only *free* occurrences of variables? I am pretty confident that I can define a new sub' function that does exactly this. Or I define sub in this way from the beginning.

We have for every expression its quotational constant as its standard name; moreover, the q describes a function assigning to each expression its quotational constant. This allows to replace the quoted variable by a name for the object for that the variable stands.

Limitations of the method

So far, $\mathcal{A}$ a theory of expressions *and*, possibly, of further objects.

Limitations of the method

So far, $\mathcal{A}$ a theory of expressions *and*, possibly, of further objects.

For the function symbol q I have employed the following axiom:

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

There are no axioms that tell us what happens when q is applied to other objects.

Limitations of the method

So far, $\mathcal{A}$ a theory of expressions *and*, possibly, of further objects.

For the function symbol q I have employed the following axiom:

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

There are no axioms that tell us what happens when q is applied to other objects.

In the best case it gives us a name of that object, if there is one.

Limitations of the method

So far, $\mathcal{A}$ a theory of expressions *and*, possibly, of further objects.

For the function symbol q I have employed the following axiom:

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

There are no axioms that tell us what happens when q is applied to other objects.

In the best case it gives us a name of that object, if there is one.

The method described above for quantifying into quoted contexts works only if q gives us a name for each object. In a theory of expressions only this condition is satisfied.

Limitations of the method

So far, $\mathcal{A}$ a theory of expressions *and*, possibly, of further objects.

For the function symbol q I have employed the following axiom:

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

There are no axioms that tell us what happens when q is applied to other objects.

In the best case it gives us a name of that object, if there is one.

The method described above for quantifying into quoted contexts works only if q gives us a name for each object. In a theory of expressions only this condition is satisfied.

If q doesn't give us a name for every object, we would have to use satisfaction (instead of unary truth) or de re-necessity (instead of unary de dicto-necessity).

The uniform T-sentences

Now I strengthen the T-sentences a little bit by adding ‘uniformity’.

The uniform T-sentences

Now I strengthen the T-sentences a little bit by adding 'uniformity'.

The theory UTB is very much like TB:

Definition

Call the theory $\mathcal{A}$ plus all the following equivalences UTB:

$$\forall x(\text{True } \overline{\varphi(\dot{x})} \leftrightarrow \varphi(x))$$

where $\varphi(x)$ is a formula of $\mathcal{L}$ with at most x free.

The uniform T-sentences

I am tempted to say that

Theorem

UTB is conservative over $\mathcal{A}$.

The claim is true at least for reasonable $\mathcal{A}$, such as the minimal $\mathcal{A}$ containing only the axioms I have mentioned so far.

The uniform T-sentences

I am tempted to say that

Theorem

UTB is conservative over $\mathcal{A}$.

The claim is true at least for reasonable $\mathcal{A}$, such as the minimal $\mathcal{A}$ containing only the axioms I have mentioned so far.

Proof: One can define partial truth predicates applying to sentences up to a certain complexity, if $\mathcal{A}$ is strong enough. Then use a compactness argument. Alternatively use partial inductive satisfaction classes, see Kaye (1991).

The uniform T-sentences and disquotationalism

I guess UTB is a disquotationalist theory – perhaps not of truth but of satisfaction.

The uniform T-sentences and disquotationalism

I guess UTB is a disquotationalist theory – perhaps not of truth but of satisfaction.

UTB doesn't prove generalisations such as

$$\forall x(\mathrm{Sent}(x) \rightarrow (\mathrm{True}\, x \vee \mathrm{True}\, \dot{\neg} x)).$$

We can even chooses PA as our base theory $\mathcal{A}$ and add all induction axioms including those with T. The resulting theory is still conservative over PA.

Defining membership from truth

I have defined $\overline{\varphi(\dot{x})}$ as the complex term $\mathrm{sub}(\overline{\varphi(x)}, \overline{x}, \mathrm{q}x)$.

Defining membership from truth

I have defined $\overline{\varphi(\dot{x})}$ as the complex term $\mathrm{sub}(\overline{\varphi(x)}, \overline{x}, \mathrm{q}x)$.

Definition

Define $x \in y$ as $\mathrm{sub}(y, \overline{\mathrm{v}}, \mathrm{q}x)$, and define $\mathrm{Set}(y)$ as $\mathrm{Sent}(\mathrm{sub}(y, \overline{\mathrm{v}}, \overline{\overline{\neg}})) \wedge \neg\mathrm{Sent}(y)$.

Thus $\mathrm{Sent}(y)$ says that y itself isn't a sentence, but the result of replacing all free occurrences of v with a constant is, ie, y is a formula with exactly v free.

Defining membership from truth

I have defined $\overline{\varphi(\overset{\bullet}{x})}$ as the complex term $\mathrm{sub}(\overline{\varphi(x)}, \overline{x}, \mathrm{q}x)$.

Definition

Define $x \in y$ as $\mathrm{sub}(y, \overline{\mathrm{v}}, \mathrm{q}x)$, and define $\mathrm{Set}(y)$ as $\mathrm{Sent}(\mathrm{sub}(y, \overline{\mathrm{v}}, \overline{\overline{\neg}})) \wedge \neg\mathrm{Sent}(y)$.

Thus $\mathrm{Sent}(y)$ says that y itself isn't a sentence, but the result of replacing all free occurrences of v with a constant is, ie, y is a formula with exactly v free.

Moreover, I write $\forall X$ for $\forall x(\mathrm{Set}(x) \to \ldots$ etc.

Defining membership from truth

I have defined $\overline{\varphi(\overset{\bullet}{x})}$ as the complex term $\mathrm{sub}(\overline{\varphi(x)}, \overline{x}, \mathrm{q}x)$.

Definition

Define $x \in y$ as $\mathrm{sub}(y, \overline{\mathrm{v}}, \mathrm{q}x)$, and define $\mathrm{Set}(y)$ as $\mathrm{Sent}(\mathrm{sub}(y, \overline{\mathrm{v}}, \overline{\overline{\neg}})) \wedge \neg\mathrm{Sent}(y)$.

Thus $\mathrm{Sent}(y)$ says that y itself isn't a sentence, but the result of replacing all free occurrences of v with a constant is, ie, y is a formula with exactly v free.

Moreover, I write $\forall X$ for $\forall x(\mathrm{Set}(x) \to \ldots$ etc.

How much second-order quantification do we get in UTB?

Comprehension in UTB

The following is the first of a couple of results that show that one can mimic quantification over sets in certain truth-theoretic theories.

Comprehension in UTB

The following is the first of a couple of results that show that one can mimic quantification over sets in certain truth-theoretic theories.

Theorem

$UTB \vdash \exists X \forall z\big(z \in X \leftrightarrow \varphi(z)\big)$, *where* $\varphi(z)$ *is some formula not containing* True .

Comprehension in UTB

The following is the first of a couple of results that show that one can mimic quantification over sets in certain truth-theoretic theories.

Theorem

$UTB \vdash \exists X \forall z\big(z \in X \leftrightarrow \varphi(z)\big)$, *where* $\varphi(z)$ *is some formula not containing* True .

Proof.

$$UTB \vdash \forall z\Big(\text{True}\,\overline{\varphi(\dot{z})} \leftrightarrow \varphi(z)\Big) \tag{16}$$

$$UTB \vdash \exists X \forall z\big(z \in X \leftrightarrow \varphi(z)\big) \tag{17}$$

⊣

Comprehension in UTB

Thus uniform T-sentences and comprehension principles are closely related.

Comprehension in UTB

Thus uniform T-sentences and comprehension principles are closely related.

For those interested in conservativeness and the case $\mathcal{A} = PA$: All this holds even in the presence of full induction. Thus parameter-free ACA_0 is conservative over PA.

Comprehension in UTB

Thus uniform T-sentences and comprehension principles are closely related.

For those interested in conservativeness and the case $\mathcal{A} = PA$: All this holds even in the presence of full induction. Thus parameter-free ACA_0 is conservative over PA.

Later I'll mention more results of this kind. Strong second-order systems of arithmetic such as $\mathrm{RA}_{<\Gamma_0}$ or the like are reducible to truth theories. Strong ontological assumptions are reduced to semantical assumptions.

If we drop the restriction that $\varphi(x)$ is a formula of $\mathcal{L}$ with at most x free in UTB, that is, in the schema

$$\forall x(\text{True } \overline{\varphi(\dot{x})} \leftrightarrow \varphi(x))$$

we get an inconsistency.

Writing this with $\in$ gives Russell's paradox:

$$\forall x(x \in \overline{\varphi(\dot{x})} \leftrightarrow \varphi(x))$$

In this setting the liar sentence $\neg T\text{sub}(\overline{\neg\text{True v}}, \overline{\text{v}}, \overline{\neg\text{True v}})$ and the 'Russell' set coincide.

If we had a binary predicate Sat, we could obtain the Russell paradox from the following axiom without any syntax theory from this schema:

$$\forall x(\mathrm{Sat}(\overline{\varphi(\mathrm{v})}, x) \leftrightarrow \varphi(x))$$

The inconsistency follows in the following way:

$$\forall x(\mathrm{Sat}(\overline{\varphi(\mathrm{v})}, x) \leftrightarrow \varphi(x))$$

$$\forall x(\mathrm{Sat}(\overline{\neg\mathrm{Sat}(\mathrm{v}, \mathrm{v})}, x) \leftrightarrow \neg\mathrm{Sat}(x, x))$$

$$\mathrm{Sat}(\overline{\neg\mathrm{Sat}(\mathrm{v}, \mathrm{v})}, \overline{\neg\mathrm{Sat}(\mathrm{v}, \mathrm{v})}) \leftrightarrow \neg\mathrm{Sat}(\overline{\neg\mathrm{Sat}(\mathrm{v}, \mathrm{v})}, \overline{\neg\mathrm{Sat}(\mathrm{v}, \mathrm{v})})$$

The need for stronger theories

So we still need to look for a stronger theory.

The need for stronger theories

So we still need to look for a stronger theory.

Davidson's proposal: turn Tarski's definition of truth into axioms.

The need for stronger theories

So we still need to look for a stronger theory.

Davidson's proposal: turn Tarski's definition of truth into axioms.

At first I present Tarski's theory in my framework (you may find that there is little resemblance).

Defining truth

Definition

'is true' is defined by induction on the complexity of $\mathcal{L}$-sentences:

1. A sentence $s = t$ is true iff the value of s is the value of t, where s and t are closed terms of the language; and so on for further predicates of $\mathcal{L}$

Defining truth

Definition

'is true' is defined by induction on the complexity of $\mathcal{L}$-sentences:

1. A sentence $s = t$ is true iff the value of s is the value of t, where s and t are closed terms of the language; and so on for further predicates of $\mathcal{L}$
2. A negated $\mathcal{L}$-sentence $\neg\varphi$ is true iff φ is not true.

Defining truth

Definition

'is true' is defined by induction on the complexity of $\mathcal{L}$-sentences:

1. A sentence $s = t$ is true iff the value of s is the value of t, where s and t are closed terms of the language; and so on for further predicates of $\mathcal{L}$
2. A negated $\mathcal{L}$-sentence $\neg\varphi$ is true iff φ is not true.
3. A conditional $\varphi \to \psi$ is true iff φ is false or and ψ is true (φ and ψ are sentences of $\mathcal{L}$)

Defining truth

Definition

'is true' is defined by induction on the complexity of $\mathcal{L}$-sentences:

1. A sentence $s = t$ is true iff the value of s is the value of t, where s and t are closed terms of the language; and so on for further predicates of $\mathcal{L}$
2. A negated $\mathcal{L}$-sentence $\neg\varphi$ is true iff φ is not true.
3. A conditional $\varphi \rightarrow \psi$ is true iff φ is false or and ψ is true (φ and ψ are sentences of $\mathcal{L}$)
4. A universally quantified $\mathcal{L}$-sentence $\forall x \varphi(x)$ is true iff $\varphi(\overline{e})$ for all objects e.

For the last axiom it's assumed that there are only expressions in the domain of our standard model (or that $\overline{e}$ is the standard name for e, whatever e is).

The mathematics needed for defining truth

Ok, this definition is not very precise. If ZF is our mathematical theory we can prove that there is a set of sentences containing exactly those sentences claimed to be true in the definition on the previous slide.

The mathematics needed for defining truth

Ok, this definition is not very precise. If ZF is our mathematical theory we can prove that there is a set of sentences containing exactly those sentences claimed to be true in the definition on the previous slide.

To that end one can prove that inductive definitions determine unique sets – under certain conditions. If one applies the definition of truth to the language of set theory, we know already from Tarski's theorem on the undefinability of truth that there is no set containing the sentences declared true by that definition.

The mathematics needed for defining truth

Ok, this definition is not very precise. If ZF is our mathematical theory we can prove that there is a set of sentences containing exactly those sentences claimed to be true in the definition on the previous slide.

To that end one can prove that inductive definitions determine unique sets – under certain conditions. If one applies the definition of truth to the language of set theory, we know already from Tarski's theorem on the undefinability of truth that there is no set containing the sentences declared true by that definition.

Before formalising the definition of truth for $\mathcal{L}$-sentences I need more expressive power in $\mathcal{A}$.

Values

I assume from now on that the language of $\mathcal{A}$ contains also a predicate $\mathrm{val}(x) = \mathrm{val}(y)$ expressing that x and y are closed terms with the same value.

Additional Axiom

$\mathrm{val}(t) = \mathrm{val}(e)$ *if and only if t and s denote closed terms with the same value in the standard model.*

$\neg\mathrm{val}(t) = \mathrm{val}(e)$ *if and only if t and s do not denote closed terms with the same value in the standard model.*

Values

I assume from now on that the language of $\mathcal{A}$ contains also a predicate $\mathrm{val}(x) = \mathrm{val}(y)$ expressing that x and y are closed terms with the same value.

Additional Axiom

$\mathrm{val}(t) = \mathrm{val}(e)$ *if and only if t and s denote closed terms with the same value in the standard model.*
$\neg\mathrm{val}(t) = \mathrm{val}(e)$ *if and only if t and s do not denote closed terms with the same value in the standard model.*

You may think of val as a symbol representing the function that gives applied to a term of the language it's value, ie, he object denoted by that term.

Example

- $\mathcal{A} \vdash \neg\mathrm{val}(\overline{\neg}) = \mathrm{val}(\overline{\neg})$ (The negation symbol isn't a closed term.)

Values

I assume from now on that the language of $\mathcal{A}$ contains also a predicate $\mathrm{val}(x) = \mathrm{val}(y)$ expressing that x and y are closed terms with the same value.

Additional Axiom

$\mathrm{val}(t) = \mathrm{val}(e)$ *if and only if t and s denote closed terms with the same value in the standard model.*
$\neg\mathrm{val}(t) = \mathrm{val}(e)$ *if and only if t and s do not denote closed terms with the same value in the standard model.*

You may think of val as a symbol representing the function that gives applied to a term of the language it's value, ie, he object denoted by that term.

Example

- $\mathcal{A} \vdash \neg\mathrm{val}(\overline{\neg}) = \mathrm{val}(\overline{\neg})$ (The negation symbol isn't a closed term.)
- $\mathcal{A} \vdash \mathrm{val}(\mathrm{q}\overline{\neg}) = \mathrm{val}(\mathrm{q}\overline{\neg})$
- $\mathcal{A} \vdash \mathrm{val}(\overline{\overline{\neg}^\frown\overline{\wedge}}) = \mathrm{val}(\overline{\overline{\neg\wedge}})$

Closed terms

I need a predicate expressing that an object is closed term of $\mathcal{L}$:

Definition

$\mathrm{ClT}(x)$ abbreviates $\mathrm{val}(x) = \mathrm{val}(x)$.

Hence we have:
$\mathcal{A} \vdash \mathrm{ClT}(t)$ if and only if t is a term denoting a closed term in the standard model.

Example

- $\mathcal{A} \vdash \neg\mathrm{ClT}(\overline{\forall})$

Closed terms

I need a predicate expressing that an object is closed term of $\mathcal{L}$:

Definition

$\mathrm{ClT}(x)$ abbreviates $\mathrm{val}(x) = \mathrm{val}(x)$.

Hence we have:
$\mathcal{A} \vdash \mathrm{ClT}(t)$ if and only if t is a term denoting a closed term in the standard model.

Example

- $\mathcal{A} \vdash \neg\mathrm{ClT}(\overline{\forall})$
- $\mathcal{A} \vdash \mathrm{ClT}(\overline{\overline{\forall}})$

Closed terms

I need a predicate expressing that an object is closed term of $\mathcal{L}$:

Definition

$\mathrm{ClT}(x)$ abbreviates $\mathrm{val}(x) = \mathrm{val}(x)$.

Hence we have:
$\mathcal{A} \vdash \mathrm{ClT}(t)$ if and only if t is a term denoting a closed term in the standard model.

Example

- $\mathcal{A} \vdash \neg\mathrm{ClT}(\overline{\forall})$
- $\mathcal{A} \vdash \mathrm{ClT}(\overline{\overline{\forall}})$
- $\mathcal{A} \vdash \mathrm{ClT}(\mathrm{q}\overline{\forall \mathrm{v}\, \mathrm{v} = \mathrm{v}})$

More dots

In the following I'll use Feferman's dot notation extensively. I want to express in $\mathcal{A}$ 'the negation of', 'the conjunction of … and …', 'the universal quantification of … with respect to variable …'.

More dots

In the following I'll use Feferman's dot notation extensively. I want to express in $\mathcal{A}$ 'the negation of', 'the conjunction of … and …', 'the universal quantification of … with respect to variable …'.

These function expressions can be introduced as new axiom or they can be defined, eg:

More dots

In the following I'll use Feferman's dot notation extensively. I want to express in $\mathcal{A}$ 'the negation of', 'the conjunction of … and …', 'the universal quantification of … with respect to variable …'.

These function expressions can be introduced as new axiom or they can be defined, eg:

Definition

- $x \dot{=} y \stackrel{\text{def}}{=} x \hat{\equiv} y$

More dots

In the following I'll use Feferman's dot notation extensively. I want to express in $\mathcal{A}$ 'the negation of', 'the conjunction of … and …', 'the universal quantification of … with respect to variable …'.

These function expressions can be introduced as new axiom or they can be defined, eg:

Definition

- $x \dot{=} y \stackrel{\text{def}}{=} x \hat{\equiv} y$
- $\dot{\neg} x \stackrel{\text{def}}{=} \overline{\neg} \hat{} x$

More dots

In the following I'll use Feferman's dot notation extensively. I want to express in $\mathcal{A}$ 'the negation of', 'the conjunction of … and …', 'the universal quantification of … with respect to variable …'.

These function expressions can be introduced as new axiom or they can be defined, eg:

Definition

- $x \dot{=} y \stackrel{\text{def}}{=} x \hat{\equiv} y$
- $\dot{\neg} x \stackrel{\text{def}}{=} \overline{\hat{\neg} x}$
- $x \dot{\rightarrow} y \stackrel{\text{def}}{=} \overline{(\hat{x} \hat{\rightarrow} \hat{y})}$

More dots

In the following I'll use Feferman's dot notation extensively. I want to express in $\mathcal{A}$ 'the negation of', 'the conjunction of … and …', 'the universal quantification of … with respect to variable …'.

These function expressions can be introduced as new axiom or they can be defined, eg:

Definition

- $x \dot{=} y \stackrel{\text{def}}{=} x \hat{\equiv} y$
- $\dot{\neg} x \stackrel{\text{def}}{=} \overline{\neg} \hat{} x$
- $x \dot{\rightarrow} y \stackrel{\text{def}}{=} \overline{(} \hat{} x \hat{} \overline{\rightarrow} \hat{} y \hat{} \overline{)}$
- $\dot{\forall} x y \stackrel{\text{def}}{=} \overline{\forall} \hat{} x \hat{} y$

Turning the clauses of the definition into axioms

Definition

The theory $\mathcal{D}$ is given by all axioms of $\mathcal{A}$ and the following axioms:

Turning the clauses of the definition into axioms

Definition

The theory $\mathcal{D}$ is given by all axioms of $\mathcal{A}$ and the following axioms:

1. $\forall x \forall y \Big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow \big(\mathrm{True}\,(x \dot{=} y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y) \big) \Big)$

Turning the clauses of the definition into axioms

Definition

The theory $\mathcal{D}$ is given by all axioms of $\mathcal{A}$ and the following axioms:

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (\mathrm{True}\,(x \doteq y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y)) \big)$

   A sentence $s = t$ is true iff the value of s is the value of t, where s and t are closed terms of the language.

Turning the clauses of the definition into axioms

Definition

The theory $\mathcal{D}$ is given by all axioms of $\mathcal{A}$ and the following axioms:

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (\mathrm{True}\,(x \doteq y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y)) \big)$

   A sentence $s = t$ is true iff the value of s is the value of t, where s and t are closed terms of the language.

2. $\forall x \big(\mathrm{ClT}(x) \rightarrow (\mathrm{True}\,(\dot{\mathrm{Sent}}(x)) \leftrightarrow \mathrm{Sent}(\mathrm{val}(x))) \big)$

Turning the clauses of the definition into axioms

Definition

The theory $\mathcal{D}$ is given by all axioms of $\mathcal{A}$ and the following axioms:

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (\mathrm{True}\,(x \dot{=} y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y)) \big)$
   A sentence $s = t$ is true iff the value of s is the value of t, where s and t are closed terms of the language.
2. $\forall x \big(\mathrm{ClT}(x) \rightarrow (\mathrm{True}\,(\dot{\mathrm{Sent}}(x)) \leftrightarrow \mathrm{Sent}(\mathrm{val}(x))) \big)$
   and so on for further predicates of $\mathcal{L}$…
3. $\forall x \big(\mathrm{Sent}(x) \rightarrow (\mathrm{True}\,\dot{\neg} x \leftrightarrow \neg \mathrm{True}\, x) \big)$

Turning the clauses of the definition into axioms

Definition

The theory $\mathcal{D}$ is given by all axioms of $\mathcal{A}$ and the following axioms:

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (\mathrm{True}\,(x \dot{=} y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y)) \big)$
   A sentence $s = t$ is true iff the value of s is the value of t, where s and t are closed terms of the language.
2. $\forall x \big(\mathrm{ClT}(x) \rightarrow (\mathrm{True}\,(\dot{\mathrm{Sent}}(x)) \leftrightarrow \mathrm{Sent}(\mathrm{val}(x))) \big)$
   and so on for further predicates of $\mathcal{L}$…
3. $\forall x \big(\mathrm{Sent}(x) \rightarrow (\mathrm{True}\,\dot{\neg} x \leftrightarrow \neg \mathrm{True}\, x) \big)$
   A negated $\mathcal{L}$-sentence $\neg\varphi$ is true iff φ is not true.

Turning the clauses of the definition into axioms

Definition

The theory $\mathcal{D}$ is given by all axioms of $\mathcal{A}$ and the following axioms:

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \to (\mathrm{True}\,(x \doteq y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y)) \big)$
   A sentence $s = t$ is true iff the value of s is the value of t, where s and t are closed terms of the language.
2. $\forall x \big(\mathrm{ClT}(x) \to (\mathrm{True}\,(\dot{\mathrm{Sent}}(x)) \leftrightarrow \mathrm{Sent}(\mathrm{val}(x))) \big)$
   and so on for further predicates of $\mathcal{L}$…
3. $\forall x \big(\mathrm{Sent}(x) \to (\mathrm{True}\,\dot{\neg} x \leftrightarrow \neg \mathrm{True}\, x) \big)$
   A negated $\mathcal{L}$-sentence $\neg\varphi$ is true iff φ is not true.
4. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \to (\mathrm{True}(x \dot{\to} y) \leftrightarrow (\mathrm{True}\, x \to \mathrm{True}\, y)) \big)$

Turning the clauses of the definition into axioms

Definition

The theory $\mathcal{D}$ is given by all axioms of $\mathcal{A}$ and the following axioms:

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \to (\mathrm{True}\,(x \doteq y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y)) \big)$
   A sentence $s = t$ is true iff the value of s is the value of t, where s and t are closed terms of the language.
2. $\forall x \big(\mathrm{ClT}(x) \to (\mathrm{True}\,(\dot{\mathrm{Sent}}(x)) \leftrightarrow \mathrm{Sent}(\mathrm{val}(x))) \big)$
   and so on for further predicates of $\mathcal{L}$…
3. $\forall x \big(\mathrm{Sent}(x) \to (\mathrm{True}\,\dot{\neg} x \leftrightarrow \neg \mathrm{True}\, x) \big)$
   A negated $\mathcal{L}$-sentence $\neg\varphi$ is true iff φ is not true.
4. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \to (\mathrm{True}(x \dot{\to} y) \leftrightarrow (\mathrm{True}\, x \to \mathrm{True}\, y)) \big)$
   A conditional $\varphi \to \psi$ is true iff φ is false or and ψ is true (φ and ψ are sentences of $\mathcal{L}$)

Turning the clauses of the definition into axioms

Definition

The theory $\mathcal{D}$ is given by all axioms of $\mathcal{A}$ and the following axioms:

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \to (\mathrm{True}\,(x \doteq y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y)) \big)$
   A sentence $s = t$ is true iff the value of s is the value of t, where s and t are closed terms of the language.
2. $\forall x \big(\mathrm{ClT}(x) \to (\mathrm{True}\,(\dot{\mathrm{Sent}}(x)) \leftrightarrow \mathrm{Sent}(\mathrm{val}(x))) \big)$
   and so on for further predicates of $\mathcal{L}$…
3. $\forall x \big(\mathrm{Sent}(x) \to (\mathrm{True}\,\dot{\neg} x \leftrightarrow \neg \mathrm{True}\, x) \big)$
   A negated $\mathcal{L}$-sentence $\neg\varphi$ is true iff φ is not true.
4. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \to (\mathrm{True}(x \dot{\to} y) \leftrightarrow (\mathrm{True}\, x \to \mathrm{True}\, y)) \big)$
   A conditional $\varphi \to \psi$ is true iff φ is false or and ψ is true (φ and ψ are sentences of $\mathcal{L}$)
5. $\forall x \forall y \big(\mathrm{Sent}(\dot{\forall} x y) \to (\mathrm{True}\,(\dot{\forall} x y) \leftrightarrow \forall z \mathrm{True}\, \mathrm{sub}(y, x, \mathrm{q}z)) \big)$

Turning the clauses of the definition into axioms

Definition

The theory $\mathcal{D}$ is given by all axioms of $\mathcal{A}$ and the following axioms:

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \to (\mathrm{True}\,(x \doteq y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y)) \big)$
   A sentence $s = t$ is true iff the value of s is the value of t, where s and t are closed terms of the language.
2. $\forall x \big(\mathrm{ClT}(x) \to (\mathrm{True}\,(\dot{\mathrm{Sent}}(x)) \leftrightarrow \mathrm{Sent}(\mathrm{val}(x))) \big)$
   and so on for further predicates of $\mathcal{L}$…
3. $\forall x \big(\mathrm{Sent}(x) \to (\mathrm{True}\,\dot{\neg} x \leftrightarrow \neg \mathrm{True}\, x) \big)$
   A negated $\mathcal{L}$-sentence $\neg\varphi$ is true iff φ is not true.
4. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \to (\mathrm{True}(x \dot{\to} y) \leftrightarrow (\mathrm{True}\, x \to \mathrm{True}\, y)) \big)$
   A conditional $\varphi \to \psi$ is true iff φ is false or and ψ is true (φ and ψ are sentences of $\mathcal{L}$)
5. $\forall x \forall y \big(\mathrm{Sent}(\dot{\forall} x y) \to (\mathrm{True}\,(\dot{\forall} x y) \leftrightarrow \forall z \mathrm{True}\, \mathrm{sub}(y, x, \mathrm{q}z)) \big)$
   A universally quantified $\mathcal{L}$-sentence $\forall x \varphi(x)$ is true iff $\varphi(\overline{e})$ for all objects e.

Turning the clauses of the definition into axioms

The grey comments on the previous slide are merely the metatheoretic counterparts of the axioms; here is the pure version:

Definition

The theory $\mathcal{D}$ is given by all axioms of $\mathcal{A}$ and the following axioms:

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \to (\mathrm{True}\,(x \doteq y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y)) \big)$
2. $\forall x \big(\mathrm{ClT}(x) \to (\mathrm{True}\,(\dot{\mathrm{Sent}}(x)) \leftrightarrow \mathrm{Sent}(\mathrm{val}(x))) \big)$

   $\ldots$
3. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \to (\mathrm{True}(x \dot{\to} y) \leftrightarrow (\mathrm{True}\, x \to \mathrm{True}\, y)) \big)$
4. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \to (\mathrm{True}\,(x \dot{\to} y) \leftrightarrow (\mathrm{True}\, x \to \mathrm{True}\, y)) \big)$
5. $\forall x \forall y \big(\mathrm{Sent}(\dot{\forall} x y) \to (\mathrm{True}\,(\dot{\forall} x y) \leftrightarrow \forall z \mathrm{True}\, \mathrm{sub}(y, x, \mathrm{q}z)) \big)$

Comments on the axioms of $\mathcal{D}$

- True is not a symbol of $\mathcal{L}$: any non-logical axiom schemata of $\mathcal{A}$ contains only substitution instances from $\mathcal{L}$, that is, without True .

Comments on the axioms of $\mathcal{D}$

- True is not a symbol of $\mathcal{L}$: any non-logical axiom schemata of $\mathcal{A}$ contains only substitution instances from $\mathcal{L}$, that is, without True .
- The last axiom makes use of 'quantifying in'.

Comments on the axioms of $\mathcal{D}$

- True is not a symbol of $\mathcal{L}$: any non-logical axiom schemata of $\mathcal{A}$ contains only substitution instances from $\mathcal{L}$, that is, without True .
- The last axiom makes use of 'quantifying in'.
- Additional axioms for every predicate symbol of $\mathcal{L}$ have to be added. Here I should say something about schematic theories and list definitions…

Comments on the axioms of $\mathcal{D}$

- True is not a symbol of $\mathcal{L}$: any non-logical axiom schemata of $\mathcal{A}$ contains only substitution instances from $\mathcal{L}$, that is, without True .
- The last axiom makes use of 'quantifying in'.
- Additional axioms for every predicate symbol of $\mathcal{L}$ have to be added. Here I should say something about schematic theories and list definitions…
- The axioms capture a *compositional* conception of truth.

Comments on the axioms of $\mathcal{D}$

- True is not a symbol of $\mathcal{L}$: any non-logical axiom schemata of $\mathcal{A}$ contains only substitution instances from $\mathcal{L}$, that is, without True .
- The last axiom makes use of 'quantifying in'.
- Additional axioms for every predicate symbol of $\mathcal{L}$ have to be added. Here I should say something about schematic theories and list definitions…
- The axioms capture a *compositional* conception of truth.
- I suppose that these are the axioms for truth Davidson alluded to when talking about turning the definitional clauses of truth into axioms, although there are some open questions…

Comments on the axioms of $\mathcal{D}$

- True is not a symbol of $\mathcal{L}$: any non-logical axiom schemata of $\mathcal{A}$ contains only substitution instances from $\mathcal{L}$, that is, without True .
- The last axiom makes use of 'quantifying in'.
- Additional axioms for every predicate symbol of $\mathcal{L}$ have to be added. Here I should say something about schematic theories and list definitions…
- The axioms capture a *compositional* conception of truth.
- I suppose that these are the axioms for truth Davidson alluded to when talking about turning the definitional clauses of truth into axioms, although there are some open questions…
- $\mathcal{D}$ proves many of the desired generalisations such as $\forall x(\mathrm{Sent}(x) \rightarrow (\mathrm{True}\, x \vee \mathrm{True}\, \dot{\neg} x))$

Deflationism and conservativeness

Some deflationists wrt truth are keen on a theory of truth that proves generalisations but that is at the same time 'insubstantial' in the sense that it is still conservative over the base theory $\mathcal{A}$, that is, it doesn't prove any new sentences in the original language $\mathcal{L}$ (ie, the language without True) (cf Shapiro (1998), Field (1999), Halbach (1999a), Ketland (1999)).

Deflationism and conservativeness

Some deflationists wrt truth are keen on a theory of truth that proves generalisations but that is at the same time 'insubstantial' in the sense that it is still conservative over the base theory $\mathcal{A}$, that is, it doesn't prove any new sentences in the original language $\mathcal{L}$ (ie, the language without True) (cf Shapiro (1998), Field (1999), Halbach (1999a), Ketland (1999)).

So we would like to show for all sentences φ of $\mathcal{L}$:

If $\mathcal{D} \vdash \varphi$, *then* $\mathcal{A} \vdash \varphi$.

Models

Several people have claimed that this can be shown by proving that any model of $\mathcal{A}$ can be extended to a model of $\mathcal{D}$.

Models

Several people have claimed that this can be shown by proving that any model of $\mathcal{A}$ can be extended to a model of $\mathcal{D}$.

It follows from a result by Lachlan (1981) that this isn't feasible:

Theorem (Lachlan (1981))

Take $\mathcal{A}$ to be PA. If $\mathcal{M}$ is a model of $\mathcal{A}$ that can be be extended to a model of D, then $\mathcal{M}$ is recursively saturated or the standard model.

A set of of formulae is recursively saturated iff every recursive finitely satisfiable set of finitely many variables is globally satisfiable.

Models

Several people have claimed that this can be shown by proving that any model of $\mathcal{A}$ can be extended to a model of $\mathcal{D}$.

It follows from a result by Lachlan (1981) that this isn't feasible:

Theorem (Lachlan (1981))

Take $\mathcal{A}$ to be PA. If $\mathcal{M}$ is a model of $\mathcal{A}$ that can be be extended to a model of D, then $\mathcal{M}$ is recursively saturated or the standard model.

A set of of formulae is recursively saturated iff every recursive finitely satisfiable set of finitely many variables is globally satisfiable.

Discuss nonstandard models of $\mathcal{A}$; existence of nonstandard sentences.

Models

Several people have claimed that this can be shown by proving that any model of $\mathcal{A}$ can be extended to a model of $\mathcal{D}$.

It follows from a result by Lachlan (1981) that this isn't feasible:

Theorem (Lachlan (1981))

Take $\mathcal{A}$ to be PA. If $\mathcal{M}$ is a model of $\mathcal{A}$ that can be be extended to a model of D, then $\mathcal{M}$ is recursively saturated or the standard model.

A set of of formulae is recursively saturated iff every recursive finitely satisfiable set of finitely many variables is globally satisfiable.

Discuss nonstandard models of $\mathcal{A}$; existence of nonstandard sentences.

This result dooms all attempts to prove conservativeness by extending any given model of $\mathcal{A}$ to a model of $\mathcal{D}$.

Conservativeness

Theorem (Kotlarski et al. 1981)

Every countable recursively saturated model of PA can be extended to a model of $\mathcal{D}$.

Conservativeness

Theorem (Kotlarski et al. 1981)

Every countable recursively saturated model of PA can be extended to a model of $\mathcal{D}$.

That is sufficient for establishing conservativeness:

Theorem

If $\mathcal{D} \vdash \varphi$, then $\mathcal{A} \vdash \varphi$.

Conservativeness

Theorem (Kotlarski et al. 1981)

Every countable recursively saturated model of PA can be extended to a model of $\mathcal{D}$.

That is sufficient for establishing conservativeness:

Theorem

If $\mathcal{D} \vdash \varphi$, then $\mathcal{A} \vdash \varphi$.

Remark (Smith 1987?)

Countability of the model is required.

Conservativeness via cut elimination

McGee (2003) complained that the model-theoretic proof of conservativeness is of little use to the deflationist: he ought to be able to prove conservativeness in his theory $\mathcal{A}$. At least for sufficiently strong $\mathcal{A}$ we get:

Conservativeness via cut elimination

McGee (2003) complained that the model-theoretic proof of conservativeness is of little use to the deflationist: he ought to be able to prove conservativeness in his theory $\mathcal{A}$. At least for sufficiently strong $\mathcal{A}$ we get:

Theorem (Leigh 2015, Enayat and Visser 2015, but not Halbach 1999b)

If $\mathcal{D} \vdash \varphi$, then $\mathcal{A} \vdash \varphi$. Moreover, this implication can be proved in PRA.

Conservativeness via cut elimination

McGee (2003) complained that the model-theoretic proof of conservativeness is of little use to the deflationist: he ought to be able to prove conservativeness in his theory $\mathcal{A}$. At least for sufficiently strong $\mathcal{A}$ we get:

Theorem (Leigh 2015, Enayat and Visser 2015, but not Halbach 1999b)

If $\mathcal{D} \vdash \varphi$, then $\mathcal{A} \vdash \varphi$. Moreover, this implication can be proved in PRA.

The theory is still fairly weak. It doesn't prove that all sentence $\overline{\neg} = \overline{\neg} \wedge \overline{\neg} = \overline{\neg} \wedge \overline{\neg} = \overline{\neg} \ldots$ are true.

Global reflection (cf Kreisel and Lévy 1968)

We can use the truth predicate to state soundness of theories. For instance, we might try to prove

$$\forall x\big(\mathrm{Sent}(x) \wedge \mathrm{Bew}_{\mathcal{A}}(x) \to \mathrm{True}\, x\big)$$

Global reflection (cf Kreisel and Lévy 1968)

We can use the truth predicate to state soundness of theories. For instance, we might try to prove

$$\forall x\big(\mathrm{Sent}(x) \wedge \mathrm{Bew}_{\mathcal{A}}(x) \to \mathrm{True}\, x\big)$$

The obvious strategy is to prove first that all axioms of $\mathcal{A}$ are true and that all rules of inference are true preserving; then by induction on the length of proofs in $\mathcal{A}$, one concludes that *all* theorems of $\mathcal{A}$ are true.

Global reflection (cf Kreisel and Lévy 1968)

We can use the truth predicate to state soundness of theories. For instance, we might try to prove

$$\forall x\big(\mathrm{Sent}(x) \wedge \mathrm{Bew}_{\mathcal{A}}(x) \to \mathrm{True}\, x\big)$$

The obvious strategy is to prove first that all axioms of $\mathcal{A}$ are true and that all rules of inference are true preserving; then by induction on the length of proofs in $\mathcal{A}$, one concludes that *all* theorems of $\mathcal{A}$ are true.

What we lack now is some induction principle…

Arithmetic in $\mathcal{A}$

Strings vvv... can be used as natural numbers, and I call expressions $\overline{\text{vvv}\ldots}$ *numerals* because they act as constants for numbers. I shall write $\overline{n}$ for $\overline{\underbrace{\text{v}\ldots\text{v}}_{n}}$ and call it the numeral for n. For instance, $\overline{4}$ is the numeral for 4 and stands for $\overline{\text{vvvv}}$.

Definition

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

The idea is that substituting the empty string for v in a string of v's gives the empty string. If the original string had contained some symbol different from v, the symbol would be left. The empty string is the empty string of v's, so it is a natural number and this is provable in $\mathcal{A}$.

Arithmetic in $\mathcal{A}$

Lemma

$\mathcal{A} \vdash \mathrm{Nat}(\overline{n})$ *for all natural numbers* n.

Proof.

$\mathrm{sub}(\underline{0}, \overline{\mathrm{v}}, \underline{0}) = \underline{0}$ is an instance of A4.

Arithmetic in $\mathcal{A}$

Lemma

$\mathcal{A} \vdash \mathrm{Nat}(\overline{n})$ *for all natural numbers* n.

Proof.

$\mathrm{sub}(\underline{0}, \overline{\mathrm{v}}, \underline{0}) = \underline{0}$ is an instance of A4.

Then proceed inductively using A8. ⊣

Arithmetic in $\mathcal{A}$

Lemma

$\mathcal{A} \vdash \mathrm{Nat}(\overline{n})$ *for all natural numbers* n.

Proof.

$\mathrm{sub}(\underline{0}, \overline{\mathrm{v}}, \underline{0}) = \underline{0}$ is an instance of A4.

Then proceed inductively using A8. ⊣

I write $\forall n \varphi(n)$ for $\forall x(\mathrm{Nat}(x) \leftrightarrow \varphi(x))$ and similarly $\exists n \varphi(n)$ for $\exists x(\mathrm{Nat}(x) \wedge \varphi(x))$.

Addition and multiplication

Addition and multiplication can be mimicked in $\mathcal{L}$ by concatenation and substitution, respectively.

Lemma

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

1. $\mathcal{A} \vdash \overline{n}^\frown \overline{k} = \overline{n+k}$
2. $\mathcal{A} \vdash \mathrm{sub}(\overline{n}, \overline{v}, \overline{k}) = \overline{n \cdot k}$

In particular we have $\mathcal{A} \vdash \overline{n}^\frown \overline{1} = \overline{n+1}$.

Induction

Now we can add induction in the full language with True to $\mathcal{D}$:

Definition

$\mathcal{D}^+$ is the theory $\mathcal{D}$ plus all of the following axioms:

$$\varphi(\overline{0}) \wedge \forall n(\varphi(n) \rightarrow \varphi(n+\overline{1})) \rightarrow \forall n \varphi(n)$$

Actually I am adding to $\mathcal{A}$ also the axioms for a predicate symbol $Cons(x)$ expressing that x is a quotational constant.

Induction

Now we can add induction in the full language with True to $\mathcal{D}$:

Definition

$\mathcal{D}^+$ is the theory $\mathcal{D}$ plus all of the following axioms:

$$\varphi(\overline{0}) \wedge \forall n(\varphi(n) \rightarrow \varphi(n + \overline{1})) \rightarrow \forall n \varphi(n)$$

Actually I am adding to $\mathcal{A}$ also the axioms for a predicate symbol $Cons(x)$ expressing that x is a quotational constant.

That will allow to define notions such as 'x is a proof'.

Proving consistency

Now with a suitable definition of $\mathrm{Bew}_{\mathcal{A}}$ and some additional axioms in $\mathcal{A}$ we are able to prove the following in $\mathcal{D}^{+}$:

$$\forall x\big(\mathrm{Sent}(x) \wedge \mathrm{Bew}_{\mathcal{A}}(x) \to \mathrm{True}\, x\big)$$

Proving consistency

Now with a suitable definition of $\mathrm{Bew}_{\mathcal{A}}$ and some additional axioms in $\mathcal{A}$ we are able to prove the following in $\mathcal{D}^{+}$:

$$\forall x\big(\mathrm{Sent}(x) \wedge \mathrm{Bew}_{\mathcal{A}}(x) \to \mathrm{True}\, x\big)$$

In particular, we'll be able to prove

$$\mathrm{Sent}(\overline{\bot}) \wedge \mathrm{Bew}_{\mathcal{A}}(\overline{\bot}) \to \mathrm{True}\, \overline{\bot}$$

where $\bot$ is some contradiction, and therefore

Proving consistency

Now with a suitable definition of $\mathrm{Bew}_{\mathcal{A}}$ and some additional axioms in $\mathcal{A}$ we are able to prove the following in $\mathcal{D}^{+}$:

$$\forall x\big(\mathrm{Sent}(x) \wedge \mathrm{Bew}_{\mathcal{A}}(x) \to \mathrm{True}\, x\big)$$

In particular, we'll be able to prove

$$\mathrm{Sent}(\overline{\bot}) \wedge \mathrm{Bew}_{\mathcal{A}}(\overline{\bot}) \to \mathrm{True}\, \overline{\bot}$$

where $\bot$ is some contradiction, and therefore

$$\neg \mathrm{Bew}_{\mathcal{A}}(\overline{\bot})$$

which isn't derivable in $\mathcal{A}$ by Gödel's second incompleteness theorem, at least if $\mathrm{Bew}_{\mathcal{A}}$ is well behaved.

How much can one do in $\mathcal{D}^+$?

We know that $\mathcal{D}^+$ is not conservative over $\mathcal{A}$. But one can do much more than just proving consistency.

Theorem

$\mathcal{D}^+$ proves $\exists X \forall n (n \in X \leftrightarrow \varphi(n))$, where $\varphi(n)$ is a formula possibly containing subformulae of the form $z \in Y$ where Y is free and where there are no further occurrences of the truth predicate in $\varphi(n)$.

How much can one do in $\mathcal{D}^+$?

We know that $\mathcal{D}^+$ is not conservative over $\mathcal{A}$. But one can do much more than just proving consistency.

Theorem

$\mathcal{D}^+$ proves $\exists X \forall n (n \in X \leftrightarrow \varphi(n))$, where $\varphi(n)$ is a formula possibly containing subformulae of the form $z \in Y$ where Y is free and where there are no further occurrences of the truth predicate in $\varphi(n)$.

In the proof the 'Tarski clauses' are needed to handle the parameters Y.

How much can one do in $\mathcal{D}^+$?

We know that $\mathcal{D}^+$ is not conservative over $\mathcal{A}$. But one can do much more than just proving consistency.

Theorem

$\mathcal{D}^+$ proves $\exists X \forall n(n \in X \leftrightarrow \varphi(n))$, where $\varphi(n)$ is a formula possibly containing subformulae of the form $z \in Y$ where Y is free and where there are no further occurrences of the truth predicate in $\varphi(n)$.

In the proof the 'Tarski clauses' are needed to handle the parameters Y.

'Tarskian truth is as strong as arithmetical comprehension.'

How much can one do in $\mathcal{D}^+$?

We know that $\mathcal{D}^+$ is not conservative over $\mathcal{A}$. But one can do much more than just proving consistency.

Theorem

$\mathcal{D}^+$ proves $\exists X \forall n(n \in X \leftrightarrow \varphi(n))$, where $\varphi(n)$ is a formula possibly containing subformulae of the form $z \in Y$ where Y is free and where there are no further occurrences of the truth predicate in $\varphi(n)$.

In the proof the 'Tarski clauses' are needed to handle the parameters Y.

'Tarskian truth is as strong as arithmetical comprehension.'

Thus is does not only prove global reflection but much more.

Truth & Paradox

V · Kripke's Theory of Truth

Volker Halbach

Nordic Logic Summer School 2017

Type restrictions

The theories TB, UTB, and $\mathcal{D}$ are *typed* theories in the sense that there is no sentence φ containing the predicate True such that, eg, $\mathcal{D} \vdash \text{True}\,\overline{\varphi}$.

Type restrictions

The theories TB, UTB, and $\mathcal{D}$ are *typed* theories in the sense that there is no sentence φ containing the predicate True such that, eg, $\mathcal{D} \vdash$ True $\overline{\varphi}$.

TB, UTB, and $\mathcal{D}$ don't say anything about sentences containing True , so they are not incompatible with a theory proving claims about the truth of sentences containing True . This is an advantage of the axiomatic approach, because on the semantic account one would have to decide for any sentence whether they are in the extension of True or not.

Hierarchical solutions

Typed truth theories can be iterated, ie, one can introduce new predicates True_1, True_2, …, True_ω, $\text{True}_{\omega+1}$, …, True_{ϵ_0} as far as you can count… At any level one could use axioms analogous to TB, UTB, and $\mathcal{D}$ or pursue a semantic approach using Tarski's definition of truth.

Hierarchical solutions

Typed truth theories can be iterated, ie, one can introduce new predicates True_1, True_2, …, True_ω, $\text{True}_{\omega+1}$, …, True_{ϵ_0} as far as you can count… At any level one could use axioms analogous to TB, UTB, and $\mathcal{D}$ or pursue a semantic approach using Tarski's definition of truth.

I did it up to ω_1^{CK}, that is, as far one can count recursively. Beyond that point the languages become non-recursive, and pushing the hierarchy further becomes an exercise in counting rather than in the theory of truth.

Hierarchical solutions

Typed truth theories can be iterated, ie, one can introduce new predicates True_1, True_2, …, True_ω, $\text{True}_{\omega+1}$, …, True_{ϵ_0} as far as you can count… At any level one could use axioms analogous to TB, UTB, and $\mathcal{D}$ or pursue a semantic approach using Tarski's definition of truth.

I did it up to ω_1^{CK}, that is, as far one can count recursively. Beyond that point the languages become non-recursive, and pushing the hierarchy further becomes an exercise in counting rather than in the theory of truth.

Kripke proposed to look at ill-founded hierarchies of truth predicates, eg, True_n applies to all sentences containing truth predicates True_i with $i > n$. Visser (1989a) axiomatised truth in such a language by axioms analogous to TB and showed that the resulting theory is ω-inconsistent.

Iterating truth without typing

Anyway, we would like to say more about the truth of sentences containing the very same truth predicate. Therefore I return to the type-free truth predicate T contained in the language $\mathcal{L}$. Clearly there are sentences that ought to be true, although they contain T:

- $T\overline{\forall x\, x = x}$

Iterating truth without typing

Anyway, we would like to say more about the truth of sentences containing the very same truth predicate. Therefore I return to the type-free truth predicate T contained in the language $\mathcal{L}$. Clearly there are sentences that ought to be true, although they contain T:

- $T\overline{\forall x\, x = x}$
- $T\overline{T\overline{\forall x\, x = x}}$

Iterating truth without typing

Anyway, we would like to say more about the truth of sentences containing the very same truth predicate. Therefore I return to the type-free truth predicate T contained in the language $\mathcal{L}$. Clearly there are sentences that ought to be true, although they contain T:

- $T\overline{\forall x\, x = x}$
- $T\overline{T\overline{\forall x\, x = x}}$
- $T\overline{T}^{\frown}\mathrm{q}\,\overline{\forall x\, x = x}$

Declaring these sentences true seems unproblematic, as they are not self-referential in any way.

Defining type-free truth

Assume we have a model $\mathcal{M}$ for the language without T; we try to define type-free truth based on that model.

Defining type-free truth

Assume we have a model $\mathcal{M}$ for the language without T; we try to define type-free truth based on that model.

Of course we know what to do with sentences not containing T: they should get into the extension of T iff they are true in $\mathfrak{M}$; otherwise their negation should be in.

Defining type-free truth

Assume we have a model $\mathcal{M}$ for the language without T; we try to define type-free truth based on that model.

Of course we know what to do with sentences not containing T: they should get into the extension of T iff they are true in $\mathfrak{M}$; otherwise their negation should be in. Moreover, if φ and ψ are already in, they are 'safe' and their conjunction $\varphi \wedge \psi$ should be thrown into the extension S and so on.

Defining type-free truth

Assume we have a model $\mathcal{M}$ for the language without T; we try to define type-free truth based on that model.

Of course we know what to do with sentences not containing T: they should get into the extension of T iff they are true in $\mathfrak{M}$; otherwise their negation should be in. Moreover, if φ and ψ are already in, they are 'safe' and their conjunction $\varphi \wedge \psi$ should be thrown into the extension S and so on.

So assume we have a model $\mathcal{M}$ for the language without T. I proceed in the style of Tarski (with a certain twist), but add the truth predicate itself to the object language.

Defining type-free truth

Definition

A set S of $\mathcal{L}$-sentences is a Kripke fixed-point iff for all sentences φ and ψ the following holds:

1. $s = t$ is true in $\mathcal{M}$ iff $(s = t) \in S$.

Defining type-free truth

Definition

A set S of $\mathcal{L}$-sentences is a Kripke fixed-point iff for all sentences φ and ψ the following holds:

1. $s = t$ is true in $\mathcal{M}$ iff $(s = t) \in S$. Similarly for all other atomic sentences without T: φ is true in $\mathcal{M}$ iff $\varphi \in S$.

Defining type-free truth

Definition

A set S of $\mathcal{L}$-sentences is a Kripke fixed-point iff for all sentences φ and ψ the following holds:

1. $s = t$ is true in $\mathcal{M}$ iff $(s = t) \in S$. Similarly for all other atomic sentences without T: φ is true in $\mathcal{M}$ iff $\varphi \in S$.
2. $s = t$ is false in $\mathcal{M}$ iff $(\neg s = t) \in S$.

Defining type-free truth

Definition

A set S of $\mathcal{L}$-sentences is a Kripke fixed-point iff for all sentences φ and ψ the following holds:

1. $s = t$ is true in $\mathcal{M}$ iff $(s = t) \in S$. Similarly for all other atomic sentences without T: φ is true in $\mathcal{M}$ iff $\varphi \in S$.
2. $s = t$ is false in $\mathcal{M}$ iff $(\neg s = t) \in S$. Similarly for all other atomic sentences without T: $\neg\varphi$ is true in $\mathcal{M}$ iff $\neg\varphi \in S$.

Defining type-free truth

Definition

A set S of $\mathcal{L}$-sentences is a Kripke fixed-point iff for all sentences φ and ψ the following holds:

1. $s = t$ is true in $\mathcal{M}$ iff $(s = t) \in S$. Similarly for all other atomic sentences without T: φ is true in $\mathcal{M}$ iff $\varphi \in S$.
2. $s = t$ is false in $\mathcal{M}$ iff $(\neg s = t) \in S$. Similarly for all other atomic sentences without T: $\neg\varphi$ is true in $\mathcal{M}$ iff $\neg\varphi \in S$.
3. ($\varphi \in S$ and $\psi \in S$) iff $(\varphi \wedge \psi) \in S$.

Defining type-free truth

Definition

A set S of $\mathcal{L}$-sentences is a Kripke fixed-point iff for all sentences φ and ψ the following holds:

1. $s = t$ is true in $\mathcal{M}$ iff $(s = t) \in S$. Similarly for all other atomic sentences without T: φ is true in $\mathcal{M}$ iff $\varphi \in S$.
2. $s = t$ is false in $\mathcal{M}$ iff $(\neg s = t) \in S$. Similarly for all other atomic sentences without T: $\neg\varphi$ is true in $\mathcal{M}$ iff $\neg\varphi \in S$.
3. ($\varphi \in S$ and $\psi \in S$) iff $(\varphi \wedge \psi) \in S$.
4. ($\neg\varphi \in S$ or $\neg\psi \in S$) iff $(\neg(\varphi \wedge \psi)) \in S$.

Defining type-free truth

Definition

A set S of $\mathcal{L}$-sentences is a Kripke fixed-point iff for all sentences φ and ψ the following holds:

1. $s = t$ is true in $\mathcal{M}$ iff $(s = t) \in S$. Similarly for all other atomic sentences without T: φ is true in $\mathcal{M}$ iff $\varphi \in S$.
2. $s = t$ is false in $\mathcal{M}$ iff $(\neg s = t) \in S$. Similarly for all other atomic sentences without T: $\neg\varphi$ is true in $\mathcal{M}$ iff $\neg\varphi \in S$.
3. ($\varphi \in S$ and $\psi \in S$) iff $(\varphi \wedge \psi) \in S$.
4. ($\neg\varphi \in S$ or $\neg\psi \in S$) iff $(\neg(\varphi \wedge \psi)) \in S$.
5. $\varphi \in S$ iff$(\neg\neg\varphi) \in S$.

Defining type-free truth

Definition

A set S of $\mathcal{L}$-sentences is a Kripke fixed-point iff for all sentences φ and ψ the following holds:

1. $s = t$ is true in $\mathcal{M}$ iff $(s = t) \in S$. Similarly for all other atomic sentences without T: φ is true in $\mathcal{M}$ iff $\varphi \in S$.
2. $s = t$ is false in $\mathcal{M}$ iff $(\neg s = t) \in S$. Similarly for all other atomic sentences without T: $\neg\varphi$ is true in $\mathcal{M}$ iff $\neg\varphi \in S$.
3. ($\varphi \in S$ and $\psi \in S$) iff $(\varphi \wedge \psi) \in S$.
4. ($\neg\varphi \in S$ or $\neg\psi \in S$) iff $(\neg(\varphi \wedge \psi)) \in S$.
5. $\varphi \in S$ iff$(\neg\neg\varphi) \in S$.
6. $\varphi(\overline{e}) \in S$ for all objects e , iff $\forall x \varphi(x) \in S$.

Defining type-free truth

Definition

A set S of $\mathcal{L}$-sentences is a Kripke fixed-point iff for all sentences φ and ψ the following holds:

1. $s = t$ is true in $\mathcal{M}$ iff $(s = t) \in S$. Similarly for all other atomic sentences without T: φ is true in $\mathcal{M}$ iff $\varphi \in S$.
2. $s = t$ is false in $\mathcal{M}$ iff $(\neg s = t) \in S$. Similarly for all other atomic sentences without T: $\neg\varphi$ is true in $\mathcal{M}$ iff $\neg\varphi \in S$.
3. ($\varphi \in S$ and $\psi \in S$) iff $(\varphi \wedge \psi) \in S$.
4. ($\neg\varphi \in S$ or $\neg\psi \in S$) iff $(\neg(\varphi \wedge \psi)) \in S$.
5. $\varphi \in S$ iff$(\neg\neg\varphi) \in S$.
6. $\varphi(\overline{e}) \in S$ for all objects e , iff $\forall x\varphi(x) \in S$.
7. $\neg\varphi(\overline{e}) \in S$ for some object e , iff $(\neg\forall x\varphi(x)) \in S$.

Defining type-free truth

Definition

A set S of $\mathcal{L}$-sentences is a Kripke fixed-point iff for all sentences φ and ψ the following holds:

1. $s = t$ is true in $\mathcal{M}$ iff $(s = t) \in S$. Similarly for all other atomic sentences without T: φ is true in $\mathcal{M}$ iff $\varphi \in S$.
2. $s = t$ is false in $\mathcal{M}$ iff $(\neg s = t) \in S$. Similarly for all other atomic sentences without T: $\neg\varphi$ is true in $\mathcal{M}$ iff $\neg\varphi \in S$.
3. ($\varphi \in S$ and $\psi \in S$) iff $(\varphi \wedge \psi) \in S$.
4. ($\neg\varphi \in S$ or $\neg\psi \in S$) iff $(\neg(\varphi \wedge \psi)) \in S$.
5. $\varphi \in S$ iff$(\neg\neg\varphi) \in S$.
6. $\varphi(\overline{e}) \in S$ for all objects e , iff $\forall x\varphi(x) \in S$.
7. $\neg\varphi(\overline{e}) \in S$ for some object e , iff $(\neg\forall x\varphi(x)) \in S$.
8. If $\varphi \in S$ and the value of the term t is φ, iff $(Tt) \in S$.

Defining type-free truth

Definition

A set S of $\mathcal{L}$-sentences is a Kripke fixed-point iff for all sentences φ and ψ the following holds:

1. $s = t$ is true in $\mathcal{M}$ iff $(s = t) \in S$. Similarly for all other atomic sentences without T: φ is true in $\mathcal{M}$ iff $\varphi \in S$.
2. $s = t$ is false in $\mathcal{M}$ iff $(\neg s = t) \in S$. Similarly for all other atomic sentences without T: $\neg\varphi$ is true in $\mathcal{M}$ iff $\neg\varphi \in S$.
3. ($\varphi \in S$ and $\psi \in S$) iff $(\varphi \wedge \psi) \in S$.
4. ($\neg\varphi \in S$ or $\neg\psi \in S$) iff $(\neg(\varphi \wedge \psi)) \in S$.
5. $\varphi \in S$ iff$(\neg\neg\varphi) \in S$.
6. $\varphi(\overline{e}) \in S$ for all objects e , iff $\forall x\varphi(x) \in S$.
7. $\neg\varphi(\overline{e}) \in S$ for some object e , iff $(\neg\forall x\varphi(x)) \in S$.
8. If $\varphi \in S$ and the value of the term t is φ, iff $(Tt) \in S$.
9. If $\neg\varphi \in S$ and the value of the term t is φ, iff $(\neg Tt) \in S$.

Comments on the definition

- Up to item 7 this is a ‘Tarskian’ definition of truth.

Comments on the definition

- Up to item 7 this is a 'Tarskian' definition of truth.
- 8 and 9 imply the following
  - $\varphi \in S$ iff $T\overline{\varphi} \in S$.

Comments on the definition

- Up to item 7 this is a 'Tarskian' definition of truth.
- 8 and 9 imply the following
  - $\varphi \in S$ iff $T\overline{\varphi} \in S$.
  - If $\neg\varphi \in S$ iff $\neg T\overline{\varphi} \in S$.
- But why having different clauses for $\varphi \wedge \psi$ and $\neg(\varphi \wedge \psi)$?

Comments on the definition

- Up to item 7 this is a 'Tarskian' definition of truth.
- 8 and 9 imply the following
  - $\varphi \in S$ iff $T\overline{\varphi} \in S$.
  - If $\neg\varphi \in S$ iff $\neg T\overline{\varphi} \in S$.
- But why having different clauses for $\varphi \wedge \psi$ and $\neg(\varphi \wedge \psi)$?

to answer that question we need to look at the proof for the existence of Kripke fixed-points.

Existence of fixed points

Theorem

For any $\mathcal{M}$ satisfying the base theory there are Kripke fixed-points.

I don't know who proved this first. It seems Kripke (1975), Martin and Woodruff (1975), and Cantini (mid 70ies) all proved similar results independently. The most notable contribution of Kripke consists in a systematic investigation into the structure of the fixed points (minimal, intrinsic, and maximal fixed points) as well as in the application to various evaluation schemata.

Existence of fixed points: the proof

Proof.

Let $\mathcal{M}$ be given. Start with $S = \emptyset$ and expand S successively by applying the clauses in the definition read from left to right only, eg,

- If $s = t$ is true in $\mathcal{M}$ then $(s = t) \in S$.
- If $s = t$ is false in $\mathcal{M}$ then $(\neg s = t) \in S$.
- If $\varphi \in S$ and $\psi \in S$ then $(\varphi \wedge \psi) \in S$.
- If $\neg\varphi \in S$ or $\neg\psi \in S$ then $(\neg(\varphi \wedge \psi)) \in S$.
- If $\varphi \in S$ and the value of the term t is φ, then $(Tt) \in S$.

$\dashv$

Existence of fixed points: the proof

Proof.

For cardinality reason there must be as stage where this procedure reaches a fixed point – given that our language including all constants has some cardinality, ie, given that our language together with constants for all objects isn't a proper class.

Existence of fixed points: the proof

Proof.

For cardinality reason there must be as stage where this procedure reaches a fixed point – given that our language including all constants has some cardinality, ie, given that our language together with constants for all objects isn't a proper class.

This yields a procedure for generating an S satisfying all the clauses of the definition.

Existence of fixed points: the proof

Proof.

For cardinality reason there must be as stage where this procedure reaches a fixed point – given that our language including all constants has some cardinality, ie, given that our language together with constants for all objects isn't a proper class.

This yields a procedure for generating an S satisfying all the clauses of the definition.

The fixed point obtained in this way is *minimal*. ⊣

Starting with other sets

One could also start with any other set of $\mathcal{L}$-sentences and then close it in the way described above.

Starting with other sets

One could also start with any other set of $\mathcal{L}$-sentences and then close it in the way described above.

In some cases this will not yields a Kripke fixed-point, eg if a sentence false in $\mathcal{M}$ is in the starting set.

Starting with other sets

One could also start with any other set of $\mathcal{L}$-sentences and then close it in the way described above.

In some cases this will not yields a Kripke fixed-point, eg if a sentence false in $\mathcal{M}$ is in the starting set.

Starting, eg with the truth teller yields another Kripke fixed point. The truth teller is a sentence τ such that $\mathcal{A} \vdash \tau \leftrightarrow \neg\tau\overline{\tau}$

Fixed points

- But why should we have different clauses for $\varphi \wedge \psi$ and $\neg(\varphi \wedge \psi)$? why don't we have a simple clause for $\neg$?

Fixed points

- But why should we have different clauses for $\varphi \wedge \psi$ and $\neg(\varphi \wedge \psi)$? why don't we have a simple clause for $\neg$?
- If we had a clause

  $$\textit{If } \varphi \notin S, \textit{ then } \neg\varphi \in S.$$

  we would have to add $\neg\varphi$ in the first step and then, once φ had been added we would have to remove $\neg\varphi$ again. So sentences might flip in and out of S.

Fixed points

- But why should we have different clauses for $\varphi \wedge \psi$ and $\neg(\varphi \wedge \psi)$? why don't we have a simple clause for $\neg$?
- If we had a clause

  $$\textit{If } \varphi \notin S, \textit{ then } \neg\varphi \in S.$$

  we would have to add $\neg\varphi$ in the first step and then, once φ had been added we would have to remove $\neg\varphi$ again. So sentences might flip in and out of S.
- Because S becomes larger and larger by applying the clauses, there must be a stage where all the sentences that can be added have been added.

Properties of fixed points

As a surrogate for the T-sentences we obtain the following metatheoretic equivalence:

Theorem

$\varphi \in S$ *iff* $T\overline{\varphi} \in S$.

This is a consequence of clause 8.

The liar

Theorem

Kripke fixed-points are not closed under classical logic.

The liar

Theorem

Kripke fixed-points are not closed under classical logic.

Proof.

The liar sentence is a sentence of the form $\neg T t$ where the value of t is $\neg T t$ itself. I distinguish two cases:

The liar

Theorem

Kripke fixed-points are not closed under classical logic.

Proof.

The liar sentence is a sentence of the form $\neg Tt$ where the value of t is $\neg Tt$ itself. I distinguish two cases:

1. $Tt \in S$. Then $\neg Tt \in S$ by clause 9, ie S contains a sentence together with its negation.

The liar

Theorem

Kripke fixed-points are not closed under classical logic.

Proof.

The liar sentence is a sentence of the form $\neg Tt$ where the value of t is $\neg Tt$ itself. I distinguish two cases:

1. $Tt \in S$. Then $\neg Tt \in S$ by clause 9, ie S contains a sentence together with its negation.
2. $Tt \notin S$. Then $\neg Tt \notin S$ by clause 9, ie S contains neither the liar nor its negation.

The liar

Theorem

Kripke fixed-points are not closed under classical logic.

Proof.

The liar sentence is a sentence of the form $\neg Tt$ where the value of t is $\neg Tt$ itself. I distinguish two cases:

1. $Tt \in S$. Then $\neg Tt \in S$ by clause 9, ie S contains a sentence together with its negation.
2. $Tt \notin S$. Then $\neg Tt \notin S$ by clause 9, ie S contains neither the liar nor its negation.

$\dashv$

The logic of Kripke fixed-points

So any Kripke fixed-point S either contains the liar sentence together with its negation (truth value glut) or it contains neither the liar or its negation (truth value gap).

The logic of Kripke fixed-points

So any Kripke fixed-point S either contains the liar sentence together with its negation (truth value glut) or it contains neither the liar or its negation (truth value gap).

The minimal fixed-point contains neither the liar or its negation. If one starts with the singleton of the liar sentence and closes off under the Kripke clauses, then one obtains a fixed point containing the liar sentence together with its negation.

The logic of Kripke fixed-points

So any Kripke fixed-point S either contains the liar sentence together with its negation (truth value glut) or it contains neither the liar or its negation (truth value gap).

The minimal fixed-point contains neither the liar or its negation. If one starts with the singleton of the liar sentence and closes off under the Kripke clauses, then one obtains a fixed point containing the liar sentence together with its negation.

In general, so far nothing rules out 'inconsistent' fixed points, ie Kripke fixed-points S such that there is a sentence φ with $\varphi \in S$ and $\neg\varphi \in S$ (truth value 'gluts'). One can describe Kripke fixed points as sets closed under a four-valued logic (true, false, gap, and glut). See Visser (1989a).

The logic of Kripke fixed-points

So any Kripke fixed-point S either contains the liar sentence together with its negation (truth value glut) or it contains neither the liar or its negation (truth value gap).

The minimal fixed-point contains neither the liar or its negation. If one starts with the singleton of the liar sentence and closes off under the Kripke clauses, then one obtains a fixed point containing the liar sentence together with its negation.

In general, so far nothing rules out 'inconsistent' fixed points, ie Kripke fixed-points S such that there is a sentence φ with $\varphi \in S$ and $\neg\varphi \in S$ (truth value 'gluts'). One can describe Kripke fixed points as sets closed under a four-valued logic (true, false, gap, and glut). See Visser (1989a).

Kripke (1975) focused on consistent fixed points, ie Kripke fixed-points without truth-value gluts.

The logic of consistent Kripke fixed-points

As shown above, Kripke fixed-points are not closed under classical logic. Consistent Kripke fixed-points are closed under a three-valued or partial logic called Strong Kleene logic: some sentences are true, some are false, some lack a truth value.

The logic of consistent Kripke fixed-points

As shown above, Kripke fixed-points are not closed under classical logic. Consistent Kripke fixed-points are closed under a three-valued or partial logic called Strong Kleene logic: some sentences are true, some are false, some lack a truth value.

Here are the truth tables for $\wedge$ and $\neg$:

φ	ψ	$(\varphi \wedge \psi)$
T	T	T
T	F	F
T	-	-
F	T	F
F	F	F
F	-	F
-	T	-
-	F	F
-	-	-

φ	$\neg\varphi$
T	F
F	T
-	-

Other consistent Kripke fixed-points

A consistent Kripke fixed-point S is *maximal* iff there is no Kripke fixed-point $S' \subsetneq S$. There is more than one maximal Kripke fixed-point.

Other consistent Kripke fixed-points

A consistent Kripke fixed-point S is *maximal* iff there is no Kripke fixed-point $S' \subsetneqq S$. There is more than one maximal Kripke fixed-point.

A consistent Kripke fixed-point S is *intrinsic* iff any $\varphi \in S$ is contained in any Kripke maximal fixed-point. There are various intrinsic Kripke-fixed points.

Other consistent Kripke fixed-points

A consistent Kripke fixed-point S is *maximal* iff there is no Kripke fixed-point $S' \subsetneqq S$. There is more than one maximal Kripke fixed-point.

A consistent Kripke fixed-point S is *intrinsic* iff any $\varphi \in S$ is contained in any Kripke maximal fixed-point. There are various intrinsic Kripke-fixed points.

The consistent Kripke fixed-point S is *maximal intrinsic* iff it all intrinsic Kripke fixed-points. There is only one such fixed point.

Consistent Kripke fixed-points

These different consistent Kripke fixed-points can be used to classify sentences:

Consistent Kripke fixed-points

These different consistent Kripke fixed-points can be used to classify sentences:

- The liar isn't in any consistent Kripke fixed-point. such sentences are called paradoxical.

Consistent Kripke fixed-points

These different consistent Kripke fixed-points can be used to classify sentences:

- The liar isn't in any consistent Kripke fixed-point. such sentences are called paradoxical.
- The truth teller is contained in some consistent Kripke fixed-points. It's not an element of an intrinsic consistent Kripke fixed-point.

Consistent Kripke fixed-points

These different consistent Kripke fixed-points can be used to classify sentences:

- The liar isn't in any consistent Kripke fixed-point. such sentences are called paradoxical.
- The truth teller is contained in some consistent Kripke fixed-points. It's not an element of an intrinsic consistent Kripke fixed-point.
- By the diagonal lemma, there is a sentence such that $\mathcal{A} \vdash \tau \leftrightarrow (T\overline{\tau} \vee \neg T\overline{\tau})$. Such a τ is an element of the maximal intrinsic consistent Kripke fixed-point, but it's not an element of the minimal fixed-point.

Other evaluation schemata

Kripke (1975) presented his theory in a different way which allows one to change the logic of the fixed-points to other logics, eg supervaluations and Weak Kleene logic.

Other evaluation schemata

Kripke (1975) presented his theory in a different way which allows one to change the logic of the fixed-points to other logics, eg supervaluations and Weak Kleene logic.

Supervaluations will make $\lambda \vee \neg\lambda$ (λ the liar true), but one looses compositionality. The disjunction will be true although both disjuncts lack truth values (and such disjunction 'usually' do not have a truth value).

Complexity considerations

- Closing a set under the operations 1-9 may take many steps. In the case of the standard model of arithmetic or a simple structure of expressions it takes ω_1^{CK} many steps.

Complexity considerations

- Closing a set under the operations 1-9 may take many steps. In the case of the standard model of arithmetic or a simple structure of expressions it takes ω_1^{CK} many steps.
- S is defined by an inductive definition, but any such definition (positive inductive definition) may be reduced to the definition of truth (Π^1_1 sets).

Complexity considerations

- Closing a set under the operations 1-9 may take many steps. In the case of the standard model of arithmetic or a simple structure of expressions it takes ω_1^{CK} many steps.
- S is defined by an inductive definition, but any such definition (positive inductive definition) may be reduced to the definition of truth (Π^1_1 sets).
- These complexity issues is fairly independent from the evaluation scheme. For the complexity considerations see Burgess (1986), Burgess (1988), McGee (1991).

Complexity considerations

- Closing a set under the operations 1-9 may take many steps. In the case of the standard model of arithmetic or a simple structure of expressions it takes ω_1^{CK} many steps.
- S is defined by an inductive definition, but any such definition (positive inductive definition) may be reduced to the definition of truth (Π_1^1 sets).
- These complexity issues is fairly independent from the evaluation scheme. For the complexity considerations see Burgess (1986), Burgess (1988), McGee (1991).
- There is a straightforward translation between the Tarskian hierarchy up to ω_1^{CK} and the minimal Kripke fixed point (Halbach (1997)).

Truth & Paradox

VI · The Kripke-Feferman Theory of Truth

Volker Halbach

Nordic Logic Summer School 2017

What has happened so far

Last time I have shown how to define a set of sentences having certain nice properties (Kripke fixed-point).

What has happened so far

Last time I have shown how to define a set of sentences having certain nice properties (Kripke fixed-point).

The definition of the set starts from a given model $\mathcal{M}$ of $\mathcal{A}$ interpreting all function and predicate symbols of $\mathcal{L}$ except for T.

What has happened so far

Last time I have shown how to define a set of sentences having certain nice properties (Kripke fixed-point).

The definition of the set starts from a given model $\mathcal{M}$ of $\mathcal{A}$ interpreting all function and predicate symbols of $\mathcal{L}$ except for T.

Here is the definition again:

Defining type-free truth

Definition

A set S of $\mathcal{L}$-sentences is a Kripke fixed-point iff for all sentences φ and ψ the following holds:

1. $s = t$ is true in $\mathcal{M}$ iff $(s = t) \in S$.

Defining type-free truth

Definition

A set S of $\mathcal{L}$-sentences is a Kripke fixed-point iff for all sentences φ and ψ the following holds:

1. $s = t$ is true in $\mathcal{M}$ iff $(s = t) \in S$. Similarly for all other atomic sentences without T: φ is true in $\mathcal{M}$ iff $\varphi \in S$.

Defining type-free truth

Definition

A set S of $\mathcal{L}$-sentences is a Kripke fixed-point iff for all sentences φ and ψ the following holds:

1. $s = t$ is true in $\mathcal{M}$ iff $(s = t) \in S$. Similarly for all other atomic sentences without T: φ is true in $\mathcal{M}$ iff $\varphi \in S$.
2. $s = t$ is false in $\mathcal{M}$ iff $(\neg s = t) \in S$.

Defining type-free truth

Definition

A set S of $\mathcal{L}$-sentences is a Kripke fixed-point iff for all sentences φ and ψ the following holds:

1. $s = t$ is true in $\mathcal{M}$ iff $(s = t) \in S$. Similarly for all other atomic sentences without T: φ is true in $\mathcal{M}$ iff $\varphi \in S$.
2. $s = t$ is false in $\mathcal{M}$ iff $(\neg s = t) \in S$. Similarly for all other atomic sentences without T: $\neg\varphi$ is true in $\mathcal{M}$ iff $\neg\varphi \in S$.

Defining type-free truth

Definition

A set S of $\mathcal{L}$-sentences is a Kripke fixed-point iff for all sentences φ and ψ the following holds:

1. $s = t$ is true in $\mathcal{M}$ iff $(s = t) \in S$. Similarly for all other atomic sentences without T: φ is true in $\mathcal{M}$ iff $\varphi \in S$.
2. $s = t$ is false in $\mathcal{M}$ iff $(\neg s = t) \in S$. Similarly for all other atomic sentences without T: $\neg\varphi$ is true in $\mathcal{M}$ iff $\neg\varphi \in S$.
3. ($\varphi \in S$ and $\psi \in S$) iff $(\varphi \wedge \psi) \in S$.

Defining type-free truth

Definition

A set S of $\mathcal{L}$-sentences is a Kripke fixed-point iff for all sentences φ and ψ the following holds:

1. $s = t$ is true in $\mathcal{M}$ iff $(s = t) \in S$. Similarly for all other atomic sentences without T: φ is true in $\mathcal{M}$ iff $\varphi \in S$.
2. $s = t$ is false in $\mathcal{M}$ iff $(\neg s = t) \in S$. Similarly for all other atomic sentences without T: $\neg\varphi$ is true in $\mathcal{M}$ iff $\neg\varphi \in S$.
3. ($\varphi \in S$ and $\psi \in S$) iff $(\varphi \wedge \psi) \in S$.
4. ($\neg\varphi \in S$ or $\neg\psi \in S$) iff $(\neg(\varphi \wedge \psi)) \in S$.

Defining type-free truth

Definition

A set S of $\mathcal{L}$-sentences is a Kripke fixed-point iff for all sentences φ and ψ the following holds:

1. $s = t$ is true in $\mathcal{M}$ iff $(s = t) \in S$. Similarly for all other atomic sentences without T: φ is true in $\mathcal{M}$ iff $\varphi \in S$.
2. $s = t$ is false in $\mathcal{M}$ iff $(\neg s = t) \in S$. Similarly for all other atomic sentences without T: $\neg\varphi$ is true in $\mathcal{M}$ iff $\neg\varphi \in S$.
3. ($\varphi \in S$ and $\psi \in S$) iff $(\varphi \wedge \psi) \in S$.
4. ($\neg\varphi \in S$ or $\neg\psi \in S$) iff $(\neg(\varphi \wedge \psi)) \in S$.
5. $\varphi \in S$ iff $(\neg\neg\varphi) \in S$.

Defining type-free truth

Definition

A set S of $\mathcal{L}$-sentences is a Kripke fixed-point iff for all sentences φ and ψ the following holds:

1. $s = t$ is true in $\mathcal{M}$ iff $(s = t) \in S$. Similarly for all other atomic sentences without T: φ is true in $\mathcal{M}$ iff $\varphi \in S$.
2. $s = t$ is false in $\mathcal{M}$ iff $(\neg s = t) \in S$. Similarly for all other atomic sentences without T: $\neg\varphi$ is true in $\mathcal{M}$ iff $\neg\varphi \in S$.
3. ($\varphi \in S$ and $\psi \in S$) iff $(\varphi \wedge \psi) \in S$.
4. ($\neg\varphi \in S$ or $\neg\psi \in S$) iff $(\neg(\varphi \wedge \psi)) \in S$.
5. $\varphi \in S$ iff $(\neg\neg\varphi) \in S$.
6. $\varphi(\overline{e}) \in S$ for all objects e , iff $\forall x \varphi(x) \in S$.

Defining type-free truth

Definition

A set S of $\mathcal{L}$-sentences is a Kripke fixed-point iff for all sentences φ and ψ the following holds:

1. $s = t$ is true in $\mathcal{M}$ iff $(s = t) \in S$. Similarly for all other atomic sentences without T: φ is true in $\mathcal{M}$ iff $\varphi \in S$.
2. $s = t$ is false in $\mathcal{M}$ iff $(\neg s = t) \in S$. Similarly for all other atomic sentences without T: $\neg\varphi$ is true in $\mathcal{M}$ iff $\neg\varphi \in S$.
3. ($\varphi \in S$ and $\psi \in S$) iff $(\varphi \wedge \psi) \in S$.
4. ($\neg\varphi \in S$ or $\neg\psi \in S$) iff $(\neg(\varphi \wedge \psi)) \in S$.
5. $\varphi \in S$ iff $(\neg\neg\varphi) \in S$.
6. $\varphi(\overline{e}) \in S$ for all objects e , iff $\forall x \varphi(x) \in S$.
7. $\neg\varphi(\overline{e}) \in S$ for some object e , iff $(\neg\forall x \varphi(x)) \in S$.

Defining type-free truth

Definition

A set S of $\mathcal{L}$-sentences is a Kripke fixed-point iff for all sentences φ and ψ the following holds:

1. $s = t$ is true in $\mathcal{M}$ iff $(s = t) \in S$. Similarly for all other atomic sentences without T: φ is true in $\mathcal{M}$ iff $\varphi \in S$.
2. $s = t$ is false in $\mathcal{M}$ iff $(\neg s = t) \in S$. Similarly for all other atomic sentences without T: $\neg\varphi$ is true in $\mathcal{M}$ iff $\neg\varphi \in S$.
3. ($\varphi \in S$ and $\psi \in S$) iff $(\varphi \wedge \psi) \in S$.
4. ($\neg\varphi \in S$ or $\neg\psi \in S$) iff $(\neg(\varphi \wedge \psi)) \in S$.
5. $\varphi \in S$ iff $(\neg\neg\varphi) \in S$.
6. $\varphi(\overline{e}) \in S$ for all objects e , iff $\forall x \varphi(x) \in S$.
7. $\neg\varphi(\overline{e}) \in S$ for some object e , iff $(\neg\forall x \varphi(x)) \in S$.
8. $\varphi \in S$ iff $(Tt) \in S$, if the value of the term t is φ.

Defining type-free truth

Definition

A set S of $\mathcal{L}$-sentences is a Kripke fixed-point iff for all sentences φ and ψ the following holds:

1. $s = t$ is true in $\mathcal{M}$ iff $(s = t) \in S$. Similarly for all other atomic sentences without T: φ is true in $\mathcal{M}$ iff $\varphi \in S$.
2. $s = t$ is false in $\mathcal{M}$ iff $(\neg s = t) \in S$. Similarly for all other atomic sentences without T: $\neg\varphi$ is true in $\mathcal{M}$ iff $\neg\varphi \in S$.
3. ($\varphi \in S$ and $\psi \in S$) iff $(\varphi \wedge \psi) \in S$.
4. ($\neg\varphi \in S$ or $\neg\psi \in S$) iff $(\neg(\varphi \wedge \psi)) \in S$.
5. $\varphi \in S$ iff $(\neg\neg\varphi) \in S$.
6. $\varphi(\overline{e}) \in S$ for all objects e , iff $\forall x\varphi(x) \in S$.
7. $\neg\varphi(\overline{e}) \in S$ for some object e , iff $(\neg\forall x\varphi(x)) \in S$.
8. $\varphi \in S$ iff $(Tt) \in S$, if the value of the term t is φ.
9. $\neg\varphi \in S$ iff $(\neg Tt) \in S$, if the value of the term t is φ.

interpreting $\mathcal{L}$

Assume $\mathcal{M}$ is a model of the language $\mathcal{L}$ without T, and assume that S is a subset of the domain of $\mathcal{M}$. Then $(\mathcal{M}, S)$ is the *classical* model of $\mathcal{L}$ interpreting all symbols in the same way as $\mathcal{M}$ and assigning to T the set S as extension.

interpreting $\mathcal{L}$

Assume $\mathcal{M}$ is a model of the language $\mathcal{L}$ without T, and assume that S is a subset of the domain of $\mathcal{M}$. Then $(\mathcal{M}, S)$ is the *classical* model of $\mathcal{L}$ interpreting all symbols in the same way as $\mathcal{M}$ and assigning to T the set S as extension.

Remark

$(\mathcal{M}, S) \vDash T\overline{\varphi}$ iff $\varphi \in S$.

interpreting $\mathcal{L}$

Assume $\mathcal{M}$ is a model of the language $\mathcal{L}$ without T, and assume that S is a subset of the domain of $\mathcal{M}$. Then $(\mathcal{M}, S)$ is the *classical* model of $\mathcal{L}$ interpreting all symbols in the same way as $\mathcal{M}$ and assigning to T the set S as extension.

Remark

$(\mathcal{M}, S) \vDash T\overline{\varphi}$ iff $\varphi \in S$.

Generally the problem is to find a 'neat' extension S for the truth predicate.

Closing off

Assume $\mathcal{M}$ is the intended model of $\mathcal{A}$ (without an interpretation of T) and assume further that S is some Kripke fixed-point. Then $(\mathcal{M}, S)$ ought to be a well behaved model for $\mathcal{L}$.

Closing off

Assume $\mathcal{M}$ is the intended model of $\mathcal{A}$ (without an interpretation of T) and assume further that S is some Kripke fixed-point. Then $(\mathcal{M}, S)$ ought to be a well behaved model for $\mathcal{L}$.

Assume $(\mathcal{M}, S)$ is such a model of $\mathcal{L}$, then the following will hold:

- $(\mathcal{M}, S) \vDash T\overline{\varphi} \leftrightarrow T\overline{T\overline{\varphi}}$
- $(\mathcal{M}, S) \vDash T\overline{\varphi \wedge \psi} \leftrightarrow (T\overline{\varphi} \wedge T\overline{\psi})$
- $(\mathcal{M}, S) \vDash \neg T\overline{\lambda}$ if S is a consistent Kripke fixed-point and if λ is the liar sentence. In this case $(\mathcal{M}, S) \vDash \lambda$.
- $(\mathcal{M}, S) \nvDash T\overline{\lambda \vee \neg\lambda}$ if S is a consistent Kripke fixed-point, but of course $(\mathcal{M}, S) \vDash \lambda \vee \neg\lambda$ as $(\mathcal{M}, S)$ is a classical model.

Again: why axioms?

When one tries to apply Kripke's theory to our overall theory, eg Zermelo–Fraenkel set theory or the like, one is confronted with the difficulty that we don't have a starting model, i.e. 'the standard model of ZF'.

Again: why axioms?

When one tries to apply Kripke's theory to our overall theory, eg Zermelo–Fraenkel set theory or the like, one is confronted with the difficulty that we don't have a starting model, i.e. 'the standard model of ZF'.

Of course one can claim that this is quite ok: ZF is our 'topmost' theory: everything sensible – including semantics – must be definable in it.

Again: why axioms?

When one tries to apply Kripke's theory to our overall theory, eg Zermelo–Fraenkel set theory or the like, one is confronted with the difficulty that we don't have a starting model, i.e. 'the standard model of ZF'.

Of course one can claim that this is quite ok: ZF is our 'topmost' theory: everything sensible – including semantics – must be definable in it.

My proposal: The topmost theory ought to contain semantics as well. This semantic theory is not reduced to set theory again, but given axiomatically. We may hope that axioms for truth that work for Peano arithmetic as base theory work also for ZFC as base theory.

The reflective closure

Feferman (1991) proposed to define the *reflective closure* of a theory (such as PA or ZF). This is obtained from the theory by adding axioms describing a Kripke fixed-point.

The reflective closure

Feferman (1991) proposed to define the *reflective closure* of a theory (such as PA or ZF). This is obtained from the theory by adding axioms describing a Kripke fixed-point.

Feferman takes *schematic theories* as his starting point, which are then applied to the language including the truth predicate. Thus the reflective closure of a theory contains always all instances of the induction scheme (induction, replacement, separation,…).

The reflective closure

Feferman thought of the reflective closure as a way of making explicit all assumptions or commitments implicit in the acceptance of a theory.

The reflective closure

Feferman thought of the reflective closure as a way of making explicit all assumptions or commitments implicit in the acceptance of a theory.

If we accept a theory S, then we are also committed to the consistency of S. We are also committed to the local reflection principle for S

$$\mathrm{Bew}_S(\overline{\varphi}) \to \varphi$$

for all sentences φ, and iterations of these reflection principles.

The reflective closure

The strongest reflection principle for a theory S is the Global reflection principle

$$\forall x\big(\mathrm{Sent}(x) \wedge \mathrm{Bew}_S(x) \to Tx\big)$$

The reflective closure

The strongest reflection principle for a theory S is the Global reflection principle

$$\forall x\big(\mathrm{Sent}(x) \wedge \mathrm{Bew}_S(x) \to Tx\big)$$

There are a couple of problems with the exact formulation of the Global reflection principle, but you get the idea…

The reflective closure

The strongest reflection principle for a theory S is the Global reflection principle

$$\forall x\big(\mathrm{Sent}(x) \wedge \mathrm{Bew}_S(x) \to Tx\big)$$

There are a couple of problems with the exact formulation of the Global reflection principle, but you get the idea…

Thus rather than adding proof-theoretic reflection principles (consistency statement, local or uniform reflection), one adds a truth theory proving Global reflection and iterations thereof.

The reflective closure

The strongest reflection principle for a theory S is the Global reflection principle

$$\forall x\big(\mathrm{Sent}(x) \wedge \mathrm{Bew}_S(x) \to Tx\big)$$

There are a couple of problems with the exact formulation of the Global reflection principle, but you get the idea…

Thus rather than adding proof-theoretic reflection principles (consistency statement, local or uniform reflection), one adds a truth theory proving Global reflection and iterations thereof.

To this end one can add iterations of Global reflection using typed truth predicates – or, much more elegantly, axioms describing the definition of Kripke fixed-points.

History of the system

Feferman (1991) proposed as system similar to the one above in 1979 in a talk, which resultedThe first published version of KF appeared in as Reinhardt (1986); Cantini's KF has weaker induction; McGee's version in (McGee 1991) is similar to Reinhardt's. Occasionally the Consistency axiom is omitted.

Axiomatising Kripke's theory

Definition

The theory KF is given by all axioms of $\mathcal{A}$ including all induction axioms and the following axioms:

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T(x \doteq y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y)) \big)$

Axiomatising Kripke's theory

Definition

The theory KF is given by all axioms of $\mathcal{A}$ including all induction axioms and the following axioms:

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T(x \doteq y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y)) \big)$
   $s = t$ is true in $\mathcal{M}$ iff $(s = t) \in S$

Axiomatising Kripke's theory

Definition

The theory KF is given by all axioms of $\mathcal{A}$ including all induction axioms and the following axioms:

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T(x \doteq y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y)) \big)$
   $s = t$ is true in $\mathcal{M}$ iff $(s = t) \in S$
2. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T \dot{\neg} (x \doteq y) \leftrightarrow \mathrm{val}(x) \neq \mathrm{val}(y)) \big)$

Axiomatising Kripke's theory

Definition

The theory KF is given by all axioms of $\mathcal{A}$ including all induction axioms and the following axioms:

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T(x \doteq y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y)) \big)$
   $s = t$ is true in $\mathcal{M}$ iff $(s = t) \in S$
2. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T \dot{\neg} (x \doteq y) \leftrightarrow \mathrm{val}(x) \neq \mathrm{val}(y)) \big)$
   $s = t$ is false in $\mathcal{M}$ iff $(\neg s = t) \in S$

Axiomatising Kripke's theory

Definition

The theory KF is given by all axioms of $\mathcal{A}$ including all induction axioms and the following axioms:

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T(x \doteq y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y)) \big)$
   $s = t$ is true in $\mathcal{M}$ iff $(s = t) \in S$
2. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T \dot{\neg} (x \doteq y) \leftrightarrow \mathrm{val}(x) \neq \mathrm{val}(y)) \big)$
   $s = t$ is false in $\mathcal{M}$ iff $(\neg s = t) \in S$
3. …and so on for all predicates other than = and T.

Axiomatising Kripke's theory

Definition

The theory KF is given by all axioms of $\mathcal{A}$ including all induction axioms and the following axioms:

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T(x \doteq y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y)) \big)$
   $s = t$ is true in $\mathcal{M}$ iff $(s = t) \in S$
2. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T \dot{\neg} (x \doteq y) \leftrightarrow \mathrm{val}(x) \neq \mathrm{val}(y)) \big)$
   $s = t$ is false in $\mathcal{M}$ iff $(\neg s = t) \in S$
3. …and so on for all predicates other than = and T.
4. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (T(x \dot{\wedge} y) \leftrightarrow (Tx \wedge Ty) \big)$

Axiomatising Kripke's theory

Definition

The theory KF is given by all axioms of $\mathcal{A}$ including all induction axioms and the following axioms:

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T(x \doteq y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y)) \big)$
   $s = t$ is true in $\mathcal{M}$ iff $(s = t) \in S$
2. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T \dot{\neg} (x \doteq y) \leftrightarrow \mathrm{val}(x) \neq \mathrm{val}(y)) \big)$
   $s = t$ is false in $\mathcal{M}$ iff $(\neg s = t) \in S$
3. …and so on for all predicates other than = and T.
4. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (T(x \dot{\wedge} y) \leftrightarrow (Tx \wedge Ty) \big)$
   $(\varphi \in S$ and $\psi \in S)$ iff $(\varphi \wedge \psi) \in S$.

Axiomatising Kripke's theory

Definition

The theory KF is given by all axioms of $\mathcal{A}$ including all induction axioms and the following axioms:

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T(x \doteq y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y)) \big)$
   $s = t$ is true in $\mathcal{M}$ iff $(s = t) \in S$
2. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T \dot{\neg} (x \doteq y) \leftrightarrow \mathrm{val}(x) \neq \mathrm{val}(y)) \big)$
   $s = t$ is false in $\mathcal{M}$ iff $(\neg s = t) \in S$
3. …and so on for all predicates other than = and T.
4. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (T(x \dot{\wedge} y) \leftrightarrow (Tx \wedge Ty) \big)$
   $(\varphi \in S$ and $\psi \in S)$ iff $(\varphi \wedge \psi) \in S$.
5. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (T \dot{\neg} (x \dot{\wedge} y) \leftrightarrow (T \dot{\neg} x \vee T \dot{\neg} y) \big)$

Axiomatising Kripke's theory

Definition

The theory KF is given by all axioms of $\mathcal{A}$ including all induction axioms and the following axioms:

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T(x \doteq y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y)) \big)$
   $s = t$ is true in $\mathcal{M}$ iff $(s = t) \in S$
2. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T \dot{\neg} (x \doteq y) \leftrightarrow \mathrm{val}(x) \neq \mathrm{val}(y)) \big)$
   $s = t$ is false in $\mathcal{M}$ iff $(\neg s = t) \in S$
3. …and so on for all predicates other than = and T.
4. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (T(x \dot{\wedge} y) \leftrightarrow (Tx \wedge Ty) \big)$
   $(\varphi \in S$ and $\psi \in S)$ iff $(\varphi \wedge \psi) \in S$.
5. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (T \dot{\neg} (x \dot{\wedge} y) \leftrightarrow (T \dot{\neg} x \vee T \dot{\neg} y) \big)$
   $(\neg\varphi \in S$ or $\neg\psi \in S)$ iff $(\neg(\varphi \wedge \psi)) \in S$

Axiomatising Kripke's theory

Definition

The theory KF is given by all axioms of $\mathcal{A}$ including all induction axioms and the following axioms:

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T(x \doteq y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y)) \big)$
   $s = t$ is true in $\mathcal{M}$ iff $(s = t) \in S$
2. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T \dot{\neg} (x \doteq y) \leftrightarrow \mathrm{val}(x) \neq \mathrm{val}(y)) \big)$
   $s = t$ is false in $\mathcal{M}$ iff $(\neg s = t) \in S$
3. …and so on for all predicates other than = and T.
4. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (T(x \dot{\wedge} y) \leftrightarrow (Tx \wedge Ty)) \big)$
   $(\varphi \in S$ and $\psi \in S)$ iff $(\varphi \wedge \psi) \in S$.
5. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (T \dot{\neg} (x \dot{\wedge} y) \leftrightarrow (T \dot{\neg} x \vee T \dot{\neg} y)) \big)$
   $(\neg\varphi \in S$ or $\neg\psi \in S)$ iff $(\neg(\varphi \wedge \psi)) \in S$
6. $\forall x \big(\mathrm{Sent}(x) \rightarrow (T \dot{\neg} \dot{\neg} x \leftrightarrow Tx) \big)$

Axiomatising Kripke's theory

Definition

The theory KF is given by all axioms of $\mathcal{A}$ including all induction axioms and the following axioms:

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T(x \doteq y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y)) \big)$
   $s = t$ is true in $\mathcal{M}$ iff $(s = t) \in S$
2. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T \dot{\neg} (x \doteq y) \leftrightarrow \mathrm{val}(x) \neq \mathrm{val}(y)) \big)$
   $s = t$ is false in $\mathcal{M}$ iff $(\neg s = t) \in S$
3. …and so on for all predicates other than = and T.
4. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (T(x \dot{\wedge} y) \leftrightarrow (Tx \wedge Ty)) \big)$
   $(\varphi \in S$ and $\psi \in S)$ iff $(\varphi \wedge \psi) \in S$.
5. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (T \dot{\neg} (x \dot{\wedge} y) \leftrightarrow (T \dot{\neg} x \vee T \dot{\neg} y)) \big)$
   $(\neg\varphi \in S$ or $\neg\psi \in S)$ iff $(\neg(\varphi \wedge \psi)) \in S$
6. $\forall x \big(\mathrm{Sent}(x) \rightarrow (T \dot{\neg} \dot{\neg} x \leftrightarrow Tx) \big)$
   $\varphi \in S$ iff $(\neg\neg\varphi) \in S$.

Axiomatising Kripke's theory

Definition

7. $\forall x \forall y \big(\mathrm{Sent}(\dot{\forall} x y) \to (T(\dot{\forall} x y) \leftrightarrow \forall z T\mathrm{sub}(\mathrm{q}z, x, y)) \big)$

Axiomatising Kripke's theory

Definition

7. $\forall x \forall y \big(\mathrm{Sent}(\dot{\forall} x y) \to (T(\dot{\forall} x y) \leftrightarrow \forall z T \mathrm{sub}(\mathrm{q}z, x, y)) \big)$
   $\varphi(\overline{e}) \in S$ for all objects e , iff $\forall x \varphi(x) \in S$

Axiomatising Kripke's theory

Definition

7. $\forall x \forall y \big(\mathrm{Sent}(\dot{\forall} x y) \to (T(\dot{\forall} x y) \leftrightarrow \forall z T \mathrm{sub}(\mathrm{q}z, x, y)) \big)$
   $\varphi(\overline{e}) \in S$ for all objects e , iff $\forall x \varphi(x) \in S$
8. $\forall x \forall y \big(\mathrm{Sent}(\dot{\forall} x y) \to (T(\dot{\neg} \dot{\forall} x y) \leftrightarrow \exists z T \dot{\neg} \mathrm{sub}(\mathrm{q}z, x, y)) \big)$

Axiomatising Kripke's theory

Definition

7. $\forall x \forall y \big(\mathrm{Sent}(\dot{\forall} x y) \to (T(\dot{\forall} x y) \leftrightarrow \forall z T \mathrm{sub}(\mathrm{q}z, x, y)) \big)$
   $\varphi(\overline{e}) \in S$ for all objects e , iff $\forall x \varphi(x) \in S$
8. $\forall x \forall y \big(\mathrm{Sent}(\dot{\forall} x y) \to (T(\dot{\neg} \dot{\forall} x y) \leftrightarrow \exists z T \dot{\neg} \mathrm{sub}(\mathrm{q}z, x, y)) \big)$
   $\neg \varphi(\overline{e}) \in S$ for some object e , iff $(\neg \forall x \varphi(x)) \in S$

Axiomatising Kripke's theory

Definition

7. $\forall x \forall y \big(\mathrm{Sent}(\dot{\forall} x y) \to (T(\dot{\forall} x y) \leftrightarrow \forall z T \mathrm{sub}(\mathrm{q}z, x, y)) \big)$
   $\varphi(\overline{e}) \in S$ for all objects e , iff $\forall x \varphi(x) \in S$
8. $\forall x \forall y \big(\mathrm{Sent}(\dot{\forall} x y) \to (T(\dot{\neg} \dot{\forall} x y) \leftrightarrow \exists z T \dot{\neg} \mathrm{sub}(\mathrm{q}z, x, y)) \big)$
   $\neg \varphi(\overline{e}) \in S$ for some object e , iff $(\neg \forall x \varphi(x)) \in S$
9. $\forall x \big(\mathrm{ClT}(x) \to (T \dot{T} x \leftrightarrow T \mathrm{val}(x)) \big)$

Axiomatising Kripke's theory

Definition

7. $\forall x \forall y \big(\mathrm{Sent}(\dot{\forall} x y) \rightarrow (T(\dot{\forall} x y) \leftrightarrow \forall z T \mathrm{sub}(\mathrm{q}z, x, y)) \big)$
   $\varphi(\overline{e}) \in S$ for all objects e , iff $\forall x \varphi(x) \in S$
8. $\forall x \forall y \big(\mathrm{Sent}(\dot{\forall} x y) \rightarrow (T(\dot{\neg} \dot{\forall} x y) \leftrightarrow \exists z T \dot{\neg} \mathrm{sub}(\mathrm{q}z, x, y)) \big)$
   $\neg \varphi(\overline{e}) \in S$ for some object e , iff $(\neg \forall x \varphi(x)) \in S$
9. $\forall x \big(\mathrm{ClT}(x) \rightarrow (T \dot{T} x \leftrightarrow T \mathrm{val}(x)) \big)$
   $\varphi \in S$ iff $(T t) \in S$, if the value of the term t is φ

Axiomatising Kripke's theory

Definition

7. $\forall x \forall y \big(\mathrm{Sent}(\dot{\forall} x y) \to (T(\dot{\forall} x y) \leftrightarrow \forall z T \mathrm{sub}(\mathrm{q}z, x, y)) \big)$
   $\varphi(\overline{e}) \in S$ for all objects e , iff $\forall x \varphi(x) \in S$
8. $\forall x \forall y \big(\mathrm{Sent}(\dot{\forall} x y) \to (T(\dot{\neg} \dot{\forall} x y) \leftrightarrow \exists z T \dot{\neg} \mathrm{sub}(\mathrm{q}z, x, y)) \big)$
   $\neg \varphi(\overline{e}) \in S$ for some object e , iff $(\neg \forall x \varphi(x)) \in S$
9. $\forall x \big(\mathrm{ClT}(x) \to (T \dot{T} x \leftrightarrow T \mathrm{val}(x)) \big)$
   $\varphi \in S$ iff $(Tt) \in S$, if the value of the term t is φ
10. $\forall x \big(\mathrm{ClT}(x) \to (T \dot{\neg} \dot{T} x \leftrightarrow T \dot{\neg} \mathrm{val}(x)) \big)$

Axiomatising Kripke's theory

Definition

7. $\forall x \forall y \big(\mathrm{Sent}(\dot{\forall} x y) \to (T(\dot{\forall} x y) \leftrightarrow \forall z T \mathrm{sub}(\mathrm{q}z, x, y)) \big)$
   $\varphi(\overline{e}) \in S$ for all objects e , iff $\forall x \varphi(x) \in S$
8. $\forall x \forall y \big(\mathrm{Sent}(\dot{\forall} x y) \to (T(\dot{\neg} \dot{\forall} x y) \leftrightarrow \exists z T \dot{\neg} \mathrm{sub}(\mathrm{q}z, x, y)) \big)$
   $\neg \varphi(\overline{e}) \in S$ for some object e , iff $(\neg \forall x \varphi(x)) \in S$
9. $\forall x \big(\mathrm{ClT}(x) \to (T \dot{T} x \leftrightarrow T \mathrm{val}(x)) \big)$
   $\varphi \in S$ iff $(T t) \in S$, if the value of the term t is φ
10. $\forall x \big(\mathrm{ClT}(x) \to (T \dot{\neg} \dot{T} x \leftrightarrow T \dot{\neg} \mathrm{val}(x)) \big)$
    $\neg \varphi \in S$ iff $(\neg T t) \in S$, if the value of the term t is φ, .

Axiomatising Kripke's theory

Definition

7. $\forall x \forall y \big(\mathrm{Sent}(\dot{\forall} x y) \rightarrow (T(\dot{\forall} x y) \leftrightarrow \forall z T \mathrm{sub}(\mathrm{q}z, x, y)) \big)$
   $\varphi(\overline{e}) \in S$ for all objects e , iff $\forall x \varphi(x) \in S$
8. $\forall x \forall y \big(\mathrm{Sent}(\dot{\forall} x y) \rightarrow (T(\dot{\neg} \dot{\forall} x y) \leftrightarrow \exists z T \dot{\neg} \mathrm{sub}(\mathrm{q}z, x, y)) \big)$
   $\neg \varphi(\overline{e}) \in S$ for some object e , iff $(\neg \forall x \varphi(x)) \in S$
9. $\forall x \big(\mathrm{ClT}(x) \rightarrow (T \dot{T} x \leftrightarrow T \mathrm{val}(x)) \big)$
   $\varphi \in S$ iff $(T t) \in S$, if the value of the term t is φ
10. $\forall x \big(\mathrm{ClT}(x) \rightarrow (T \dot{\neg} \dot{T} x \leftrightarrow T \dot{\neg} \mathrm{val}(x)) \big)$
    $\neg \varphi \in S$ iff $(\neg T t) \in S$, if the value of the term t is φ, .
11. (Consistency axiom)
    $\forall x \big(\mathrm{Sent}(x) \rightarrow (T \dot{\neg} x \rightarrow \neg T x) \big)$

The KF axioms on one page …

1. $\forall x \forall y \Big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow \big(T(x \doteq y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y) \big) \Big)$

The KF axioms on one page . . .

1. $\forall x \forall y \Big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow \big(T(x \doteq y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y) \big) \Big)$
2. $\forall x \forall y \Big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow \big(T \neg (x \doteq y) \leftrightarrow \mathrm{val}(x) \neq \mathrm{val}(y) \big) \Big)$

The KF axioms on one page . . .

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow \big(T(x \doteq y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y) \big) \big)$
2. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow \big(T \dot{\neg} (x \doteq y) \leftrightarrow \mathrm{val}(x) \neq \mathrm{val}(y) \big) \big)$
3. . . . and so on for all predicates other than = and T.
4. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow \big(T(x \dot{\wedge} y) \leftrightarrow (Tx \wedge Ty) \big) \big)$

The KF axioms on one page . . .

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow \big(T(x \doteq y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y) \big) \big)$
2. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow \big(T \dot{\neg} (x \doteq y) \leftrightarrow \mathrm{val}(x) \neq \mathrm{val}(y) \big) \big)$
3. . . . and so on for all predicates other than = and T.
4. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow \big(T(x \dot{\wedge} y) \leftrightarrow (Tx \wedge Ty) \big) \big)$
5. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow \big(T \dot{\neg} (x \dot{\wedge} y) \leftrightarrow (T \dot{\neg} x \vee T \dot{\neg} y) \big) \big)$

The KF axioms on one page . . .

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow \big(T(x \doteq y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y) \big) \big)$
2. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow \big(T \dot{\neg} (x \doteq y) \leftrightarrow \mathrm{val}(x) \neq \mathrm{val}(y) \big) \big)$
3. . . . and so on for all predicates other than = and T.
4. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow \big(T(x \dot{\wedge} y) \leftrightarrow (Tx \wedge Ty) \big) \big)$
5. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow \big(T \dot{\neg} (x \dot{\wedge} y) \leftrightarrow (T \dot{\neg} x \vee T \dot{\neg} y) \big) \big)$
6. $\forall x \big(\mathrm{Sent}(x) \rightarrow \big(T \dot{\neg} \dot{\neg} x \leftrightarrow Tx \big) \big)$

The KF axioms on one page . . .

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T(x \doteq y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y)) \big)$
2. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T \dot{\neg} (x \doteq y) \leftrightarrow \mathrm{val}(x) \neq \mathrm{val}(y)) \big)$
3. . . . and so on for all predicates other than = and T.
4. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (T(x \dot{\wedge} y) \leftrightarrow (Tx \wedge Ty)) \big)$
5. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (T \dot{\neg} (x \dot{\wedge} y) \leftrightarrow (T \dot{\neg} x \vee T \dot{\neg} y)) \big)$
6. $\forall x \big(\mathrm{Sent}(x) \rightarrow (T \dot{\neg} \dot{\neg} x \leftrightarrow Tx) \big)$
7. $\forall x \forall y \big(\mathrm{Sent}(\dot{\forall} x y) \rightarrow (T(\dot{\forall} x y) \leftrightarrow \forall z T \mathrm{subsub}(\mathrm{q}z, x, y)) \big)$

The KF axioms on one page . . .

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T(x \doteq y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y)) \big)$
2. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T \dot{\neg} (x \doteq y) \leftrightarrow \mathrm{val}(x) \neq \mathrm{val}(y)) \big)$
3. . . . and so on for all predicates other than = and T.
4. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (T(x \dot{\wedge} y) \leftrightarrow (Tx \wedge Ty)) \big)$
5. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (T \dot{\neg} (x \dot{\wedge} y) \leftrightarrow (T \dot{\neg} x \vee T \dot{\neg} y)) \big)$
6. $\forall x \big(\mathrm{Sent}(x) \rightarrow (T \dot{\neg} \dot{\neg} x \leftrightarrow Tx) \big)$
7. $\forall x \forall y \big(\mathrm{Sent}(\dot{\forall} x y) \rightarrow (T(\dot{\forall} x y) \leftrightarrow \forall z T \mathrm{subsub}(\mathrm{q}z, x, y)) \big)$
8. $\forall x \forall y \big(\mathrm{Sent}(\dot{\forall} x y) \rightarrow (T(\dot{\neg} \dot{\forall} x y) \leftrightarrow \exists z T \dot{\neg} \mathrm{sub}(\mathrm{q}z, x, y)) \big)$

The KF axioms on one page . . .

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T(x \dot{=} y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y)) \big)$
2. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T \dot{\neg} (x \dot{=} y) \leftrightarrow \mathrm{val}(x) \neq \mathrm{val}(y)) \big)$
3. . . . and so on for all predicates other than = and T.
4. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (T(x \dot{\wedge} y) \leftrightarrow (Tx \wedge Ty)) \big)$
5. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (T \dot{\neg} (x \dot{\wedge} y) \leftrightarrow (T \dot{\neg} x \vee T \dot{\neg} y)) \big)$
6. $\forall x \big(\mathrm{Sent}(x) \rightarrow (T \dot{\neg} \dot{\neg} x \leftrightarrow Tx) \big)$
7. $\forall x \forall y \big(\mathrm{Sent}(\dot{\forall} x y) \rightarrow (T(\dot{\forall} x y) \leftrightarrow \forall z T \mathrm{subsub}(\mathrm{q}z, x, y)) \big)$
8. $\forall x \forall y \big(\mathrm{Sent}(\dot{\forall} x y) \rightarrow (T(\dot{\neg} \dot{\forall} x y) \leftrightarrow \exists z T \dot{\neg} \mathrm{sub}(\mathrm{q}z, x, y)) \big)$
9. $\forall x \big(\mathrm{ClT}(x) \rightarrow (T \dot{T} x \leftrightarrow T \mathrm{val}(x)) \big)$

The KF axioms on one page . . .

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T(x \doteq y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y)) \big)$
2. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T \dot{\neg} (x \doteq y) \leftrightarrow \mathrm{val}(x) \neq \mathrm{val}(y)) \big)$
3. . . . and so on for all predicates other than = and T.
4. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (T(x \dot{\wedge} y) \leftrightarrow (Tx \wedge Ty)) \big)$
5. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (T \dot{\neg} (x \dot{\wedge} y) \leftrightarrow (T \dot{\neg} x \vee T \dot{\neg} y)) \big)$
6. $\forall x \big(\mathrm{Sent}(x) \rightarrow (T \dot{\neg} \dot{\neg} x \leftrightarrow Tx) \big)$
7. $\forall x \forall y \big(\mathrm{Sent}(\dot{\forall} x y) \rightarrow (T(\dot{\forall} x y) \leftrightarrow \forall z T \mathrm{subsub}(\mathrm{q}z, x, y)) \big)$
8. $\forall x \forall y \big(\mathrm{Sent}(\dot{\forall} x y) \rightarrow (T(\dot{\neg} \dot{\forall} x y) \leftrightarrow \exists z T \dot{\neg} \mathrm{sub}(\mathrm{q}z, x, y)) \big)$
9. $\forall x \big(\mathrm{ClT}(x) \rightarrow (T \dot{T} x \leftrightarrow T \mathrm{val}(x)) \big)$
10. $\forall x \big(\mathrm{ClT}(x) \rightarrow (T \dot{\neg} \dot{T} x \leftrightarrow T \dot{\neg} \mathrm{val}(x)) \big)$

The KF axioms on one page . . .

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T(x \dot{=} y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y)) \big)$
2. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T \dot{\neg} (x \dot{=} y) \leftrightarrow \mathrm{val}(x) \neq \mathrm{val}(y)) \big)$
3. . . . and so on for all predicates other than = and T.
4. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (T(x \dot{\wedge} y) \leftrightarrow (Tx \wedge Ty)) \big)$
5. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (T \dot{\neg} (x \dot{\wedge} y) \leftrightarrow (T \dot{\neg} x \vee T \dot{\neg} y)) \big)$
6. $\forall x \big(\mathrm{Sent}(x) \rightarrow (T \dot{\neg} \dot{\neg} x \leftrightarrow Tx) \big)$
7. $\forall x \forall y \big(\mathrm{Sent}(\dot{\forall} x y) \rightarrow (T(\dot{\forall} x y) \leftrightarrow \forall z T \mathrm{subsub}(\mathrm{q}z, x, y)) \big)$
8. $\forall x \forall y \big(\mathrm{Sent}(\dot{\forall} x y) \rightarrow (T(\dot{\neg} \dot{\forall} x y) \leftrightarrow \exists z T \dot{\neg} \mathrm{sub}(\mathrm{q}z, x, y)) \big)$
9. $\forall x \big(\mathrm{ClT}(x) \rightarrow (T \dot{T} x \leftrightarrow T \mathrm{val}(x)) \big)$
10. $\forall x \big(\mathrm{ClT}(x) \rightarrow (T \dot{\neg} \dot{T} x \leftrightarrow T \dot{\neg} \mathrm{val}(x)) \big)$
11. (Consistency axiom)
    $\forall x \big(\mathrm{Sent}(x) \rightarrow (T \dot{\neg} x \rightarrow \neg Tx) \big)$

The mathematical content of KF

Theorem

KF with PA as base theory $\mathcal{A}$ proves the same arithmetical theorems as $\epsilon_0 = \omega^{\omega^{:\omega}}$-iterated arithmetical comprehension $RA_{<\epsilon_0}$ or ϵ_0-times iterated 'Tarskian' truth.

The mathematical content of KF

Theorem

KF with PA as base theory $\mathcal{A}$ proves the same arithmetical theorems as $\epsilon_0 = \omega^{\omega^{:\omega}}$-iterated arithmetical comprehension $RA_{<\epsilon_0}$ or ϵ_0-times iterated 'Tarskian' truth.

Feferman has tried to characterise the reflective closure of PA is various ways. KF fits into the picture.

The mathematical content of KF

Theorem

KF with PA as base theory $\mathcal{A}$ proves the same arithmetical theorems as $\epsilon_0 = \omega^{\omega^{:\omega}}$-iterated arithmetical comprehension $RA_{<\epsilon_0}$ or ϵ_0-times iterated 'Tarskian' truth.

Feferman has tried to characterise the reflective closure of PA is various ways. KF fits into the picture.

Anyway, KF allows one to carry out reasoning that would otherwise require quite a lot of typed truth predicates.

The mathematical content of KF

Theorem

KF with PA as base theory $\mathcal{A}$ proves the same arithmetical theorems as $\epsilon_0 = \omega^{\omega^{:\omega}}$-iterated arithmetical comprehension $RA_{<\epsilon_0}$ or ϵ_0-times iterated 'Tarskian' truth.

Feferman has tried to characterise the reflective closure of PA is various ways. KF fits into the picture.

Anyway, KF allows one to carry out reasoning that would otherwise require quite a lot of typed truth predicates.

But is it a nice theory?

Consistency

Over KF without the Consistency axiom, the following are equivalent:

(i) CONS, that is, $\forall x\big(\mathrm{Sent}(x) \to \neg(\mathrm{True}\, x \wedge \mathrm{True}\, \dot{\neg} x)\big)$

(ii) $\forall x\big(\mathrm{Sent}(x) \to (\mathrm{True}\, \dot{\neg} x \to \neg \mathrm{True}\, x)\big)$

(iii) the schema $\forall \vec{x}(\mathrm{True}\, \overline{\varphi(\dot{\vec{x}})} \to \varphi(\vec{x}))$ for all formulas $\varphi(\vec{x})$; $\vec{x}$ stands here for a string $x_1, \ldots, x_n$ of variables

(iv) the schema $\forall x\big(\mathrm{True}\, \overline{\varphi(\dot{x})} \to \varphi(x)\big)$ for all formulas $\varphi(x)$; this schema allows only one free variable in the respective instantiating formula.

(v) $\forall x(\mathrm{True}\, \overline{\neg \mathrm{True}\, \dot{x}} \to \neg \mathrm{True}\, x)$

The liar in KF

In particular $KF \vdash T\overline{\varphi} \to \varphi$ (previous page (iii))

The liar in KF

In particular $KF \vdash T\overline{\varphi} \rightarrow \varphi$ (previous page (iii))

Since KF is formulated in classical logic we have a sentence λ s.t.:

$KF \vdash \lambda \leftrightarrow \neg T\overline{\lambda}$

The liar in KF

In particular $KF \vdash T\overline{\varphi} \to \varphi$ (previous page (iii))
Since KF is formulated in classical logic we have a sentence λ s.t.:
$KF \vdash \lambda \leftrightarrow \neg T\overline{\lambda}$

Lemma

KF proves the liar sentence, ie $KF \vdash \lambda$.

Given the intended semantics, this shouldn't be too surprising; see 132.

Proof.

$KF \vdash T\overline{\lambda} \to \lambda$	previous page (iii)
$KF \vdash T\overline{\lambda} \to \neg\lambda$	diagonal lemma
$KF \vdash \neg T\overline{\lambda}$	
$KF \vdash \lambda$	diagonal lemma

$\dashv$

Worries about the consistency axiom

If the consistency axiom

$$\forall x\Big(\mathrm{Sent}(x) \to \big(T\dot{\neg}x \to \neg\, Tx\big)\Big)$$

is dropped, then the liar is no longer provable.

Worries about the consistency axiom

If the consistency axiom

$$\forall x\big(\mathrm{Sent}(x) \to (T\dot{\neg}x \to \neg Tx)\big)$$

is dropped, then the liar is no longer provable.

KF without the consistency axiom can be seen as an axiomatisation of arbitrary Kripke fixed-points, ie, fixed-points with truth-value gluts as well as gaps, while the original KF axiomatises consistent Kripke fixed-points (note that neither forces the minimal fixed point).

Internal and external logic

Still, even after dropping the consistency axiom KF describes a non-classical notion of truth in classical logic.

Internal and external logic

Still, even after dropping the consistency axiom KF describes a non-classical notion of truth in classical logic.

The internal logic $\{\varphi : \ KF \vdash T\overline{\varphi}\}$ of KF and its external logic, that is, the set of all KF-provable sentences are highly disparate.

Internal and external logic

Still, even after dropping the consistency axiom KF describes a non-classical notion of truth in classical logic.

The internal logic $\{\varphi : KF \vdash T\overline{\varphi}\}$ of KF and its external logic, that is, the set of all KF-provable sentences are highly disparate.

Example

- $KF \vdash \lambda \vee \neg\lambda$, but $KF \nvdash T\overline{\lambda \vee \neg\lambda}$ (also for KF without consistency axiom)

Internal and external logic

Still, even after dropping the consistency axiom KF describes a non-classical notion of truth in classical logic.

The internal logic $\{\varphi : KF \vdash T\overline{\varphi}\}$ of KF and its external logic, that is, the set of all KF-provable sentences are highly disparate.

Example

- $KF \vdash \lambda \vee \neg\lambda$, but $KF \nvdash T\overline{\lambda \vee \neg\lambda}$ (also for KF without consistency axiom)
- $KF \vdash \tau \vee \neg\tau$, but $KF \nvdash T\overline{\tau \vee \neg\tau}$ (also for KF without consistency axiom; τ the truth teller)

Internal and external logic

Still, even after dropping the consistency axiom KF describes a non-classical notion of truth in classical logic.

The internal logic $\{\varphi : KF \vdash T\overline{\varphi}\}$ of KF and its external logic, that is, the set of all KF-provable sentences are highly disparate.

Example

- $KF \vdash \lambda \vee \neg\lambda$, but $KF \nvdash T\overline{\lambda \vee \neg\lambda}$ (also for KF without consistency axiom)
- $KF \vdash \tau \vee \neg\tau$, but $KF \nvdash T\overline{\tau \vee \neg\tau}$ (also for KF without consistency axiom; τ the truth teller)
- $KF \vdash \lambda$ but $KF \vdash \neg T\overline{\lambda}$ (this requires the consistency axiom)

Internal and external logic

Thus KF with and without consistency axiom is a theory in classical logic of non-classical truth.

Internal and external logic

Thus KF with and without consistency axiom is a theory in classical logic of non-classical truth.

Isn't that against the spirit of Kripke's theory? Isn't what we are really interested an axiomatisation of S rather than of $(\mathcal{M}, S)$? Shouldn't we just stick to non-classical logic when adopting and axiomatising a non-classical notion of truth?

Internal and external logic

Thus KF with and without consistency axiom is a theory in classical logic of non-classical truth.

Isn't that against the spirit of Kripke's theory? Isn't what we are really interested an axiomatisation of S rather than of $(\mathcal{M}, S)$? Shouldn't we just stick to non-classical logic when adopting and axiomatising a non-classical notion of truth?

Feferman (to some extent) and Reinhardt seem to agree: what we are really interested in is a system allowing us to derive sentences that are elements of all Kripke fixed points S. Reinhardt thought one shouldn't 'trust' a classical metatheory, but that one should try to show that it is safe because it's dispensable.

Internal and external logic

Thus KF with and without consistency axiom is a theory in classical logic of non-classical truth.

Isn't that against the spirit of Kripke's theory? Isn't what we are really interested an axiomatisation of S rather than of $(\mathcal{M}, S)$? Shouldn't we just stick to non-classical logic when adopting and axiomatising a non-classical notion of truth?

Feferman (to some extent) and Reinhardt seem to agree: what we are really interested in is a system allowing us to derive sentences that are elements of all Kripke fixed points S. Reinhardt thought one shouldn't 'trust' a classical metatheory, but that one should try to show that it is safe because it's dispensable.

But at the same time they agree that one shouldn't try to think in Strong Kleene logic, because nothing like sustained ordinary reasoning can be carried out in such non-classical logics.

Axiomatising Kripke fixed-points

And what happened when your parents told you that you shouldn't try to do something?

Axiomatising Kripke fixed-points

And what happened when your parents told you that you shouldn't try to do something?

So Leon and I Halbach and Horsten (2006) tried to write down a theory directly axiomatising the conditions on Kripke fixed-points in three valued logic.

Axiomatising Kripke fixed-points

And what happened when your parents told you that you shouldn't try to do something?

So Leon and I Halbach and Horsten (2006) tried to write down a theory directly axiomatising the conditions on Kripke fixed-points in three valued logic.

Thus our intended model is not $(\mathcal{M}, S)$ but we want to generate all sentences in all Kripke fixed-points (over the standard model), ie, all sentences in the minimal Kripke fixed point.

Axiomatising Kripke fixed-points

And what happened when your parents told you that you shouldn't try to do something?

So Leon and I Halbach and Horsten (2006) tried to write down a theory directly axiomatising the conditions on Kripke fixed-points in three valued logic.

Thus our intended model is not $(\mathcal{M}, S)$ but we want to generate all sentences in all Kripke fixed-points (over the standard model), ie, all sentences in the minimal Kripke fixed point.

We tried … and perished.

Axiomatising Kripke fixed-points

First Leon and I tried to prove that the system PKF in Strong-Kleene logic is as strong as the classical system KF, ie that both systems prove the same T-free sentences. After all Feferman's Feferman (1991) analysis of KF makes use of the 'grounded' sentences only.

Axiomatising Kripke fixed-points

First Leon and I tried to prove that the system PKF in Strong-Kleene logic is as strong as the classical system KF, ie that both systems prove the same T-free sentences. After all Feferman's Feferman (1991) analysis of KF makes use of the 'grounded' sentences only.

When we tried to carry out Feferman's proof in PKF we didn't experience problems with the derivation of the truth-theoretic sentences (eg it wasn't a problem that KF proves the liar, while it's gappy in PKF). The problem was the proof of a mathematical schema: transfinite induction up to ϵ_0.

Axiomatising Kripke fixed-points

First Leon and I tried to prove that the system PKF in Strong-Kleene logic is as strong as the classical system KF, ie that both systems prove the same T-free sentences. After all Feferman's Feferman (1991) analysis of KF makes use of the 'grounded' sentences only.

When we tried to carry out Feferman's proof in PKF we didn't experience problems with the derivation of the truth-theoretic sentences (eg it wasn't a problem that KF proves the liar, while it's gappy in PKF). The problem was the proof of a mathematical schema: transfinite induction up to ϵ_0.

Theorem

PKF proves the same arithmetical sentences as RA_{ω^ω} *or* ω^ω*-times iterated 'Tarskian' truth* $\mathcal{D}^+$.

Using alternative evaluation schemata

I have focused on Kripke's theory with the Strong-Kleene evaluation scheme for handling truth value gaps.

Using alternative evaluation schemata

I have focused on Kripke's theory with the Strong-Kleene evaluation scheme for handling truth value gaps.

Kripke showed how to carry out his construction for other policies on truth-value gaps. There have been proposals to give KF-like axiomatisations of these theories.

Conclusions

- the use of classical logic isn't a matter of mere convenience: doing everything in Strong-Kleene logic cripples our mathematical reasoning.

Conclusions

- the use of classical logic isn't a matter of mere convenience: doing everything in Strong-Kleene logic cripples our mathematical reasoning.
- Reinhardt's hopes to justify the use of classical logic and of KF by an 'instrumentalist' interpretation are doomed.

Conclusions

- the use of classical logic isn't a matter of mere convenience: doing everything in Strong-Kleene logic cripples our mathematical reasoning.
- Reinhardt's hopes to justify the use of classical logic and of KF by an 'instrumentalist' interpretation are doomed.
- In order to evaluate an axiomatic theory of truth in non-classical logic it's not sufficient to look at the liar sentence and other truth-theoretic problems. One should check to what extent one has to give up classical patterns of reasoning.

Supervaluations

Cantini proposed a theory VF axiomatising Kripke's theory with supervaluations.

Supervaluations

Cantini proposed a theory VF axiomatising Kripke's theory with supervaluations.

Theorem

VF is proof-theoretically equivalent to ID_1 *and thus much stronger than KF.*

Weak Kleene

Feferman (1991) mentioned also that KF with Weak-Kleene logic should have the same proof-theoretic strength as KF.

Weak Kleene

Feferman (1991) mentioned also that KF with Weak-Kleene logic should have the same proof-theoretic strength as KF.

Example

$KF \vdash T\overline{\lambda \vee 0 = 0}$ but
$WKF \nvdash T\overline{\lambda \vee 0 = 0}$

Theorem (Fujimoto)

KF and WKF are proof-theoretically equivalent.

Classical logic again

I doubt more and more that KF-like systems are the way to go.

Classical logic again

I doubt more and more that KF-like systems are the way to go.

These systems describe a non-classical notion of truth, and if we go for such a notion we should revise our informal metatheory accordingly and start to think in non-classical logic. I tried it once and it proved to be a traumatic experience.

Classical logic again

I doubt more and more that KF-like systems are the way to go.

These systems describe a non-classical notion of truth, and if we go for such a notion we should revise our informal metatheory accordingly and start to think in non-classical logic. I tried it once and it proved to be a traumatic experience.

Thus I favour very much systems that are formulated in classical logic and describe a classical notion of truth…

Truth & Paradox

VII · Classical Symmetric Truth

Volker Halbach

Nordic Logic Summer School 2017

What has happened so far

I discussed the Kripke-Feferman theory and its variants. KF has the following drawbacks:

- Internal and external logic are incompatible. That is, we don't have the following for all sentences φ and ψ:

$$KF \vdash T\overline{\varphi} \text{ iff } KF \vdash \varphi$$

What has happened so far

I discussed the Kripke-Feferman theory and its variants. KF has the following drawbacks:

- Internal and external logic are incompatible. That is, we don't have the following for all sentences φ and ψ:

  $$KF \vdash T\overline{\varphi} \text{ iff } KF \vdash \varphi$$

  Call a system *symmetric* iff internal and external logic are identical.

What has happened so far

I discussed the Kripke-Feferman theory and its variants. KF has the following drawbacks:

- Internal and external logic are incompatible. That is, we don't have the following for all sentences φ and ψ:

  $$KF \vdash T\overline{\varphi} \text{ iff } KF \vdash \varphi$$

  Call a system *symmetric* iff internal and external logic are identical.
- The internal logic of KF is non-classical. If we force KF to be symmetric, we have to axiomatise KF in partial logic and abandon classical logic completely.

There may be good reasons to think about a partial notion of truth in classical logic. In fact, that's the typical situation in semantic theories of truth: people describe all kinds of funny non-classical systems by using classical systems as metatheories.

Symmetry

If we want object- and metatheory to be identical or at least compatible, then Kripke's theory forces us to abandon classical logic. In the end it's a non-classical theory.

Symmetry

If we want object- and metatheory to be identical or at least compatible, then Kripke's theory forces us to abandon classical logic. In the end it's a non-classical theory.

But can we have a system based on classical logic that is symmetric? That is, is there a system that is fully classical? Is there is system with a classical internal and external logic?

Symmetry

If we want object- and metatheory to be identical or at least compatible, then Kripke's theory forces us to abandon classical logic. In the end it's a non-classical theory.

But can we have a system based on classical logic that is symmetric? That is, is there a system that is fully classical? Is there is system with a classical internal and external logic?

Such a theory would have to satisfy two criteria:

Symmetry

If we want object- and metatheory to be identical or at least compatible, then Kripke's theory forces us to abandon classical logic. In the end it's a non-classical theory.

But can we have a system based on classical logic that is symmetric? That is, is there a system that is fully classical? Is there is system with a classical internal and external logic?

Such a theory would have to satisfy two criteria:

1. The system is formulated in classical logic.

Symmetry

If we want object- and metatheory to be identical or at least compatible, then Kripke's theory forces us to abandon classical logic. In the end it's a non-classical theory.

But can we have a system based on classical logic that is symmetric? That is, is there a system that is fully classical? Is there is system with a classical internal and external logic?

Such a theory would have to satisfy two criteria:

1. The system is formulated in classical logic.
2. Internal and external logic should coincide.

Symmetry

These requirements are easily satisfied: We formulate our theory in classical logic (as I am always doing in these lectures) and then close the system under the following two rules:

$$\text{NEC}\ \frac{\varphi}{T\overline{\varphi}} \qquad\qquad \frac{T\overline{\varphi}}{\varphi}\ \text{CONEC}$$

NEC forces the internal logic to contain the external logic, while CONEC forces the external logic to contain the internal logic, that is:

$$\{\varphi : \mathcal{A} \vdash \varphi\} \;=\; \{\varphi : \mathcal{A} \vdash T\overline{\varphi}\}$$

Obviously then internal and external logic are closed under classical logic.

Symmetry and the liar

Lemma

If $\mathcal{A}$ is consistent and closed under NEC and CONEC, neither the liar nor its negation is provable in $\mathcal{A}$.

Symmetry and the liar

Lemma

If $\mathcal{A}$ is consistent and closed under NEC and CONEC, neither the liar nor its negation is provable in $\mathcal{A}$.

The proof is that of the 'first incompleteness theorem' above:

Proof.

$\mathcal{A} \vdash \vdash \lambda \leftrightarrow \neg\overline{\lambda}$	fixed point lemma
$\mathcal{A} \vdash \vdash \lambda$	assumption
$\mathcal{A} \vdash \vdash \neg T\overline{\lambda}$	previous line
$\mathcal{A} \vdash \vdash T\overline{\lambda}$	NEC
so λ cannot be provable	
$\mathcal{A} \vdash \vdash \neg\lambda$	assumption
$\mathcal{A} \vdash \vdash T\overline{\lambda}$	first line
$\mathcal{A} \vdash \vdash \lambda$	CONEC
so $\neg\lambda$ cannot be provable	

Symmetry and the liar

Thus classical symmetric systems (ie classical systems closed under NEC and CONEC) don't decide the liar.

Symmetry and the liar

Thus classical symmetric systems (ie classical systems closed under NEC and CONEC) don't decide the liar.

Don't get worried: I don't claim that neither λ nor $\neg\lambda$ can be consistently added to $\mathcal{A}$, but after adding the liar sentence the system cannot be closed again under NEC and CONEC.

Symmetry and the liar

Thus classical symmetric systems (ie classical systems closed under NEC and CONEC) don't decide the liar.

Don't get worried: I don't claim that neither λ nor $\neg\lambda$ can be consistently added to $\mathcal{A}$, but after adding the liar sentence the system cannot be closed again under NEC and CONEC.

I take the undecidability of the liar as a nice property of symmetric systems: its provability in KF is an odd feature of KF.

Symmetry

One can easily show that closing $\mathcal{A}$ under NEC and CONEC yields a consistent theory (if $\mathcal{A}$ isn't crazily behaved), but the resulting system will be weak.

Symmetry

One can easily show that closing $\mathcal{A}$ under NEC and CONEC yields a consistent theory (if $\mathcal{A}$ isn't crazily behaved), but the resulting system will be weak.

Basically closing under NEC and CONEC means adding the 'rule version' of the T-sentences.

Symmetry

One can easily show that closing $\mathcal{A}$ under NEC and CONEC yields a consistent theory (if $\mathcal{A}$ isn't crazily behaved), but the resulting system will be weak.

Basically closing under NEC and CONEC means adding the 'rule version' of the T-sentences.

I want a stronger theory, eg a type-free version of 'Tarskian' truth, that is, the truth predicate ought to commute with the connectives and quantifiers. So I write down the $\mathcal{D}$ axioms without type restriction…

The Friedman-Sheard theory

Definition

The theory FS is given by all axioms of $\mathcal{A}$ including all induction axioms and the following axioms:

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T(x \dot{=} y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y)) \big)$

The Friedman-Sheard theory

Definition

The theory FS is given by all axioms of $\mathcal{A}$ including all induction axioms and the following axioms:

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T(x \dot{=} y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y)) \big)$
2. …and so on for all predicates other than = and T.

The Friedman-Sheard theory

Definition

The theory FS is given by all axioms of $\mathcal{A}$ including all induction axioms and the following axioms:

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T(x \doteq y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y)) \big)$
2. …and so on for all predicates other than = and T.
3. $\forall x \big(\mathrm{Sent}(x) \rightarrow (T \dot{\neg} x \leftrightarrow \neg T x \big)$

The Friedman-Sheard theory

Definition

The theory FS is given by all axioms of $\mathcal{A}$ including all induction axioms and the following axioms:

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T(x \dot{=} y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y)) \big)$
2. …and so on for all predicates other than = and T.
3. $\forall x \big(\mathrm{Sent}(x) \rightarrow (T \dot{\neg} x \leftrightarrow \neg T x \big)$
4. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (T(x \dot{\wedge} y) \leftrightarrow (Tx \wedge Ty)) \big)$

The Friedman-Sheard theory

Definition

The theory FS is given by all axioms of $\mathcal{A}$ including all induction axioms and the following axioms:

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T(x \doteq y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y)) \big)$
2. …and so on for all predicates other than = and T.
3. $\forall x \big(\mathrm{Sent}(x) \rightarrow (T \dot{\neg} x \leftrightarrow \neg T x \big)$
4. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (T(x \dot{\wedge} y) \leftrightarrow (Tx \wedge Ty) \big)$
5. $\forall x \forall y \big(\mathrm{Sent}(\dot{\forall} x y) \rightarrow (T(\dot{\forall} x y) \leftrightarrow \forall z T \mathrm{sub}(y, x \mathrm{q} z)) \big)$

The Friedman-Sheard theory

Definition

The theory FS is given by all axioms of $\mathcal{A}$ including all induction axioms and the following axioms:

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T(x \doteq y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y)) \big)$
2. …and so on for all predicates other than = and T.
3. $\forall x \big(\mathrm{Sent}(x) \rightarrow (T \dot{\neg} x \leftrightarrow \neg T x \big)$
4. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (T(x \dot{\wedge} y) \leftrightarrow (Tx \wedge Ty) \big)$
5. $\forall x \forall y \big(\mathrm{Sent}(\dot{\forall} x y) \rightarrow (T(\dot{\forall} x y) \leftrightarrow \forall z T \mathrm{sub}(y, x \mathrm{q} z)) \big)$
6. FS is closed under NEC and CONEC.

History of the Friedman-Sheard theory

Friedman & Sheard studied the theory in their 1987 paper Friedman and Sheard (1987) under a slightly different axiomatisation. They proved the consistency of the theory.

History of the Friedman-Sheard theory

Friedman & Sheard studied the theory in their 1987 paper Friedman and Sheard (1987) under a slightly different axiomatisation. They proved the consistency of the theory.

Before its actual birth (in publication) McGee damaged its reputation in McGee (1985).

Duality axioms

For FS we don't need the 'dual' axioms such as

$$\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \to \big(T \dot{\neg} (x \dot{\wedge} y) \leftrightarrow \big(T \dot{\neg} x \vee T \dot{\neg} y \big) \big) \big)$$

They are already consequences of FS.

Complete and consistent truth

Lemma

FS proves the following theorems:

1. $\forall x\big(Sent(x) \to \neg(Tx \wedge T\dot{\neg}x)\big)$ *(consistency)*
2. $\forall x\big(Sent(x) \to (Tx \vee T\dot{\neg}x)\big)$ *(completeness)*

Complete and consistent truth

Lemma

FS proves the following theorems:

1. $\forall x\big(Sent(x) \to \neg(Tx \wedge T\dot{\neg}x)\big)$ *(consistency)*
2. $\forall x\big(Sent(x) \to (Tx \vee T\dot{\neg}x)\big)$ *(completeness)*

Thus FS explicitly denies the existence of truth value gluts and gaps.

Complete and consistent truth

Lemma

FS proves the following theorems:

1. $\forall x\big(Sent(x) \rightarrow \neg(Tx \wedge T\dot{\neg}x)\big)$ *(consistency)*
2. $\forall x\big(Sent(x) \rightarrow (Tx \vee T\dot{\neg}x)\big)$ *(completeness)*

Thus FS explicitly denies the existence of truth value gluts and gaps.

The consistency axiom drops out as a consequence of Axiom 3, while it caused so much trouble in KF. Generally KF-like systems are incompatible with Axiom 3 which excludes truth value gluts and gaps.

Complete and consistent truth

Lemma

FS proves the following theorems:

1. $\forall x\big(Sent(x) \rightarrow \neg(Tx \wedge T\dot{\neg}x)\big)$ *(consistency)*
2. $\forall x\big(Sent(x) \rightarrow (Tx \vee T\dot{\neg}x)\big)$ *(completeness)*

Thus FS explicitly denies the existence of truth value gluts and gaps.

The consistency axiom drops out as a consequence of Axiom 3, while it caused so much trouble in KF. Generally KF-like systems are incompatible with Axiom 3 which excludes truth value gluts and gaps.

This shows that FS describes a thoroughly classical notion of truth.

Truth clauses

In KF we had the following axiom:

$$\forall x\Big(\mathrm{ClT}(x) \to \big(T\dot{T}x \leftrightarrow T\mathrm{val}(x)\big)\Big)$$

Such an axiom would nicely complement Axiom 1. But because of Axiom 3, this axiom also allows one to derive the full T-sentences, which makes adding such axioms impossible:

Lemma

The schema $T\overline{T\overline{\varphi}} \leftrightarrow T\overline{\varphi}$ *for all sentences is inconsistent with FS.*

Proof-theoretic analysis

Theorem (Halbach 1994)

FS is proof-theoretically equivalent to ω-iterated Tarskian truth, ie to $RA_{<\omega}$.

Proof-theoretic analysis

Theorem (Halbach 1994)

FS is proof-theoretically equivalent to ω-iterated Tarskian truth, ie to $RA_{<\omega}$.

Thus FS is much weaker than KF (or KF in partial logic), which are equivalent to $\epsilon_0 = \omega^{\omega^{:\omega}}$ and ω^ω-iterated truth.

Proof-theoretic analysis

Theorem (Halbach 1994)

FS is proof-theoretically equivalent to ω-iterated Tarskian truth, ie to $RA_{<\omega}$.

Thus FS is much weaker than KF (or KF in partial logic), which are equivalent to $\epsilon_0 = \omega^{\omega^{:\omega}}$ and ω^ω-iterated truth.

Basically an application of NEC adds an other level of Tarskian truth (although CONEC proved to be powerful on its own – Sheard (2001)).

Iterating truth

The proof-theoretical analysis reveals that FS is not good at iterations of truth. The axiom

$$\forall x\big(\mathrm{ClT}(x) \to (T\dot{T}x \leftrightarrow T\mathrm{val}(x))\big)$$

of KF allows one to prove (by induction) that long iterations of application of the truth predicate are true.

Iterating truth

The proof-theoretical analysis reveals that FS is not good at iterations of truth. The axiom

$$\forall x\big(\mathrm{ClT}(x) \to (T\dot{T}x \leftrightarrow T\mathrm{val}(x))\big)$$

of KF allows one to prove (by induction) that long iterations of application of the truth predicate are true.

In FS we can only add one truth predicate at a time by applying NEC.

Iterating truth into the transfinite

It would be nice if we had a way of proving iterations of the truth predicate into the transfinite, so we are able to show in FS that eg all sentences

$$T\overline{0 = 0},\ T\overline{T\overline{0 = 0}},\ T\overline{T\overline{T\overline{0 = 0}}}, \ldots$$

are true.

Iterating truth into the transfinite

It would be nice if we had a way of proving iterations of the truth predicate into the transfinite, so we are able to show in FS that eg all sentences

$$T\overline{0 = 0},\ T\overline{T\overline{0 = 0}},\ T\overline{T\overline{T\overline{0 = 0}}}, \ldots$$

are true.

One can try to replace NEC by reflection principles.

Iterating truth into the transfinite

Define FS_0 as the system FS but without NEC or CONEC.

Iterating truth into the transfinite

Define FS_0 as the system FS but without NEC or CONEC.

FS_{n+1} is the system FS_n with global reflection for FS_n, ie with

$$\forall x\big(\mathrm{Sent}(x) \to (\mathrm{Bew}_{\mathrm{FS}_n}(x) \to Tx)\big)$$

Iterating truth into the transfinite

Define FS_0 as the system FS but without NEC or CONEC.

FS_{n+1} is the system FS_n with global reflection for FS_n, ie with

$$\forall x\big(\mathrm{Sent}(x) \to (\mathrm{Bew}_{\mathrm{FS}_n}(x) \to Tx)\big)$$

FS_ω is then the system $\bigcup_{n\in\omega} \mathrm{FS}_n$.

$\mathrm{FS}_{\omega+1}$ is then FS_ω plus global reflection for FS_ω:

$$\forall x\big(\mathrm{Sent}(x) \to (\mathrm{Bew}_{\mathrm{FS}_\omega}(x) \to Tx)\big)$$

Iterating truth into the transfinite

Theorem

$FS_{\omega+1}$ is inconsistent.

Iterating truth into the transfinite

Theorem

$FS_{\omega+1}$ is inconsistent.

Thus one cannot consistently iterate truth beyond all finite levels in FS. It's possible to reformulate NEC to allow for more than finitely many applications in a sense, but doing it leads to an inconsistent theory.

Iterating truth into the transfinite

Theorem

$FS_{\omega+1}$ is inconsistent.

Thus one cannot consistently iterate truth beyond all finite levels in FS. It's possible to reformulate NEC to allow for more than finitely many applications in a sense, but doing it leads to an inconsistent theory.

That's strange: a theory that cannot consistently reflect on itself!

Iterating truth into the transfinite

Theorem

$FS_{\omega+1}$ is inconsistent.

Thus one cannot consistently iterate truth beyond all finite levels in FS. It's possible to reformulate NEC to allow for more than finitely many applications in a sense, but doing it leads to an inconsistent theory.

That's strange: a theory that cannot consistently reflect on itself!

Theorem

FS plus global reflection for FS

$$\forall x\big(Sent(x) \to (Bew_{FS}(x) \to Tx)\big)$$

is inconsistent.

McGee's theorem on ω-inconsistency

The inconsistency results are a consequence of a theorem due to McGee McGee (1985).

McGee's theorem on ω-inconsistency

The inconsistency results are a consequence of a theorem due to McGee McGee (1985).

So far all paradoxes have been purely 'modal', in the sense that they did not involve quantification. McGe's paradox involves a quantifier.

McGee's theorem on ω-inconsistency

Theorem (McGee 1985)

Assume that $\mathcal{A}$ proves the following schemes for all sentences φ and ψ and all formulas $\chi(\mathrm{v})$ having at most v *free.*

(i) NEC

(ii) $T\overline{\varphi \to \psi} \to (T\overline{\varphi} \to T\overline{\psi})$

(iii) $T\overline{\neg\varphi} \to \neg T\overline{\varphi}$

(iv) $\forall x T\overline{\chi(\dot{x})} \to T\overline{\forall x \chi(x)}$

Then $\mathcal{A}$ is ω-inconsistent, that is, $\mathcal{A} \vdash \zeta(\overline{n})$ for all $n \in \omega$ and $\mathcal{A} \vdash \neg\forall n \zeta(n)$

FS proves all these sentences and is closed under NEC.

McGee's theorem on ω-inconsistency: the proof

I assume throughout this section that $\mathcal{A}$ satisfies all axioms listed on the previous page. I shall apply the diagonalization theorem twice; this is the first instance:

(18)

$$\mathcal{A} \vdash \gamma(n, x, y) \leftrightarrow \exists k(n = k\hat{}\overline{1} \wedge x = \overline{\forall x(\gamma(\dot{k}, x, \dot{y}) \rightarrow Tx)}) \vee (n = \overline{0} \wedge x = y)$$

McGee's theorem on ω-inconsistency: the proof

I assume throughout this section that $\mathcal{A}$ satisfies all axioms listed on the previous page. I shall apply the diagonalization theorem twice; this is the first instance:

(18)

$$\mathcal{A} \vdash \gamma(n, x, y) \leftrightarrow \exists k(n = k^\frown \overline{1} \wedge x = \overline{\forall x(\gamma(\dot{k}, x, \dot{y}) \rightarrow Tx)}) \vee (n = \overline{0} \wedge x = y)$$

The free variables n, x, y do not present a problem because nothing prevents free variables from occurring in the diagonalized formula. The relativization to number variables n is unproblematic, as well.

McGee's theorem on ω-inconsistency: the proof

Roughly speaking, the use of γ is the following. Assume a sentence σ has been fixed. Then $\gamma(\underline{0}, x, \overline{\sigma})$ holds, if x is the sentence σ. $\gamma(\overline{1}, x, \overline{\sigma})$ holds if x is the sentence $\forall x(\gamma(\underline{0}, x, \overline{\sigma}) \to Tx)$, that is, x is a sentence equivalent to $T\overline{\sigma}$. Similarly, $\gamma(\overline{2}, x, \overline{\sigma})$ says that x is a certain sentence equivalent to $T\overline{T\overline{\sigma}}$.

McGee's theorem on ω-inconsistency: the proof

The second application of the diagonalization theorem 12 is straightforward:

(19) $$\sigma \leftrightarrow \neg \forall n \forall x (\gamma(n, x, \overline{\sigma}) \rightarrow Tx)$$

McGee's theorem on ω-inconsistency: the proof

The following lemma holds also for arbitrary sentences in place of σ.

Lemma

(i) $\mathcal{A} \vdash \gamma(\underline{0}, x, \overline{\sigma}) \leftrightarrow x = \overline{\sigma}$

(ii) $\mathcal{A} \vdash \gamma(n^\frown\overline{1}, x, \overline{\sigma}) \leftrightarrow x = \overline{\forall x(\gamma(\dot{n}, x, \overline{\sigma}) \to Tx)}$

McGee's theorem on ω-inconsistency: the proof

Proof.

From A6 we we derive $n = \underline{0} \to \neg\exists k\, n = k\hat{}\overline{1}$, which in turn implies the following:

$$n = \underline{0} \to (\gamma(n, x, y) \leftrightarrow (n = \overline{0} \wedge x = y))$$

This yields (i). (ii) is proved similarly. ⊣

McGee's theorem on ω-inconsistency: the proof

From (19) we obtain the following:

$$\begin{aligned} \mathcal{A} \vdash \neg\sigma &\leftrightarrow \forall n \forall x(\gamma(n, x, \overline{\sigma}) \to Tx) \\ &\to \forall x(\gamma(\underline{0}, x, \overline{\sigma}) \to Tx) \\ &\to T\overline{\sigma} \end{aligned} \tag{20}$$

Lemma 78 (i)

McGee's theorem on ω-inconsistency: the proof

An application of NEC and axiom (ii) of theorem 77 to (19) yields the following equivalence:

$$
\begin{aligned}
\mathcal{A} \vdash T\overline{\sigma} &\to T\overline{\neg\forall n\forall x(\gamma(n,x,\overline{\sigma}) \to Tx)} \\
&\to \neg T\overline{\forall n\forall x(\gamma(n,x,\overline{\sigma}) \to Tx)} && \text{axiom (iii)} \\
&\to \neg\forall n T\overline{\forall x(\gamma(\overset{\bullet}{n},x,\overline{\sigma}) \to Tx)} && \text{axiom (iv)} \\
&\to \neg\forall n\forall x(x = \overline{\forall x(\gamma(\overset{\bullet}{n},x,\overline{\sigma}) \to Tx)} \to Tx) \\
&\to \neg\forall n\forall x(\gamma(n^\frown\overline{1},x,\overline{\sigma}) \to Tx) && \text{Lemma 78 (ii)} \\
&\to \neg\forall n\forall x(\gamma(n,x,\overline{\sigma}) \to Tx) && \text{Lemma } \mathbf{??} \\
(21) \qquad &\to \sigma
\end{aligned}
$$

McGee's theorem on ω-inconsistency: the proof

Taken together, (20) and (21) imply $\mathcal{A} \vdash \sigma$, which in turn implies by (19)

(22) $$\mathcal{A} \vdash \neg \forall n \forall x (\gamma(n, x, \overline{\sigma}) \rightarrow Tx)$$

McGee's theorem on ω-inconsistency: the proof

Taken together, (20) and (21) imply $\mathcal{A} \vdash \sigma$, which in turn implies by (19)

(22) $$\mathcal{A} \vdash \neg \forall n \forall x (\gamma(n, x, \overline{\sigma}) \rightarrow Tx)$$

From $\mathcal{A} \vdash \sigma$, however, we have also by NEC:

McGee's theorem on ω-inconsistency: the proof

	$\mathcal{A} \vdash T\overline{\sigma}$	
(23)	$\mathcal{A} \vdash \forall x(\gamma(\underline{0}, x, \overline{\sigma}) \to Tx)$	Lemma 78 (i)
	$\mathcal{A} \vdash T\overline{\forall x(\gamma(\underline{0}, x, \overline{\sigma}) \to Tx)}$	NEC
	$\mathcal{A} \vdash \forall x(x = \overline{\forall x(\gamma(\underline{0}, x, \overline{\sigma}) \to Tx)} \to Tx)$	
	$\mathcal{A} \vdash \forall x(\gamma(\underline{0}\hat{}\,\overline{1}, x, \overline{\sigma}) \to Tx)$	Lemma 78 (ii)
(24)	$\mathcal{A} \vdash \forall x(\gamma(\overline{1}, x, \overline{\sigma}) \to Tx)$	axiom A2
	$\mathcal{A} \vdash T\overline{\forall x(\gamma(\overline{1}, x, \overline{\sigma}) \to Tx)}$	NEC
	$\mathcal{A} \vdash \forall x(x = \overline{\forall x(\gamma(\overline{1}, x, \overline{\sigma}) \to Tx)} \to Tx)$	
	$\mathcal{A} \vdash \forall x(\gamma(\overline{1}\hat{}\,\overline{1}, x, \overline{\sigma}) \to Tx)$	Lemma 78 (ii)
(25)	$\mathcal{A} \vdash \forall x(\gamma(\overline{2}, x, \overline{\sigma}) \to Tx)$	axiom A2
	$\vdots$	

McGee's theorem on ω-inconsistency: the proof

(22) and the sequence of lines continuing (23), (24) and (25) establish the ω-inconsistency of $\mathcal{A}$ according to definition **??**.

More ω-inconsistencies

Further ω-inconsistencies of truth theories have been established by Visser (1989b), Yablo (1993) and Leitgeb (2001).

Truth & Paradox

VIII · The Real Culprit

Volker Halbach

Nordic Logic Summer School 2017

The cases

In the lectures we have encountered various paradoxes and inconsistencies:

- The T-sentences

The cases

In the lectures we have encountered various paradoxes and inconsistencies:

- The T-sentences
- Montague's paradox

The cases

In the lectures we have encountered various paradoxes and inconsistencies:

- The T-sentences
- Montague's paradox
- Gödel's second incompleteness theorem (inconsistency of $\neg T\overline{\bot}$)

The cases

In the lectures we have encountered various paradoxes and inconsistencies:

- The T-sentences
- Montague's paradox
- Gödel's second incompleteness theorem (inconsistency of $\neg T\overline{\bot}$)
- McGee's paradox

The cases

In the lectures we have encountered various paradoxes and inconsistencies:

- The T-sentences
- Montague's paradox
- Gödel's second incompleteness theorem (inconsistency of $\neg T\ulcorner\bot\urcorner$)
- McGee's paradox

and there are many more paradoxes (Yablo's paradox, also various inconsistencies in Friedman and Sheard (1987)).

Yablo's paradox

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.

$\vdots$

The victims

Various theories have been proved to be inconsistent. The paradoxes do not only concern truth but also various forms of necessity, temporal notions, deontic notions, if treated as diagonalisable predicates.

The victims

Various theories have been proved to be inconsistent. The paradoxes do not only concern truth but also various forms of necessity, temporal notions, deontic notions, if treated as diagonalisable predicates.

The unrestricted T-sentences, the predicate version of the modal system T didn't survive. FS got away scarred for life.

The suspects

- The liar? He is certainly involved in Montague's paradox and the liar paradox. But is he pulling the strings?

The suspects

- The liar? He is certainly involved in Montague's paradox and the liar paradox. But is he pulling the strings?
- McGee? The liar is hardly sophisticated enough to have eliminated FS via ω-inconsistency. McGee's fixed point looks more sophisticated.

The suspects

- The liar? He is certainly involved in Montague's paradox and the liar paradox. But is he pulling the strings?
- McGee? The liar is hardly sophisticated enough to have eliminated FS via ω-inconsistency. McGee's fixed point looks more sophisticated.
- Löb? He is a good citizen; after all it's Löb's *theorem*, not Löb's paradox.

The possible worlds perspective

Many of the paradoxes can be visualised by possible worlds semantics.

The possible worlds perspective

Many of the paradoxes can be visualised by possible worlds semantics.

Truth-theoretic axioms relate to modal principles, eg, in Montague's paradox.
The liar paradox is just one of many modal paradoxes.

The possible worlds perspective

Many of the paradoxes can be visualised by possible worlds semantics.

Truth-theoretic axioms relate to modal principles, eg, in Montague's paradox. The liar paradox is just one of many modal paradoxes.

How can one develop possible worlds semantics, if necessity is treated as a predicate rather than as a modal operator?

The possible worlds perspective

Many of the paradoxes can be visualised by possible worlds semantics.

Truth-theoretic axioms relate to modal principles, eg, in Montague's paradox. The liar paradox is just one of many modal paradoxes.

How can one develop possible worlds semantics, if necessity is treated as a predicate rather than as a modal operator?

All the 'problems' for the predicate approach arise only because of the increased expressive power of modal predicates.

The following is based on Halbach et al. (2003). The idea is simple but …

The syntactic vocabulary will be interpreted at all worlds standardly.

The syntactic vocabulary will be interpreted at all worlds standardly.

From now on we assume that our language also contains 'contingent' vocabulary, namely a sentence letter p. More contingent vocabulary is ok.

Call the language $\mathcal{L}_{\text{min}}$.

A standard model for the language $\mathcal{L}_{\text{min}}$ without T interprets the syntactic vocabulary in the intended way and assigns some interpretations to the other vocabulary, in particular to p. The domain of a standard model is always the set of all $\mathcal{L}_{\text{min}}$-expressions (finite strings of symbols).

A standard model for the language $\mathcal{L}_{\text{min}}$ without T interprets the syntactic vocabulary in the intended way and assigns some interpretations to the other vocabulary, in particular to p. The domain of a standard model is always the set of all $\mathcal{L}_{\text{min}}$-expressions (finite strings of symbols).

Let S be a set of expressions and $\mathcal{E}$ a standard model. A sentence φ of $\mathcal{L}_{\text{min}}$ holds in $\langle \mathcal{E}, S \rangle$—formally $\langle \mathcal{E}, S \rangle \vDash \varphi$ if and only if φ holds when all nonlogical symbols of $\mathcal{L}_{\text{min}}$ except T are interpreted according to the model $\mathcal{E}$ and T is interpreted by S. S is called the extension of T in the model $\langle \mathcal{E}, S \rangle$.

A standard model for the language $\mathcal{L}_{\text{min}}$ without T interprets the syntactic vocabulary in the intended way and assigns some interpretations to the other vocabulary, in particular to p. The domain of a standard model is always the set of all $\mathcal{L}_{\text{min}}$-expressions (finite strings of symbols).

Let S be a set of expressions and $\mathcal{E}$ a standard model. A sentence φ of $\mathcal{L}_{\text{min}}$ holds in $\langle \mathcal{E}, S \rangle$—formally $\langle \mathcal{E}, S \rangle \vDash \varphi$ if and only if φ holds when all nonlogical symbols of $\mathcal{L}_{\text{min}}$ except T are interpreted according to the model $\mathcal{E}$ and T is interpreted by S. S is called the extension of T in the model $\langle \mathcal{E}, S \rangle$.

We have:

$$\langle \mathcal{E}, S \rangle \vDash T\overline{e} \text{ iff } e \in S$$

Frames

Frames are defined as in operator modal logic (ie the usual modal logic):

Definition

A frame is an ordered pair $\langle W, R \rangle$ where

- $W \neq \emptyset$ (the set of worlds),
- $R \subseteq W \times W$ (the accessibility relation).

Frames

Frames are defined as in operator modal logic (ie the usual modal logic):

Definition

A frame is an ordered pair $\langle W, R \rangle$ where

- $W \neq \emptyset$ (the set of worlds),
- $R \subseteq W \times W$ (the accessibility relation).

T should behave like the box in modal logic: $T\overline{\varphi}$ should be true at a world w iff φ is true at all worlds that can be 'seen' from w.

Definition

A valuation V for a frame $\langle W, R \rangle$ is a function that assigns to every $w \in W$ a standard model $\mathcal{E}$.

Definition

A valuation V for a frame $\langle W, R \rangle$ is a function that assigns to every $w \in W$ a standard model $\mathcal{E}$.

Definition

A T-interpretation is a function that assigns to each world w a set S of expressions.

Thus a T-interpretation tells us how interpret T a t a given world w.

Definition

A *PW-model* is a quadruple $\langle W, R, V, B\rangle$ such that $\langle W, R\rangle$ is a frame, V is a valuation for $\langle W, R\rangle$ and B is a T-interpretation for $\langle W, R\rangle$ satisfying the following condition:

$$B(w) = \{\varphi \in \mathcal{L}_{\text{min}} : \text{ for all } u \in W : (\text{if } wRu, \text{ then } \langle V(u), B(u)\rangle \vDash \varphi)\}$$

Truth at a world cannot be *defined* by induction on the complexity of sentences.

$\langle V(u), B(u)\rangle$ is always a standard model and $\langle V(u), B(u)\rangle \vDash \varphi$ means that φ is true in the standard model $\langle V(u), B(u)\rangle$ in the usual sense of first-order predicate logic.

Definition

A frame $\langle W, R\rangle$ admits a PW-model on *every* valuation iff for every valuation V on $\langle W, R\rangle$ there is a such that B such that $\langle W, R, V, B\rangle$ is a PW-model.

Definition

A frame $\langle W, R \rangle$ admits a PW-model on *every* valuation iff for every valuation V on $\langle W, R \rangle$ there is a such that B such that $\langle W, R, V, B \rangle$ is a PW-model.

Definition

A frame admits a PW-model iff the frame admits a PW-model on some valuation, that is, iff there is a valuation V such that $\langle W, R, V, B \rangle$ is a PW-model model.

Definition

A frame $\langle W, R \rangle$ admits a PW-model on *every* valuation iff for every valuation V on $\langle W, R \rangle$ there is a such that B such that $\langle W, R, V, B \rangle$ is a PW-model.

Definition

A frame admits a PW-model iff the frame admits a PW-model on some valuation, that is, iff there is a valuation V such that $\langle W, R, V, B \rangle$ is a PW-model model.

Strong Characterization problem

Which frames admit a PW-model on every valuation?

Lemma (normality)

Suppose $\langle W, R, V, B\rangle$ *is a* PW*-model,* $w \in W$ *and* φ, ψ *sentences of* $\mathcal{L}_{min}$*. Then the following holds for all* $w \in W$*:*

1. *If* $\langle V(u), B(u)\rangle \vDash \varphi$ *for all* $u \in W$*, then* $\langle V(w), B(w)\rangle \vDash T\overline{\varphi}$*.*
2. $\langle V(w), B(w)\rangle \vDash T\overline{\varphi \to \psi} \to (T\overline{\varphi} \to T\overline{\psi})$

Lemma

1. *If a frame* $\langle W, R\rangle$ *is transitive and* $\langle W, R, V, B\rangle$ *a* PW*-model on that frame, we have for all sentences* φ *in* $\mathcal{L}_{min}$ *and worlds* $w \in W$*:*

$$\langle V(w), B(w)\rangle \vDash T\overline{\varphi} \rightarrow T\overline{T\overline{\varphi}}$$

2. *If a frame* $\langle W, R\rangle$ *is reflexive and* $\langle W, R, V, B\rangle$ *a* PW*-model on that frame, we have for all sentences* φ *in* $\mathcal{L}_{min}$ *and worlds* $w \in W$*:*

$$\langle V(w), B(w)\rangle \vDash T\overline{\varphi} \rightarrow \varphi$$

In a frame where every world sees exactly itself we have the T-sentences at any $w \in W$.

The paradoxes

Theorem (liar paradox)

The frame $\langle W_1, R_1 \rangle$ does not admit a PW*-model.*

Example (Montague)

If $\langle W, R\rangle$ admits a PW-model, then $\langle W, R\rangle$ is not reflexive.

Proof.

Assume $\langle W, R, V, B\rangle$ is a PW-model with R reflexive. φ is the liar sentence, i.e., PA $\vdash \varphi \leftrightarrow \neg T\overline{\varphi}$. For arbitrary $w \in W$ we have:

(26) $V(w) \vDash T\overline{\varphi} \rightarrow \neg\varphi$

(27) $V(w) \vDash T\overline{\varphi} \rightarrow \varphi$

(28) $V(w) \vDash \neg T\overline{\varphi}$

(29) $V(w) \vDash \varphi$

By normality $V(w) \vDash T\overline{\varphi}$. Contradiction!

Example (Montague)

If $\langle W, R\rangle$ admits a PW-model, then $\langle W, R\rangle$ is not reflexive.

Proof.

Assume $\langle W, R, V, B\rangle$ is a PW-model with R reflexive. φ is the liar sentence, i.e., $\mathrm{PA} \vdash \varphi \leftrightarrow \neg T\overline{\varphi}$. For arbitrary $w \in W$ we have:

$$V(w) \vDash T\overline{\varphi} \rightarrow \neg\varphi \tag{26}$$

$$V(w) \vDash T\overline{\varphi} \rightarrow \varphi \tag{27}$$

$$V(w) \vDash \neg T\overline{\varphi} \tag{28}$$

$$V(w) \vDash \varphi \tag{29}$$

By normality $V(w) \vDash T\overline{\varphi}$. Contradiction! ⊣

Montague's paradox is more general than the liar paradox. But there are even more general results.

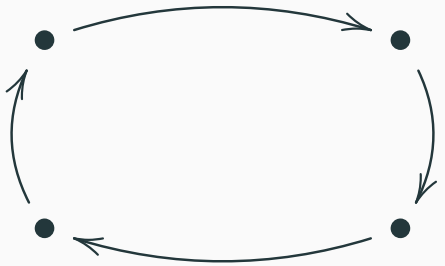
Example

The frame ‘two worlds see each another’ displayed above does not admit a PW-model.

Proof.

fixed point $\varphi \leftrightarrow \neg T\overline{T\overline{\varphi}}$. ⊣

The following frame does not admit a PW-model.



Example

The frame 'one world sees itself and one other world' does not admit a PW-model.

We call the frame $\langle W_3, R_3 \rangle$.

One can use the fixed point $\varphi \leftrightarrow (T\overline{\varphi} \to T\overline{\neg\varphi})$

Gödel's second theorem

Proposition

In a transitive frame every world must either see a dead end or be a dead end.

Gödel's second theorem

Proposition

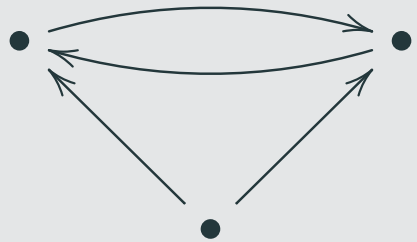
In a transitive frame every world must either see a dead end or be a dead end.

Proof.

We have proved in $T\overline{\bot} \vee \neg T\overline{\neg T\overline{\bot}}$ in K4. ⊣

Proposition

The following frame does not admit a valuation:



Proof.

$\varphi \leftrightarrow \neg T\overline{T\overline{\varphi}} \wedge \neg T\overline{\varphi}$ ⊣

Example

The frame $\langle \omega, \mathrm{Pre} \rangle$ does not admit a PW-model. Here ω is the set of all natural numbers and Pre is the relation with all pairs $\langle n, n+1 \rangle$. Hence every world n sees $n+1$ but no other world.

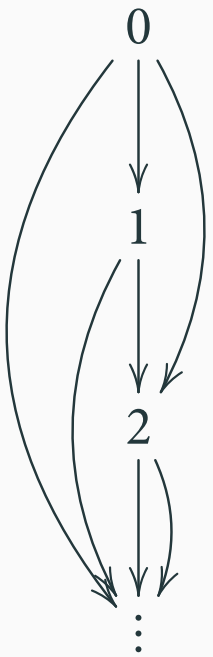
The frame $\langle \omega, \mathrm{Pre} \rangle$ can be displayed by the following diagram:

This can be shown with the fixed point used in McGee's ω-inconsistency proof.

Theorem

The frame $\langle \omega, < \rangle$ does not admit a PW*-model. Here $<$ is the usual 'smaller than' relation on the natural numbers:*

The frame $\langle \omega, < \rangle$ can be displayed by the following diagram:



Frames with possible worlds models

Example

The frame $\langle \{w\}, \emptyset \rangle$ admits a PW-model on every valuation.

By Suc we denote the successor relation $\{\langle k, n\rangle : \ k = n + 1\}$ on the set ω of natural numbers.

Example

The frame $\langle \omega, \mathrm{Suc}\rangle$ admits a PW-model on every valuation.

Definition

A frame $\langle W, R \rangle$ is converse wellfounded (or Noetherian) iff for every non-empty $M \subseteq W$ there is a $w \in M$ that is R-maximal in M.

Definition

A frame $\langle W, R \rangle$ is converse wellfounded (or Noetherian) iff for every non-empty $M \subseteq W$ there is a $w \in M$ that is R-maximal in M.

Lemma

Every converse wellfounded frame $\langle W, R \rangle$ admits a PW*-model on every valuation.*

The Strong Characterization Problem

Theorem (Strong Characterization theorem)

A frame admits a PW*-model on every valuation iff it is converse wellfounded.*

We define T^*x as $\forall n T \underbrace{\dot{T} \ldots \dot{T}}_{n} x$.

T^* can be defined properly using the diagonal lemma.

T^* corresponds ot the transitive closure of R.

Lemma

For all sentences φ and ψ und PW-models $\langle W, R, V, B\rangle$ the following holds:

1. *If $\langle V(w), B(w)\rangle \vDash \varphi$ for all $w \in W$, then $\langle V(w), B(w)\rangle \vDash T^*\overline{\varphi}$.*
2. $\langle V(w), B(w)\rangle \vDash T^*\overline{\varphi \to \psi} \to (T^*\overline{\varphi} \to T^*\overline{\psi})$
3. $\langle V(w), B(w)\rangle \vDash T^*\overline{\varphi} \to T^*\overline{T^*\overline{\varphi}}$
4. $\langle V(w), B(w)\rangle \vDash T^*\overline{T^*\overline{\varphi} \to \varphi} \to T^*\overline{\varphi}$

Lemma

The transitive closure R^* of the accessibility relation R of any PW*-model that admits a* PW*-model on every valuation is converse wellfounded.*

Lemma

A frame $\langle W, R\rangle$ is converse wellfounded iff its transitive closure $\langle W, R^*\rangle$ is converse wellfounded.

This concludes the proof of the Strong Characterization theorem.

What happens if we drop the sentence letter p from the language?

A sufficient condition

Definition

The *depth* of a world w is defined to be the rank of R^{-1} restricted to the converse wellfounded part of $\{v \in W : wRv\}$.

A sufficient condition

Definition

The *depth* of a world w is defined to be the rank of R^{-1} restricted to the converse wellfounded part of $\{v \in W : wRv\}$.

Let γ be least so that L_γ has a transitive Σ_1-end extension.

A sufficient condition

Definition

The *depth* of a world w is defined to be the rank of R^{-1} restricted to the converse wellfounded part of $\{v \in W : wRv\}$.

Let γ be least so that L_γ has a transitive Σ_1-end extension.

Proposition (Aczel Aczel and Richter (1973))

γ is the least ordinal such such that L_γ is first order-reflecting, that is, if φ is any formula (with parameters from L_γ) in the language of set theory, then

$$L_\gamma \vDash \varphi \quad \Longrightarrow \quad \exists \alpha < \gamma \; L_\alpha \vDash \varphi.$$

Remark

γ is an admissible that is much larger than ω_1^{CK}.

A sufficient condition

Theorem

Assume $\langle W, R \rangle$ is transitive and every converse illfounded world in $\langle W, R \rangle$ has depth at least γ. Then $\langle W, R \rangle$ admits a valuation.

A necessary condition

Definition

ADM^* is the class of all admissible ordinals without ω together with all of its limit points.

A necessary condition

Definition

ADM^* is the class of all admissible ordinals without ω together with all of its limit points.

Theorem

If $\langle W, R\rangle$ admits a valuation, then in the transitive closure $\langle W, R^*\rangle$ of $\langle W, R\rangle$ every converse illfounded world in W has depth α with $\alpha \in \mathrm{ADM}^*$ or $\alpha \geq \gamma$.

A necessary condition

Definition

ADM^* is the class of all admissible ordinals without ω together with all of its limit points.

Theorem

If $\langle W, R\rangle$ admits a valuation, then in the transitive closure $\langle W, R^*\rangle$ of $\langle W, R\rangle$ every converse illfounded world in W has depth α with $\alpha \in \mathrm{ADM}^*$ or $\alpha \geq \gamma$.

This is proved via Löb's theorem.

A necessary condition

Definition

ADM^* is the class of all admissible ordinals without ω together with all of its limit points.

Theorem

If $\langle W, R\rangle$ admits a valuation, then in the transitive closure $\langle W, R^*\rangle$ of $\langle W, R\rangle$ every converse illfounded world in W has depth α with $\alpha \in \mathrm{ADM}^*$ or $\alpha \geq \gamma$.

This is proved via Löb's theorem.

Remark

There are PW-models with converse illfounded worlds of depth ω_1^{CK}.

A necessary condition

Theorem

If $\langle W, R\rangle$ *admits a valuation, then in the transitive closure* $\langle W, R^*\rangle$ *of* $\langle W, R\rangle$ *every converse illfounded world in* W *has depth* α *with* $\alpha \in \mathrm{ADM}^*$ *or* $\alpha \geq \gamma$.

A necessary condition

Theorem

If $\langle W, R \rangle$ *admits a valuation, then in the transitive closure* $\langle W, R^* \rangle$ *of* $\langle W, R \rangle$ *every converse illfounded world in* W *has depth* α *with* $\alpha \in \mathrm{ADM}^*$ *or* $\alpha \geq \gamma$.

This theorem implies all other limitative results mentioned so far: the liar paradox, Montague's paradox, McGee's paradox, Löb's theorem, Gödel's second theorem, 'two worlds see one another',…

Transitive frames

Theorem

If a transitive frame $\langle W, R \rangle$ admits a valuation, then:

1. *All worlds w with depth smaller than ω_1^{CK} have to be converse wellfounded.*
2. *For worlds with depth greater or equal to γ there are no restrictions.*
3. *All converse illfounded worlds with depth between ω_1^{CK} and γ must have depth in ADM^*. Only for those worlds our results don't provide full information.*

General frames

Theorem 86 cannot be generalized to non-transitive frames:

General frames

Theorem 86 cannot be generalized to non-transitive frames:

Proposition

There are frames that do not admit a valuation, although their transitive closure does.

General frames

Theorem 86 cannot be generalized to non-transitive frames:

Proposition

There are frames that do not admit a valuation, although their transitive closure does.

Proposition

If $\langle W, R \rangle$ is converse wellfounded, then $\langle W, R \rangle$ admits a valuation.

Extensions

All worlds share the same domain, the set of all expressions. The definition of a standard model can be generalized, so that a standard model also specifies a set of 'contingent' objects. Since some of them may lack names, the unary predicate T is replaced with a binary predicate applying to formulae and sequences of objects (variable assignments). This requires a theory sequences of arbitrary objects (symbols or contingent objects).

If we take de re-modality into account, we shouldn't use a unary predicate True , but rather a binary predicate applying to formulae and variable assignments.

de re-truth is just satisfaction.

We could consider a bimodal setting with two modal predicates.

Conclusion

Löb's theorem is the mother of all paradoxes – as long as the modality is 'normal'.

Various predicate modalities aren't normal in this sense, for instance, KF.

References

Peter Aczel and Wayne Richter. Inductive definitions and reflecting properties of admissible ordinals. In Jens E. Fenstad and Peter Hinman, editors, *Generalized Recursion Theory*, pages 301–381. North Holland, 1973.

George Boolos. *The Logic of Provability*. Cambridge University Press, Cambridge, 1993.

John P. Burgess. The Truth Is Never Simple. *Journal of Symbolic Logic*, 51:663–81, September 1986.

John P. Burgess. Addendum to 'The Truth Is Never Simple' . *Journal of Symbolic Logic*, 53: 390–92, June 1988.

Cezary Cieśliński. Deflationism, Conservativeness and Maximality. *Journal of Philosophical Logic*, 36:695–705, 2007.

Ali Enayat and Albert Visser. New Constructions of Satisfaction Classes. In T. Achourioti, H. Galinon, K. Fujimoto, and J. Martinez-Fernandez, editors, *Unifying the Philosophy of Truth*, pages 321–335. Springer, 2015.

Solomon Feferman. Reflecting on Incompleteness. *Journal of Symbolic Logic*, 56:1–49, 1991.

Hartry Field. Deflating the Conservativeness Argument. *Journal of Philosophy*, 96:533–540, 1999.

Harvey Friedman and Michael Sheard. An Axiomatic Approach to Self-Referential Truth. *Annals of Pure and Applied Logic*, 33:1–21, 1987.

Volker Halbach. A system of complete and consistent truth. *Notre Dame Journal of Formal Logic*, 35:311–327, 1994.

Volker Halbach. Tarskian and Kripkean truth. *Journal of Philosophical Logic*, 26:69–80, 1997.

Volker Halbach. Disquotationalism and infinite conjunctions. *Mind*, 108:1–22, 1999a.

Volker Halbach. Conservative Theories of Classical Truth. *Studia Logica*, 62:353–70, 1999b.

Volker Halbach and Leon Horsten. Axiomatizing Kripke's Theory of Truth. *Journal of Symbolic Logic*, 71:677–712, 2006.

Volker Halbach, Hannes Leitgeb, and Philip Welch. Possible Worlds Semantics For Modal Notions Conceived As Predicates. *Journal of Philosophical Logic*, 32:179–223, 2003.

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

Paul Horwich. *Truth*. Basil Blackwell, Oxford, first edition, 1990.

Richard Kaye. *Models of Peano Arithmetic*. Oxford Logic Guides. Oxford University Press, Oxford, 1991.

Jeffrey Ketland. Deflationism and Tarski's Paradise. *Mind*, 108:69–94, 1999.

Henryk Kotlarski, Stanislav Krajewski, and Alistair Lachlan. Construction of Satisfaction Classes for Nonstandard Models. *Canadian Mathematical Bulletin*, 24:283–293, 1981.

Georg Kreisel and Azriel Lévy. Reflection principles and their use for establishing the complexity of axiomatic systems. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, 14:97–142, 1968.

Saul Kripke. Outline of a Theory of Truth. *Journal of Philosophy*, 72:690–716, 1975. reprinted in **?**.

Alistair Lachlan. Full Satisfaction Classes and Recursive Saturation. *Canadian Mathematical Bulletin*, 24:295–297, 1981.

Graham E. Leigh. Conservativity for theories of compositional truth via cut elimination. *jsl*, 80: 825–865, 2015.

Hannes Leitgeb. Theories of Truth which have no Standard Models. *Studia Logica*, 21:69–87, 2001.

Robert L. Martin and Peter W. Woodruff. On representing 'true-in-L' in L. *Philosophia*, 5: 213–217, 1975. Reprinted in **?**.

Vann McGee. How Truthlike Can a Predicate Be? A Negative Result. *Journal of Philosophical Logic*, 14:399–410, 1985.

Vann McGee. *Truth, Vagueness, and Paradox: An Essay on the Logic of Truth*. Hackett Publishing, Indianapolis and Cambridge, 1991.

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

Vann McGee. In Praise of the Free Lunch: why Disquotationalists Should Embrace Compositional Semantics. manuscript, 2003.

Richard Montague. Syntactical Treatments of Modality, with Corollaries on Reflexion Principles and Finite Axiomatizability. *Acta Philosophica Fennica*, 16:153–67, 1963. Reprinted in (**?**, 286–302).

William Reinhardt. Some Remarks on Extending and Interpreting Theories with a Partial Predicate for Truth. *Journal of Philosophical Logic*, 15:219–251, 1986.

Stewart Shapiro. Proof and Truth: Through Thick and Thin. *Journal of Philosophy*, 95:493–521, 1998.

Michael Sheard. Weak and Strong Theories of Truth. *Studia Logica*, 68:89–101, 2001.

Stuart Smith. Nonstandard characterizations of recursive saturation and resplendency. *Journal of Symbolic Logic*, 52:842–863, 1987.

Richmond H. Thomason. A Note on Syntactical Treatments of Modality. *Synthese*, 44:391–396, 1980.

Albert Visser. Semantics and the Liar Paradox. In D[ov] M. Gabbay and F[ranz] Guenthner, editors, *Topics in the Philosophy of Language*, volume 4 of *Handbook of Philosophical Logic*, pages 617–706. Reidel, Dordrecht, 1989a.

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

Stephen Yablo. Paradox Without Self-Reference. *Analysis*, 53:251–252, 1993.