SESSION 1

Truth, Modality and Paradox

Volker Halbach

Fudan

15th May 2012

Why formal theories of truth & modality?

- Solutions to the liar paradox and other paradoxes

Why formal theories of truth & modality?

- 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 & modality?

- 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?
- In particular, is truth a substantial notion that allows one new insights into non-semantic issues?

Why formal theories of truth & modality?

- 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?
- 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 & modality?

- 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?
- 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 & modality?

- 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?
- 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
- formal framework for thinking about truth and modalities

Concepts I'm interested in

- truth

Concepts I'm interested in

- truth
- necessity

Concepts I'm interested in

- truth
- necessity
- analyticity

Concepts I'm interested in

- truth
- necessity
- analyticity
- a priori

Concepts I'm interested in

- truth
- necessity
- analyticity
- a priori
- knowledge

Concepts I'm interested in

- truth
- necessity
- analyticity
- a priori
- knowledge
- future and past truth

Concepts I'm interested in

- truth
- necessity
- analyticity
- a priori
- knowledge
- future and past truth

I'll focus mainly on truth but many remarks can generalised.

The predicate and operator analyses

Modal notions such as necessity can be formalised as (sentential) operators or as predicates.

The predicate and operator analyses

Modal notions such as necessity can be formalised as (sentential) operators or as predicates.

Necessity is formalised as an operator in modal logic; truth is usually treated as a predicate.

The operator analysis

'It is necessary that' is conceived as an operator or adverbial phrase: combined with a sentence it gives another sentence.

$$\underbrace{\text{It is necessary that}}_{\text{operator, adverb}}\ \underbrace{\text{water is } H_2O}_{\text{sentence}}.$$

The operator analysis

'It is necessary that' is conceived as an operator or adverbial phrase: combined with a sentence it gives another sentence.

$$\underbrace{\text{It is necessary that}}_{\text{operator, adverb}}\ \underbrace{\text{water is } H_2O}_{\text{sentence}}.$$

According to the Operator Analysis sentences with 'that' can be parsed into an operator (or adverbial phrase) and a sentence.

The operator analysis

'It is necessary that' is conceived as an operator or adverbial phrase: combined with a sentence it gives another sentence.

$$\underbrace{\text{It is necessary that}}_{\text{operator, adverb}}\ \underbrace{\text{water is } H_2O}_{\text{sentence}}.$$

According to the Operator Analysis sentences with 'that' can be parsed into an operator (or adverbial phrase) and a sentence.

In English 'necessarily' is an adverb acting in this way.

The operator analysis

'It is necessary that' is conceived as an operator or adverbial phrase: combined with a sentence it gives another sentence.

$$\underbrace{\text{It is necessary that}}_{\text{operator, adverb}}\ \underbrace{\text{water is } H_2O}_{\text{sentence}}.$$

According to the Operator Analysis sentences with 'that' can be parsed into an operator (or adverbial phrase) and a sentence.

In English 'necessarily' is an adverb acting in this way.

In modal logic necessity is conceived as the sentential operator $\Box$: If φ is a sentence $\Box\varphi$ is also a sentence.

The operator analysis

'It is necessary that' is conceived as an operator or adverbial phrase: combined with a sentence it gives another sentence.

$$\underbrace{\text{It is necessary that}}_{\text{operator, adverb}}\ \underbrace{\text{water is H}_2\text{O}}_{\text{sentence}}.$$

According to the Operator Analysis sentences with 'that' can be parsed into an operator (or adverbial phrase) and a sentence.

In English 'necessarily' is an adverb acting in this way.

In modal logic necessity is conceived as the sentential operator $\Box$: If φ is a sentence $\Box\varphi$ is also a sentence.

$$\underbrace{\text{That}}_{\text{fpoo}}\ \underbrace{\text{snow is white}}_{\text{sentence}}\ \underbrace{\text{is true}}_{\text{operator}}.$$

The predicate analysis

$$\underbrace{\text{It is necessary}}_{\text{predicate}}\ \underbrace{\text{that water is } H_2O}_{\text{singular term}}.$$

The predicate analysis

$$\underbrace{\text{It is necessary}}_{\text{predicate}}\ \underbrace{\text{that water is } H_2O}_{\text{singular term}}.$$

$$\underbrace{\text{That water is } H_2O}_{\text{singular term}}\ \underbrace{\text{is necessary}}_{\text{predicate}}.$$

The predicate analysis

$$\underbrace{\text{It is necessary}}_{\text{predicate}}\ \underbrace{\text{that water is H}_2\text{O}}_{\text{singular term}}.$$

$$\underbrace{\text{That water is H}_2\text{O}}_{\text{singular term}}\ \underbrace{\text{is necessary}}_{\text{predicate}}.$$

$$\underbrace{\text{The proposition that water is H}_2\text{O}}_{\text{singular term}}\ \underbrace{\text{is necessary}}_{\text{predicate}}.$$

Ontological commitment

According to the predicate analysis, ‘that’-sentences and similar constructions are singular terms denoting objects (propositions or sentences).

Ontological commitment

According to the predicate analysis, ‘that’-sentences and similar constructions are singular terms denoting objects (propositions or sentences).

Therefore the predicate analysis commits one to an ontology of objects that can be necessary, analytic, be known etc.

Advantages of the operator analysis

1. There is no ontological commitment to objects that can be necessary or analytic.

Advantages of the operator analysis

1. There is no ontological commitment to objects that can be necessary or analytic.
2. Paradoxes of modal predicates are avoided.

Advantages of the operator analysis

1. There is no ontological commitment to objects that can be necessary or analytic.
2. Paradoxes of modal predicates are avoided.
3. Paradoxes arising from the interaction of modal predicates are avoided.

Advantages of the operator analysis

1. There is no ontological commitment to objects that can be necessary or analytic.
2. Paradoxes of modal predicates are avoided.
3. Paradoxes arising from the interaction of modal predicates are avoided.
4. Modal logic and its possible worlds semantics can be retained as the main tool in the theory of modalities.

Advantages of the operator analysis

1. There is no ontological commitment to objects that can be necessary or analytic.
2. Paradoxes of modal predicates are avoided.
3. Paradoxes arising from the interaction of modal predicates are avoided.
4. Modal logic and its possible worlds semantics can be retained as the main tool in the theory of modalities.

The quantification problem

The advantages of the operator analysis
Quantified claims such as the following cannot be expressed without nasty ploys on the Operator Analysis:

- All laws of nature are necessary.

The quantification problem

The advantages of the operator analysis
Quantified claims such as the following cannot be expressed without nasty ploys on the Operator Analysis:

- All laws of nature are necessary.
- Some necessary propositions are not analytic.
- John has a true a posteriori belief.

The quantification problem

The advantages of the operator analysis
Quantified claims such as the following cannot be expressed without nasty ploys on the Operator Analysis:

- All laws of nature are necessary.
- Some necessary propositions are not analytic.
- John has a true a posteriori belief.

On the predicate account of modalities, in contrast, these sentences can readily be formalized

- $\forall x(Law(x) \rightarrow Nx)$
- $\exists x(N(x) \wedge \neg Ax)$
- $\exists x(Bvx \wedge Tx \wedge Apost(x))$

Quantified statements of this kind are of particular interest to the philosopher.

The predicate

The predicate and the operator analysis may be compatible in the end. Halbach and Welch (2009) proposed a reduction of predicates to operators using a truth predicate.

The predicate

The predicate and the operator analysis may be compatible in the end. Halbach and Welch (2009) proposed a reduction of predicates to operators using a truth predicate.

Here I will stick to the predicate analysis. It gives greater expressive power, but yields also paradoxes.

Paradoxes of modal predicates

Montague's (1963) Paradox

If analyticity is conceived as predicates of sentences, then the following assumptions are inconsistent on the basis of a weak syntax theory.

1. axioms: If '*A*' is analytic, then *A*.
2. rule of inference: If '*A*' has been proved, one may infer '"*A*" is analytic.'

Paradoxes of modal predicates

Montague's (1963) Paradox

If analyticity is conceived as predicates of sentences, then the following assumptions are inconsistent on the basis of a weak syntax theory.

1. axioms: If '*A*' is analytic, then *A*.
2. rule of inference: If '*A*' has been proved, one may infer ' "*A*" is analytic.'

The same paradox applies to necessity, a prioricity etc.

Paradoxes of modal predicates

Montague's (1963) Paradox

If analyticity is conceived as predicates of sentences, then the following assumptions are inconsistent on the basis of a weak syntax theory.

1. axioms: If 'A' is analytic, then A.
2. rule of inference: If 'A' has been proved, one may infer '"A" is analytic.'

The same paradox applies to necessity, a prioricity etc.

If analyticity is conceived as a predicate of propositions, the theory of syntax is replaced by a theory of propositions, which may prove the diagonal lemma.

Paradoxes from interactions of modal predicates

One might hope to solve Montague's paradox by typing the modal predicates; this move blocks Montague's paradox.

'Necessity–analyticity paradox'

The following axioms and rules are inconsistent with a basic theory of syntax:

- If '*A*' is analytic, then *A* (where *A* does not contain 'analytic')
- rule of inference: If '*A*' has been proved, one may infer ' "*A*" is analytic.' (where *A* does not contain 'analytic')
- If '*A*' is necessary, then *A* (where *A* does not contain 'necessary')
- rule of inference: If '*A*' has been proved, one may infer ' "*A*" is necessary.' (where *A* does not contain 'necessary')

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. This is not to say that truth cannot be defined in terms of correspondence etc.

SESSION 2
A Formal Theory of Syntax

Volker Halbach

Fudan

17th May 2012

There is something fishy about the liar paradox:

(1) (1) is not true.

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

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

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

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

This sentence is not true.

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

Arithmetic

The approach via arithmetic is indirect. Arithmetic talks about numbers, not about sentences. Coding sentences and expressions by numbers allows 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 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 (hopefully) 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.

Self-reference

Self-reference might look like a pathological phenomenon causing work for philosophical logicians only, which bears little interest elsewhere. Self-reference, however, has not only ‘detrimental’ applications as in the liar paradox, but also mathematical useful applications as in the proof of the Recursion Theorem (see, e.g., Rogers (1967)).

Self-reference

Self-reference might look like a pathological phenomenon causing work for philosophical logicians only, which bears little interest elsewhere. Self-reference, however, has not only 'detrimental' applications as in the liar paradox, but also mathematical useful applications as in the proof of the Recursion Theorem (see, e.g., Rogers (1967)).

Therefore most logicians agree nowadays that it is not self-reference that causes trouble in (1), but rather the truth predicate. The initial impression was wrong: not self-reference is to be blamed for the paradox but rather our concept of truth.

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

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. 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. 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. 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 $\mathrm{q}(t)$, $(s^\frown t)$, $\mathrm{sub}(r, s, t)$ are terms, and similarly for all further function symbols

The empty string

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

Formulæ

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

Example

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

The predicate symbol N can be read as necessity, truth, or still something else.

Non-linearity

Remark

The reason for overlining expressions $\overline{e}$ rather than to include them in quotation marks like this $\ulcorner e \urcorner$ is not a stylistic one. $\ulcorner \mathrm{v} \urcorner \ulcorner \mathrm{v} \urcorner$ may be parsed in the following two ways:

$$\underbrace{\overbrace{\ulcorner \mathrm{v} \urcorner}\,\overbrace{\ulcorner \mathrm{v} \urcorner}}$$

On the first reading indicated by the parentheses above, the expression if of the form $s^\frown t$; on the second reading it is an atomic term $\ulcorner \ldots \urcorner$. If quotations are marked by overlining the quoted strings, this ambiguity does not arise.

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

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

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 $\overline{a}^\frown\overline{b} = \overline{ab}$, where a and b are arbitrary strings of symbols.
- A3 $\mathrm{q}(\overline{a}) = \overline{\overline{a}}$
- A4 $\mathrm{sub}(\overline{a}, \overline{b}, \overline{c}) = \overline{d}$, where a and c are arbitrary strings of symbols, b is a single symbol (or, equivalently, a string of symbols of length 1), and d is the string of symbols obtained from a by replacing all occurrences of the symbol b by the strings c.

The axioms

Definition

A5 $\forall x \forall y \forall z ((x^\frown y)^\frown z) = (x^\frown(y^\frown z))$

The axioms

Definition

A5 $\forall x \forall y \forall z((x^\frown y)^\frown z) = (x^\frown(y^\frown z))$

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

The axioms

Definition

A5 $\forall x \forall y \forall z((x ^\frown y) ^\frown z) = (x ^\frown (y ^\frown z))$

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

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

The axioms

Definition

A5 $\forall x \forall y \forall z ((x^\frown y)^\frown z) = (x^\frown (y^\frown z))$

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

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

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

The axioms

Definition

A5 $\forall x \forall y \forall z((x^\frown y)^\frown z) = (x^\frown(y^\frown z))$

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

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

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

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 $\frown$ 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, $\neg\neg$v is the concatenation of $\neg$ and $\neg$v.

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.

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, ¬¬v is the concatenation of ¬ and ¬v.

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

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

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 13.

Comments

A1-A4 are already sufficient for proving the diagonalization Theorem 13.

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

A6–A8 will only be invoked in the section on arithmetic.

Comments

A6–A8 will only be invoked in the section on arithmetic.

A6 tells us that is either *a* or *b* is not the empty expression then the concatenation of *a* and *b* is non-empty as well.

Comments

A6–A8 will only be invoked in the section on arithmetic.

A6 tells us that is either a or b is not the empty expression then the concatenation of a and b is non-empty as well.

A7 postulates that only the empty string does not change an object if it is concatenated with this object.

Comments

In the following I shall need only the following instantiation of A8:

$$\forall x_1 \forall x_2 \forall z(\mathrm{sub}(x_1, \overline{\mathrm{v}}, z)^\frown \mathrm{sub}(x_2, \overline{\mathrm{v}}, z)) = \mathrm{sub}(x_1{}^\frown x_2, \overline{\mathrm{v}}, z)$$

I just thought that the universally quantified version looks less ad hoc.

Comments

In the following I shall need only the following instantiation of A8:

$$\forall x_1 \forall x_2 \forall z(\mathrm{sub}(x_1, \overline{\mathrm{v}}, z)^\frown \mathrm{sub}(x_2, \overline{\mathrm{v}}, z)) = \mathrm{sub}(x_1{}^\frown x_2, \overline{\mathrm{v}}, z)$$

I just thought that the universally quantified version looks less ad hoc.

The variable y is not restricted to single symbols, mostly. So A8 claims that if the problem of the uniqueness of the result of substituting complex expressions is solved, then the following holds: The result of concatenating the result of substituting c for b in a_1 with the result of substituting c for b in a_2 is the same as the result of substituting c for b in $a_1{}^\frown a_2$.

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

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

Comments

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

Example

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

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:

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

$\dashv$

SESSION 3

Axiomatic Approaches to Truth and Modality

Volker Halbach

Fudan

22nd May 2012

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

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

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 $N\overline{\psi} \leftrightarrow \psi$ for all sentences ψ of $\mathcal{L}$ is inconsistent.

The liar in $\mathcal{A}$

Theorem (liar paradox)

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

Proof.

Apply the diagonalization theorem 13 to the formula $\neg N\mathrm{v}$. Then theorem 13 implies the existence of a sentence γ such that the following holds: $\mathcal{A} \vdash \gamma \leftrightarrow \neg N\overline{\gamma}$. Together with the instance $N\overline{\gamma} \leftrightarrow \gamma$ of the T-scheme this yields an inconsistency. γ is called the 'liar sentence'. $\dashv$

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 13 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 5 do not allow for a definition of such truth predciate $\tau(\mathrm{v})$. According to Definition 5, 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 13 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 13 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 $\bot$ in the theory TB can be transformed into a proof of $\bot$ in $\mathcal{A}$. In the given proof only finitely many axioms with True can occur; let

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

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

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

The proof

The idea for the proof is due to Tarski.

I shall show that a given proof of a contradiction $\bot$ in the theory TB can be transformed into a proof of $\bot$ 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 $\bot$ in the theory TB can be transformed into a proof of $\bot$ 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 $\bot$.

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

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

(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.*

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 $\perp$ by φ in the proof above. $\dashv$

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. $\dashv$

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:

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

Tarski on TB

(Tarski, 1935, p. 256) proves the consistency result for TB, and one should expect that the resulting theory is attractive because it satisfies Convention T, but Tarski says:

Tarski on TB

(Tarski, 1935, p. 256) proves the consistency result for TB, and one should expect that the resulting theory is attractive because it satisfies Convention T, but Tarski says:

The value of the result obtained is considerably diminished by the fact that the axioms mentioned in Th. III have a very restricted deductive power. A theory of truth founded on them would be a highly incomplete system, which would lack the most important and most fruitful general theorems. Let us show this in more detail by a concrete example.

Consider the sentential function '$x \in Tr$ or $\overline{x} \in Tr$'. ["$\in Tr$" is the truth predicate, "$\overline{\in} Tr$" the negated truth predicate; "$\overline{x}$" designates the negation of "x".] If in this function we substitute for the variable 'x' structural-descriptive names of sentences, we obtain an infinite number of theorems, the proof of which on the basis of the axioms obtained from the convention ***T*** *presents not the slightest difficulty.*

Tarski on *TB*

But the situation changes fundamentally as soon as we pass to the generalization of this sentential function, i.e. to the general principle of contradiction. From the intuitive standpoint the truth of all those theorems is itself already a proof of the general principle; this principle represents, so to speak, an 'infinite logical product' of those special theorems. But this does not at all mean that we can actually derive the principle of contradiction from the axioms or theorems mentioned by means of the normal modes of inference usually employed. On the contrary, by a slight modification of Th. III it can be shown that the principle of contradiction is not a consequence (at least in the existing sense of the word) of the axiom system described.

Tarski on TB

It seems Tarski admits here that Convention T is highly incomplete: a good theory of truth should not only yield the T-sentences, it should also yield the general principle of contradiction ‘For any sentence (without True) either the sentence or its negation is true.’

Tarski on TB

It seems Tarski admits here that Convention T is highly incomplete: a good theory of truth should not only yield the T-sentences, it should also yield the general principle of contradiction 'For any sentence (without True) either the sentence or its negation is true.'

Perhaps Tarski thought that *definitions* of truth always yield the generalisations, if they satisfy Convention T. But TB trivially provides a definition of truth satisfying Convention T: one defines truth as True .

Tarski on TB

It seems Tarski admits here that Convention T is highly incomplete: a good theory of truth should not only yield the T-sentences, it should also yield the general principle of contradiction 'For any sentence (without True) either the sentence or its negation is true.'

Perhaps Tarski thought that *definitions* of truth always yield the generalisations, if they satisfy Convention T. But TB trivially provides a definition of truth satisfying Convention T: one defines truth as True .

Later serious examples were proved: BG defines truth for the language of ZF. The truth definition satisfies Convention T, but it does not yield the generalisation Tarski expects from a good definition of truth.

More notation

In order to prove the result that TB doesn't prove those generalisations, I need more axioms.

More notation

In order to prove the result that TB doesn't prove those generalisations, I need more axioms.

Sent(x) is a unary predicate, $\underset{\cdot}{\neg}$ a unary function symbol. I assume that Sent(x) represents the property of being a sentence of $\mathcal{L}$, $\underset{\cdot}{\neg}$ represents the function that takes a sentence and returns its negation:

Additional Axiom

$\mathcal{A} \vdash Sent(\overline{\varphi})$ iff φ is a sentence of $\mathcal{L}$.

Additional Axiom

$\mathcal{A} \vdash \underset{\cdot}{\neg}\overline{\varphi} = \overline{\neg\varphi}$

No generalisations

I can now formulate and proof Tarski's (and Gupta's) complaint:

Theorem

$TB \nvdash \forall x(Sent(x) \rightarrow (\text{True}\, x \vee \text{True}\, \dot{\neg} x))$ (assuming that $\mathcal{A}$ is consistent).

Proof

Assume otherwise. Then there is a proof of $\forall x(\mathrm{Sent}(x) \to (\mathrm{True}\, x \vee \mathrm{True}\, \dot{\neg} x))$ in from a finite subtheory S of TB. Only finitely many T-sentences can be in S. Let

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

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

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

As above, True v can be interpreted as $\tau(\mathrm{v})$.

Proof

Assume otherwise. Then there is a proof of $\forall x(\mathrm{Sent}(x) \rightarrow (\mathrm{True}\, x \vee \mathrm{True}\, \dot{\neg} x))$ in from a finite subtheory S of TB. Only finitely many T-sentences can be in S. Let

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

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

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

As above, True v can be interpreted as $\tau(\mathrm{v})$.

If χ is none of the $\psi_0, \ldots, \psi_n$, we have $\mathcal{A} \vdash \neg\tau(\overline{\chi}) \wedge \neg\tau(\dot{\neg}\overline{\chi})$.

Finite axiomatisability

Davidson made a big fuss about the following observation.

Theorem

TB is not finitely axiomatisiable over $\mathcal{A}$.

That is, there is no sentence σ such that $\mathcal{A} + \sigma$ proves the same sentences as TB.

Finite axiomatisability

Davidson made a big fuss about the following observation.

Theorem

TB is not finitely axiomatisiable over $\mathcal{A}$.

That is, there is no sentence σ such that $\mathcal{A} + \sigma$ proves the same sentences as TB.

Let a proof of σ in TB be given. Replace the truth predicate again in that proof by the partial truth predicate

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

This shows that the truth predicate of $\mathcal{A} + \sigma$ is definable in $\mathcal{A}$ already. But by assumption $\mathcal{A} + \sigma$ proves all T-sentences, and by Tarski's theorem this truth predicate cannot be $\mathcal{A}$-definable.

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

$$\forall x(\mathrm{Sent}(x) \to (\mathrm{True}\, x \lor \mathrm{True}\, \dot{\neg} x)) \text{ or}$$

$$\forall x \forall y\big(\mathrm{Sent}(x) \land \mathrm{Sent}(y) \to (\mathrm{True}\,(x \dot{\land} y) \leftrightarrow (\mathrm{True}\, x \land \mathrm{True}\, y))\big)$$

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

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

  $\forall x \forall y\big(\text{Sent}(x) \wedge \text{Sent}(y) \rightarrow (\text{True}\,(x \dot{\wedge} y) \leftrightarrow (\text{True}\, x \wedge \text{True}\, y))\big)$

- 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

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

  $\forall x \forall y\big(\text{Sent}(x) \wedge \text{Sent}(y) \rightarrow (\text{True}\, (x \dot{\wedge} y) \leftrightarrow (\text{True}\, x \wedge \text{True}\, y))\big)$

- 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.

Philosophical moral

The truth predicate of TB may have its merits: it allows one to axiomatise certain generalisations finitely Horwich (1998) Halbach (1999). But it doesn't prove the generalisations Tarski expected from a decent theory of truth.

Philosophical moral

The truth predicate of TB may have its merits: it allows one to axiomatise certain generalisations finitely Horwich (1998) Halbach (1999). But it doesn't prove the generalisations Tarski expected from a decent theory of truth.

Moreover, TB has been criticised, because the object-/metalanguage distinction seems to restrictive.

Weakening the T-sentences

Forget about the ‘new’ truth predicate True . I return to our old language $\mathcal{L}$ again, which contains already a truth predicate N.

Weakening the T-sentences

Forget about the 'new' truth predicate True . I return to our old language $\mathcal{L}$ again, which contains already a truth predicate N.

The next theorem is a strengthening of Theorem 14. For notions like necessity one will not postulate axioms of the form $N\overline{\psi} \leftrightarrow \psi$ but only that. However, one would expect a rule of necessitation to hold: once ψ has been derived, one may conclude $N\overline{\psi}$. With these weakened assumptions on N the liar paradox can still be derived.

Montague's paradox

Theorem (Montague's paradox (1963))

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

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

Montague's paradox

Theorem (Montague's paradox (1963))

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

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

Proof.

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

$\dashv$

Montague's paradox

Theorem (Montague's paradox (1963))

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

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

Proof.

$\gamma \leftrightarrow \neg N\overline{\gamma}$	diagonalization
$N\overline{\gamma} \leftrightarrow \neg\gamma$	
$N\overline{\gamma} \to \gamma$	axiom

$\dashv$

Montague's paradox

Theorem (Montague's paradox (1963))

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

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

Proof.

$\gamma \leftrightarrow \neg N\overline{\gamma}$	diagonalization
$N\overline{\gamma} \leftrightarrow \neg\gamma$	
$N\overline{\gamma} \rightarrow \gamma$	axiom
$\neg N\overline{\gamma}$	logic

⊣

Montague's paradox

Theorem (Montague's paradox (1963))

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

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

Proof.

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

$\dashv$

Montague's paradox

Theorem (Montague's paradox (1963))

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

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

Proof.

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

$\dashv$

Another formulation of the T-sentences

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

$$N\overline{\psi} \leftrightarrow \psi$$

where ψ must not contain N. 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 $N\overline{\varphi} \leftrightarrow \varphi$ *for all sentences* φ *of* $\mathcal{L}$ *not containing* N*.*
- 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

$\gamma \leftrightarrow \neg N\overline{N\overline{\gamma}}$	diagonalisation
$N\overline{N\overline{\gamma}} \leftrightarrow \neg\gamma$	
$N\overline{\gamma} \rightarrow \neg\gamma$	
$N\overline{\gamma} \rightarrow \gamma$	(N1)
$\neg N\overline{\gamma}$	two previous lines
$\neg N\overline{N\overline{\gamma}}$	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

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

$$\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (\mathrm{True}\,(x \dot{\wedge} y) \leftrightarrow (\mathrm{True}\, x \wedge \mathrm{True}\, y))\big)$$

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

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

$$\forall x \forall y\big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (\mathrm{True}\,(x \dot{\wedge} y) \leftrightarrow (\mathrm{True}\, x \wedge \mathrm{True}\, y))\big)$$

- 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

  $\forall x(\mathrm{Sent}(x) \rightarrow (\mathrm{True}\, x \vee \mathrm{True}\, \underset{\cdot}{\neg} x))$ or

  $\forall x \forall y\big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (\mathrm{True}\,(x \underset{\cdot}{\wedge} y) \leftrightarrow (\mathrm{True}\, x \wedge \mathrm{True}\, y))\big)$

- 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

$$\forall x(\mathrm{Sent}(x) \to (\mathrm{True}\, x \vee \mathrm{True}\, \underset{\cdot}{\neg} x)) \text{ or}$$

$$\forall x \forall y\big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \to (\mathrm{True}\, (x \underset{\cdot}{\wedge} y) \leftrightarrow (\mathrm{True}\, x \wedge \mathrm{True}\, y))\big)$$

- 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 $N\overline{\varphi} \leftrightarrow \varphi$ such that

- The set is consistent.

Maximal consistent instances of schema T

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

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

Maximal consistent instances of schema T

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

- The set is consistent.
- The set is maximal, ie no further sentences of the form $N\overline{\varphi} \leftrightarrow \varphi$ 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

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

Maximal consistent instances of schema T

Theorem (McGee)

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

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

Proof.

$$\begin{aligned} \mathcal{A} &\vdash \gamma \leftrightarrow (N\overline{\gamma} \leftrightarrow \varphi) && \text{diagonalisation} \\ \mathcal{A} &\vdash \varphi \leftrightarrow (N\overline{\gamma} \leftrightarrow \gamma) && \text{propositional logic} \end{aligned}$$

⊣

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.*

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

The liar again

Theorem

Assume $\mathcal{A} \vdash \varphi$ if and only if $\mathcal{A} \vdash N\overline{\varphi}$ 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)

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

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.

The Löb derivability conditions

Let K be the following scheme:

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

The Löb derivability conditions

Let K be the following scheme:

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

4 is the following scheme:

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

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

The Löb derivability conditions

Let K be the following scheme:

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

4 is the following scheme:

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

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.

Weaker reflection

One could try to escape Montague's paradox by postulating reflection only for certain sentences, eg:

Weaker reflection

Theorem

K4 *is inconsistent with the scheme* $N\overline{N\overline{\varphi}} \to \varphi$. *The same holds for* $N\overline{N\overline{N\overline{\varphi}}} \to \varphi$ *etc.*

Weaker reflection

Theorem

K4 *is inconsistent with the scheme* $N\overline{N\overline{\varphi}} \to \varphi$. *The same holds for* $N\overline{N\overline{N\overline{\varphi}}} \to \varphi$ *etc.*

Proof.

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

⊣

Internal inconsistency

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

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 N) if and only if $\mathcal{A} \vdash N\overline{\bot}$.

Theorem (Thomason 1980)

K4 *plus the scheme* $N\overline{N\overline{\varphi}} \to \varphi$ *is internally inconsistent.*

Internal inconsistency

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

Theorem (Thomason 1980)

K4 *plus the scheme* $N\overline{N\overline{\varphi}} \to \varphi$ *is internally inconsistent.*

Proof.

One runs the proof of Montague's theorem in the scope of N. ⊣

Löb's theorem

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

Theorem (Löb's theorem)

$K4 \vdash N\overline{N\overline{\varphi} \to \varphi} \to N\overline{\varphi}$

Löb's theorem

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

Theorem (Löb's theorem)

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

The corresponding rule follows as well:

Theorem

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

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

Theorem (Löb's theorem)

$K4 \vdash N\overline{N\overline{\varphi} \to \varphi} \to N\overline{\varphi}$

The corresponding rule follows as well:

Theorem

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

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

Löb's theorem: the proof

Proof.

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

$\dashv$

Gödel's second theorem

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

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* $\neg N\ulcorner\bot\urcorner$. *Thus* K4 $\nvdash \neg N\ulcorner\bot\urcorner$ *if* 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* $\neg N\overline{\bot}$. *Thus* K4 $\nvdash \neg N\overline{\bot}$ *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.

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* $\neg N\overline{\bot}$. *Thus* K4 $\nvdash \neg N\overline{\bot}$ *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 $\vdash N\overline{\bot} \vee \neg N\overline{\neg N\overline{\bot}}$.

*References

George Boolos. *The Logic of Provability*. Cambridge University Press, Cambridge, 1993.

Volker Halbach. Disquotationalism and infinite conjunctions. *Mind*, 108:1–22, 1999.

Volker Halbach and Philip Welch. Necessities and necessary truths: A prolegomenon to the metaphysics of modality. *Mind*, 118, 2009.

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

Paul Horwich. *Truth*. Basil Blackwell, Oxford, first edition, 1990.

Paul Horwich. *Truth*. Oxford University Press, Oxford, second edition edition, 1998. first edition 1990.

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

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

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