session 3
axioms for truth

Volker Halbach

Fudan

22nd May 2012

Last time I proved that the (formalisation of the) following sentence cannot be proved in TB:

> For all sentences (of the language without the new truth predicate): the sentence is not true iff the negation of the sentence is true.

But we cannot prove this from the Tarski-biconditionals: in any given argument we can use only finitely many of them, but the generalisation requires all of them (Tarski gave a formal proof and rejected TB because of its deductive weakness).

To get a stronger theory of truth we could just add all required generalizations as axioms to the syntax theory. The new theory should be stronger than TB.

Of course the new axioms should not only include

For all sentences (of the language without the new truth predicate): the sentence is not true iff the negation of the sentence is true.

but also corresponding axioms for other connectives and quantifiers.

To get a stronger theory of truth we could just add all required generalizations as axioms to the syntax theory. The new theory should be stronger than TB.

Of course the new axioms should not only include

> For all sentences (of the language without the new truth predicate): the sentence is not true iff the negation of the sentence is true.

but also corresponding axioms for other connectives and quantifiers.

For instance, we would like to have. Sentences are always understood in the sense of sentences of the original language without the new truth predicate:

1. An identity sentence $s = t$ is true iff s and t denote the same thing (where s and t are closed terms) of the language; and so on for further predicates of $\mathcal{L}$
2. A negation of a sentence is true iff the sentence is not true.
3. A conjunction is true iff both conjuncts are true.
4. A sentence 'Everything is A' is true iff the result of applying A to any name is true.

The last clause uses a substitutional interpretation of quantification, which shouldn't matter if names for all objects are available.

All this can be written down more formally.

For instance, we would like to have. Sentences are always understood in the sense of sentences of the original language without the new truth predicate:

1. An identity sentence $s = t$ is true iff s and t denote the same thing (where s and t are closed terms) of the language; and so on for further predicates of $\mathcal{L}$
2. A negation of a sentence is true iff the sentence is not true.
3. A conjunction is true iff both conjuncts are true.
4. A sentence 'Everything is A' is true iff the result of applying A to any name is true.

The last clause uses a substitutional interpretation of quantification, which shouldn't matter if names for all objects are available.

All this can be written down more formally.

For instance, we would like to have. Sentences are always understood in the sense of sentences of the original language without the new truth predicate:

1. An identity sentence $s = t$ is true iff s and t denote the same thing (where s and t are closed terms) of the language; and so on for further predicates of $\mathcal{L}$
2. A negation of a sentence is true iff the sentence is not true.
3. A conjunction is true iff both conjuncts are true.
4. A sentence 'Everything is A' is true iff the result of applying A to any name is true.

The last clause uses a substitutional interpretation of quantification, which shouldn't matter if names for all objects are available.

All this can be written down more formally.

For instance, we would like to have. Sentences are always understood in the sense of sentences of the original language without the new truth predicate:

1. An identity sentence $s = t$ is true iff s and t denote the same thing (where s and t are closed terms) of the language; and so on for further predicates of $\mathcal{L}$
2. A negation of a sentence is true iff the sentence is not true.
3. A conjunction is true iff both conjuncts are true.
4. A sentence 'Everything is A' is true iff the result of applying A to any name is true.

The last clause uses a substitutional interpretation of quantification, which shouldn't matter if names for all objects are available.

All this can be written down more formally.

For instance, we would like to have. Sentences are always understood in the sense of sentences of the original language without the new truth predicate:

1. An identity sentence $s = t$ is true iff s and t denote the same thing (where s and t are closed terms) of the language; and so on for further predicates of $\mathcal{L}$
2. A negation of a sentence is true iff the sentence is not true.
3. A conjunction is true iff both conjuncts are true.
4. A sentence 'Everything is A' is true iff the result of applying A to any name is true.

The last clause uses a substitutional interpretation of quantification, which shouldn't matter if names for all objects are available.

All this can be written down more formally.

If we assume that truth is a new (previously unused) predicate, we add the (formalisations of the) following axioms as axioms for truth:

Definition

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).

In the formal language we have $\neg$ and $\rightarrow$ as connectives and $\forall$ as quantifier symbol.

If we assume that truth is a new (previously unused) predicate, we add the (formalisations of the) following axioms as axioms for truth:

Definition

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).

In the formal language we have $\neg$ and $\rightarrow$ as connectives and $\forall$ as quantifier symbol.

If we assume that truth is a new (previously unused) predicate, we add the (formalisations of the) following axioms as axioms for truth:

Definition

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 $\mathcal{L}$)
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).

In the formal language we have $\neg$ and $\to$ as connectives and $\forall$ as quantifier symbol.

If we assume that truth is a new (previously unused) predicate, we add the (formalisations of the) following axioms as axioms for truth:

Definition

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).

In the formal language we have $\neg$ and $\rightarrow$ as connectives and $\forall$ as quantifier symbol.

Adding these axioms to $\mathcal{A}$ yields a theory often known at CT$\upharpoonright$ ('compositional truth'). Here I'll call it $\mathcal{D}$.

I'll now turn these axioms into a formal theory. You can actually skip the pages up to the definition of $\mathcal{D}$.

If you prefer to skip the bad formal stuff click here. I'll probably do the same in the lecture, but I thought I include it for the sake of those who want to see the details.

Before formalising the definition of truth for $\mathcal{L}$-sentences I need more expressive power in $\mathcal{A}$. This is actually the hard part. I apologize…

I'll now turn these axioms into a formal theory. You can actually skip the pages up to the definition of $\mathcal{D}$.

If you prefer to skip the bad formal stuff click here. I'll probably do the same in the lecture, but I thought I include it for the sake of those who want to see the details.

Before formalising the definition of truth for $\mathcal{L}$-sentences I need more expressive power in $\mathcal{A}$. This is actually the hard part. I apologize…

I'll now turn these axioms into a formal theory. You can actually skip the pages up to the definition of $\mathcal{D}$.

If you prefer to skip the bad formal stuff click here. I'll probably do the same in the lecture, but I thought I include it for the sake of those who want to see the details.

Before formalising the definition of truth for $\mathcal{L}$-sentences I need more expressive power in $\mathcal{A}$. This is actually the hard part. I apologize…

Quantifying into quotational context is notoriously problematic. Occurrences of variables within quotation marks cannot be bound from ‘outside’. For instance, the quantifier $\forall$v is not binding into the overlined expression: $\forall v \, \Box \, \overline{v = v}$. In some cases, however, it is possible to bind quoted variables in a sense to be explained.

Assume $\Box$ is read as 'necessary' and we want to say that every expression is necessarily identical with itself, we cannot do this by $\forall \mathrm{v} \, \Box \, \overline{\mathrm{v} = \mathrm{v}}$, 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 necessary.*

This can be formalized by the following expression:

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

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

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

Values

I assume from now on that the language of $\mathcal{A}$ contains also a unary function symbol val and that $\mathcal{A}$ proves the following equations:

Additional Axiom

$\mathrm{val}(t) = e$ *if and only if t denotes a term denoting the expression denoted by e in the standard model.*

val represents the function that gives applied to a term of the language it's value, ie, he object denoted by that term.

Example

- $\mathcal{A} \vdash \mathrm{val}\overline{\overline{\neg}} = \overline{\neg}$ translated into the metalanguage: 'the value of '$\overline{\neg}$' is '$\neg$''.
- $\mathcal{A} \vdash \mathrm{val}\ \mathrm{q}\overline{\forall\neg} = \overline{\forall\neg}$, so val disquotes terms (but not sentences).
- $\mathcal{A} \vdash \mathrm{val}\ \mathrm{q}(\overline{\neg}{}^\frown\overline{\forall\forall}) = \overline{\neg\forall\forall}$

Values

I assume from now on that the language of $\mathcal{A}$ contains also a unary function symbol val and that $\mathcal{A}$ proves the following equations:

Additional Axiom

$\mathrm{val}(t) = e$ *if and only if t denotes a term denoting the expression denoted by e in the standard model.*

val represents the function that gives applied to a term of the language it's value, ie, he object denoted by that term.

Example

- $\mathcal{A} \vdash \mathrm{val}\overline{\overline{\neg}} = \overline{\neg}$ translated into the metalanguage: 'the value of '$\overline{\neg}$' is '$\neg$''.
- $\mathcal{A} \vdash \mathrm{val}\ q\overline{\forall\neg} = \overline{\forall\neg}$, so val disquotes terms (but not sentences).
- $\mathcal{A} \vdash \mathrm{val}\ q(\overline{\overline{\neg}}{}^\frown\overline{\forall\forall}) = \overline{\neg\forall\forall}$

Values

I assume from now on that the language of $\mathcal{A}$ contains also a unary function symbol val and that $\mathcal{A}$ proves the following equations:

Additional Axiom

$\mathrm{val}(t) = e$ *if and only if t denotes a term denoting the expression denoted by e in the standard model.*

val represents the function that gives applied to a term of the language it's value, ie, he object denoted by that term.

Example

- $\mathcal{A} \vdash \mathrm{val}\overline{\overline{\neg}} = \overline{\neg}$ translated into the metalanguage: 'the value of '$\overline{\neg}$' is '$\neg$''.
- $\mathcal{A} \vdash \mathrm{val}\ \mathrm{q}\overline{\forall\neg} = \overline{\forall\neg}$, so val disquotes terms (but not sentences).
- $\mathcal{A} \vdash \mathrm{val}\ \mathrm{q}(\overline{\neg}{}^\frown\overline{\forall\forall}) = \overline{\neg\forall\forall}$

Values

I assume from now on that the language of $\mathcal{A}$ contains also a unary function symbol val and that $\mathcal{A}$ proves the following equations:

Additional Axiom

$\mathrm{val}(t) = e$ *if and only if t denotes a term denoting the expression denoted by e in the standard model.*

val represents the function that gives applied to a term of the language it's value, ie, he object denoted by that term.

Example

- $\mathcal{A} \vdash \mathrm{val}\overline{\overline{\neg}} = \overline{\neg}$ translated into the metalanguage: 'the value of '$\overline{\neg}$' is '$\neg$''.
- $\mathcal{A} \vdash \mathrm{val}\ \mathrm{q}\overline{\forall\neg} = \overline{\forall\neg}$, so val disquotes terms (but not sentences).
- $\mathcal{A} \vdash \mathrm{val}\ \mathrm{q}(\overline{\neg}{}^\frown\overline{\forall\forall}) = \overline{\neg\forall\forall}$

More predicates

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

Additional Axiom

$\mathcal{A} \vdash Sent(t)$ if and only if t is a term denoting a sentence in the standard model.

Example

- $\mathcal{A} \vdash \mathrm{Sent}(\overline{\forall v\, v = v})$
- $\mathcal{A} \vdash \mathrm{Sent}(\mathrm{val}\overline{\overline{\forall v\, v = v}})$

More predicates

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

Additional Axiom

$\mathcal{A} \vdash Sent(t)$ if and only if t is a term denoting a sentence in the standard model.

Example

- $\mathcal{A} \vdash \mathrm{Sent}(\overline{\forall v\, v = v})$
- $\mathcal{A} \vdash \mathrm{Sent}(\mathrm{val}\overline{\overline{\forall v\, v = v}})$

More predicates

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

Additional Axiom

$\mathcal{A} \vdash Sent(t)$ if and only if t is a term denoting a sentence in the standard model.

Example

- $\mathcal{A} \vdash \mathrm{Sent}(\overline{\forall \mathrm{v}\, \mathrm{v} = \mathrm{v}})$
- $\mathcal{A} \vdash \mathrm{Sent}(\mathrm{val}\overline{\overline{\forall \mathrm{v}\, \mathrm{v} = \mathrm{v}}})$

More predicates

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

Additional Axiom

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

Example

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

More predicates

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

Additional Axiom

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

Example

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

More predicates

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

Additional Axiom

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

Example

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

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 …'.

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

Definition

- $x \underset{\cdot}{=} y \stackrel{\text{def}}{=} x^\frown \overline{=}^\frown y$
- $\underset{\cdot}{\neg} x \stackrel{\text{def}}{=} \overline{\neg}^\frown x$
- $x \underset{\cdot}{\rightarrow} y \stackrel{\text{def}}{=} \overline{(}^\frown x^\frown \overline{\rightarrow}^\frown y^\frown \overline{)}$
- $\underset{\cdot}{\forall} x y \stackrel{\text{def}}{=} \overline{\forall}^\frown x^\frown y$

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

- $x \underset{\cdot}{=} y \overset{\text{def}}{=} x^\frown \overline{=}^\frown y$
- $\underset{\cdot}{\neg} x \overset{\text{def}}{=} \overline{\neg}^\frown x$
- $x \underset{\cdot}{\rightarrow} y \overset{\text{def}}{=} \overline{(}^\frown x^\frown \overline{\rightarrow}^\frown y^\frown \overline{)}$
- $\underset{\cdot}{\forall} x y \overset{\text{def}}{=} \overline{\forall}^\frown x^\frown y$

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

- $x \underset{\cdot}{=} y \stackrel{\text{def}}{=} x \frown \equiv \frown y$
- $\underset{\cdot}{\neg} x \stackrel{\text{def}}{=} \overline{\neg} \frown x$
- $x \underset{\cdot}{\to} y \stackrel{\text{def}}{=} \overline{(} \frown x \frown \overline{\to} \frown y \frown \overline{)}$
- $\underset{\cdot}{\forall} x y \stackrel{\text{def}}{=} \overline{\forall} \frown x \frown y$

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

- $x \underset{\cdot}{=} y \overset{\text{def}}{=} x \frown \overline{=} \frown y$
- $\underset{\cdot}{\neg} x \overset{\text{def}}{=} \overline{\neg} \frown x$
- $x \underset{\cdot}{\rightarrow} y \overset{\text{def}}{=} \overline{(} \frown x \frown \overline{\rightarrow} \frown y \frown \overline{)}$
- $\underset{\cdot}{\forall} x y \overset{\text{def}}{=} \overline{\forall} \frown x \frown y$

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

- $x \mathbin{\underset{\cdot}{=}} y \stackrel{\text{def}}{=} x ^\frown \overline{=} ^\frown y$
- $\underset{\cdot}{\neg} x \stackrel{\text{def}}{=} \overline{\neg} ^\frown x$
- $x \mathbin{\underset{\cdot}{\rightarrow}} y \stackrel{\text{def}}{=} \overline{(} ^\frown x ^\frown \overline{\rightarrow} ^\frown y ^\frown \overline{)}$
- $\underset{\cdot}{\forall} x y \stackrel{\text{def}}{=} \overline{\forall} ^\frown x ^\frown y$

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

- $x \underset{\cdot}{=} y \overset{\text{def}}{=} x^\frown \overline{=}^\frown y$
- $\underset{\cdot}{\neg} x \overset{\text{def}}{=} \overline{\neg}^\frown x$
- $x \underset{\cdot}{\to} y \overset{\text{def}}{=} \overline{(}^\frown x^\frown \overline{\to}^\frown y^\frown \overline{)}$
- $\underset{\cdot}{\forall} x y \overset{\text{def}}{=} \overline{\forall}^\frown x^\frown y$

Definition

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

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T(x \underset{\cdot}{=} y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y))\big)$
   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. $\forall x \big(\mathrm{ClT}(x) \rightarrow (T(\underset{\cdot}{\mathrm{Sent}}(x)) \leftrightarrow \mathrm{Sent}(\mathrm{val}(x)))\big)$
   and so on for further predicates of $\mathcal{L}$...
3. $\forall x \big(\mathrm{Sent}(x) \rightarrow (T \underset{\cdot}{\neg} x \leftrightarrow \neg T x)\big)$
   A negated $\mathcal{L}$-sentence $\neg\varphi$ is true iff φ is not true.
4. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (T(x \underset{\cdot}{\rightarrow} y) \leftrightarrow (Tx \rightarrow Ty)\big)$
   A conditional $\varphi \rightarrow \psi$ is true iff φ is false or and ψ is true (φ and ψ are sentences of $\mathcal{L}$)
5. $\forall x \forall y \big(\mathrm{Sent}(\underset{\cdot}{\forall} x y) \rightarrow (T(\underset{\cdot}{\forall} x y) \leftrightarrow \forall z T \mathrm{sub}(y, x, \mathrm{q}z))\big)$
   A universally quantified $\mathcal{L}$-sentence $\forall x \varphi(x)$ is true iff $\varphi(\overline{e})$ for all objects e.

Definition

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

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T(x \underset{\cdot}{=} y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y))\big)$
   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. $\forall x \big(\mathrm{ClT}(x) \rightarrow (T(\underset{\cdot}{\mathrm{Sent}}(x)) \leftrightarrow \mathrm{Sent}(\mathrm{val}(x)))\big)$
   and so on for further predicates of $\mathcal{L}$...
3. $\forall x \big(\mathrm{Sent}(x) \rightarrow (T \underset{\cdot}{\neg} x \leftrightarrow \neg T x)\big)$
   A negated $\mathcal{L}$-sentence $\neg\varphi$ is true iff φ is not true.
4. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (T(x \underset{\cdot}{\rightarrow} y) \leftrightarrow (Tx \rightarrow Ty)\big)$
   A conditional $\varphi \rightarrow \psi$ is true iff φ is false or and ψ is true (φ and ψ are sentences of $\mathcal{L}$)
5. $\forall x \forall y \big(\mathrm{Sent}(\underset{\cdot}{\forall} x y) \rightarrow (T(\underset{\cdot}{\forall} x y) \leftrightarrow \forall z T \mathrm{sub}(y, x, \mathrm{q}z))\big)$
   A universally quantified $\mathcal{L}$-sentence $\forall x \varphi(x)$ is true iff $\varphi(\overline{e})$ for all objects e.

Definition

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

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T(x \dot{=} y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y))\big)$
   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. $\forall x \big(\mathrm{ClT}(x) \rightarrow (T(\dot{\mathrm{Sent}}(x)) \leftrightarrow \mathrm{Sent}(\mathrm{val}(x)))\big)$
   and so on for further predicates of $\mathcal{L}$...
3. $\forall x \big(\mathrm{Sent}(x) \rightarrow (T \dot{\neg} x \leftrightarrow \neg T x)\big)$
   A negated $\mathcal{L}$-sentence $\neg\varphi$ is true iff φ is not true.
4. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (T(x \dot{\rightarrow} y) \leftrightarrow (Tx \rightarrow Ty)\big)$
   A conditional $\varphi \rightarrow \psi$ is true iff φ is false or and ψ is true (φ and ψ are sentences of $\mathcal{L}$)
5. $\forall x \forall y \big(\mathrm{Sent}(\dot{\forall} xy) \rightarrow (T(\dot{\forall} xy) \leftrightarrow \forall z T \mathrm{sub}(y, x, \mathrm{q}z))\big)$
   A universally quantified $\mathcal{L}$-sentence $\forall x \varphi(x)$ is true iff $\varphi(\overline{e})$ for all objects e.

Definition

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

1. $\forall x \forall y\big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T(x \underset{\cdot}{=} y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y))\big)$
   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. $\forall x\big(\mathrm{ClT}(x) \rightarrow (T(\underset{\cdot}{\mathrm{Sent}}(x)) \leftrightarrow \mathrm{Sent}(\mathrm{val}(x)))\big)$
   and so on for further predicates of $\mathcal{L}$…
3. $\forall x\big(\mathrm{Sent}(x) \rightarrow (T \underset{\cdot}{\neg} x \leftrightarrow \neg T x)\big)$
   A negated $\mathcal{L}$-sentence $\neg\varphi$ is true iff φ is not true.
4. $\forall x \forall y\big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (T(x \underset{\cdot}{\rightarrow} y) \leftrightarrow (Tx \rightarrow Ty))\big)$
   A conditional $\varphi \rightarrow \psi$ is true iff φ is false or and ψ is true (φ and ψ are sentences of $\mathcal{L}$)
5. $\forall x \forall y\big(\mathrm{Sent}(\underset{\cdot}{\forall} x y) \rightarrow (T(\underset{\cdot}{\forall} x y) \leftrightarrow \forall z T \mathrm{sub}(y, x, \mathrm{q}z))\big)$
   A universally quantified $\mathcal{L}$-sentence $\forall x \varphi(x)$ is true iff $\varphi(\overline{e})$ for all objects e.

Definition

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

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T(x \dot{=} y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y))\big)$
   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. $\forall x \big(\mathrm{ClT}(x) \rightarrow (T(\dot{\mathrm{Sent}}(x)) \leftrightarrow \mathrm{Sent}(\mathrm{val}(x)))\big)$
   and so on for further predicates of $\mathcal{L}$...
3. $\forall x \big(\mathrm{Sent}(x) \rightarrow (T \dot{\neg} x \leftrightarrow \neg T x)\big)$
   A negated $\mathcal{L}$-sentence $\neg\varphi$ is true iff φ is not true.
4. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (T(x \dot{\rightarrow} y) \leftrightarrow (Tx \rightarrow Ty))\big)$
   A conditional $\varphi \rightarrow \psi$ is true iff φ is false or and ψ is true (φ and ψ are sentences of $\mathcal{L}$)
5. $\forall x \forall y \big(\mathrm{Sent}(\dot{\forall} x y) \rightarrow (T(\dot{\forall} x y) \leftrightarrow \forall z T \mathrm{sub}(y, x, \mathrm{q}z))\big)$
   A universally quantified $\mathcal{L}$-sentence $\forall x \varphi(x)$ is true iff $\varphi(\overline{e})$ for all objects e.

Definition

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

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T(x \dot{=} y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y))\big)$
   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. $\forall x \big(\mathrm{ClT}(x) \rightarrow (T(\dot{\mathrm{Sent}}(x)) \leftrightarrow \mathrm{Sent}(\mathrm{val}(x)))\big)$
   and so on for further predicates of $\mathcal{L}$…
3. $\forall x \big(\mathrm{Sent}(x) \rightarrow (T \dot{\neg} x \leftrightarrow \neg T x)\big)$
   A negated $\mathcal{L}$-sentence $\neg\varphi$ is true iff φ is not true.
4. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (T(x \dot{\rightarrow} y) \leftrightarrow (Tx \rightarrow Ty))\big)$
   A conditional $\varphi \rightarrow \psi$ is true iff φ is false or and ψ is true (φ and ψ are sentences of $\mathcal{L}$)
5. $\forall x \forall y \big(\mathrm{Sent}(\dot{\forall} x y) \rightarrow (T(\dot{\forall} x y) \leftrightarrow \forall z T \mathrm{sub}(y, x, \mathrm{q}z))\big)$
   A universally quantified $\mathcal{L}$-sentence $\forall x \varphi(x)$ is true iff $\varphi(\overline{e})$ for all objects e.

Definition

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

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T(x \dot{=} y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y))\big)$
   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. $\forall x \big(\mathrm{ClT}(x) \rightarrow (T(\dot{\mathrm{Sent}}(x)) \leftrightarrow \mathrm{Sent}(\mathrm{val}(x)))\big)$
   and so on for further predicates of $\mathcal{L}$…
3. $\forall x \big(\mathrm{Sent}(x) \rightarrow (T \dot{\neg} x \leftrightarrow \neg T x)\big)$
   A negated $\mathcal{L}$-sentence $\neg\varphi$ is true iff φ is not true.
4. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (T(x \dot{\rightarrow} y) \leftrightarrow (Tx \rightarrow Ty)\big)$
   A conditional $\varphi \rightarrow \psi$ is true iff φ is false or and ψ is true (φ and ψ are sentences of $\mathcal{L}$)
5. $\forall x \forall y \big(\mathrm{Sent}(\dot{\forall} x y) \rightarrow (T(\dot{\forall} x y) \leftrightarrow \forall z T \mathrm{sub}(y, x, \mathrm{q}z))\big)$
   A universally quantified $\mathcal{L}$-sentence $\forall x \varphi(x)$ is true iff $\varphi(\overline{e})$ for all objects e.

Definition

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

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \to (T(x \dot{=} y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y))\big)$
   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. $\forall x \big(\mathrm{ClT}(x) \to (T(\dot{\mathrm{Sent}}(x)) \leftrightarrow \mathrm{Sent}(\mathrm{val}(x)))\big)$
   and so on for further predicates of $\mathcal{L}$...
3. $\forall x \big(\mathrm{Sent}(x) \to (T \dot{\neg} x \leftrightarrow \neg T x)\big)$
   A negated $\mathcal{L}$-sentence $\neg\varphi$ is true iff φ is not true.
4. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \to (T(x \dot{\to} y) \leftrightarrow (T x \to T y)\big)$
   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. $\forall x \forall y \big(\mathrm{Sent}(\dot{\forall} x y) \to (T(\dot{\forall} x y) \leftrightarrow \forall z T \mathrm{sub}(y, x, \mathrm{q}z))\big)$
   A universally quantified $\mathcal{L}$-sentence $\forall x \varphi(x)$ is true iff $\varphi(\overline{e})$ for all objects e.

Definition

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

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T(x \dot{=} y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y))\big)$
   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. $\forall x \big(\mathrm{ClT}(x) \rightarrow (T(\dot{\mathrm{Sent}}(x)) \leftrightarrow \mathrm{Sent}(\mathrm{val}(x)))\big)$
   and so on for further predicates of $\mathcal{L}$...
3. $\forall x \big(\mathrm{Sent}(x) \rightarrow (T \dot{\neg} x \leftrightarrow \neg T x)\big)$
   A negated $\mathcal{L}$-sentence $\neg\varphi$ is true iff φ is not true.
4. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (T(x \dot{\rightarrow} y) \leftrightarrow (Tx \rightarrow Ty)\big)$
   A conditional $\varphi \rightarrow \psi$ is true iff φ is false or and ψ is true (φ and ψ are sentences of $\mathcal{L}$)
5. $\forall x \forall y \big(\mathrm{Sent}(\dot{\forall} x y) \rightarrow (T(\dot{\forall} x y) \leftrightarrow \forall z T \mathrm{sub}(y, x, \mathrm{q}z))\big)$
   A universally quantified $\mathcal{L}$-sentence $\forall x \varphi(x)$ is true iff $\varphi(\overline{e})$ for all objects e.

Definition

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

1. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T(x \underset{\cdot}{=} y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y))\big)$
   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. $\forall x \big(\mathrm{ClT}(x) \rightarrow (T(\underset{\cdot}{\mathrm{Sent}}(x)) \leftrightarrow \mathrm{Sent}(\mathrm{val}(x)))\big)$
   and so on for further predicates of $\mathcal{L}$…
3. $\forall x \big(\mathrm{Sent}(x) \rightarrow (T \underset{\cdot}{\neg} x \leftrightarrow \neg T x)\big)$
   A negated $\mathcal{L}$-sentence $\neg\varphi$ is true iff φ is not true.
4. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (T(x \underset{\cdot}{\rightarrow} y) \leftrightarrow (Tx \rightarrow Ty))\big)$
   A conditional $\varphi \rightarrow \psi$ is true iff φ is false or and ψ is true (φ and ψ are sentences of $\mathcal{L}$)
5. $\forall x \forall y \big(\mathrm{Sent}(\underset{\cdot}{\forall} x y) \rightarrow (T(\underset{\cdot}{\forall} x y) \leftrightarrow \forall z T \mathrm{sub}(y, x, \mathrm{q}z))\big)$
   A universally quantified $\mathcal{L}$-sentence $\forall x \varphi(x)$ is true iff $\varphi(\overline{e})$ for all objects e.

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. $\forall x \forall y \big(\mathrm{ClT}(x) \wedge \mathrm{ClT}(y) \rightarrow (T(x \underset{\cdot}{=} y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y))\big)$
2. $\forall x \big(\mathrm{ClT}(x) \rightarrow (T(\underset{\cdot}{\mathrm{Sent}}(x)) \leftrightarrow \mathrm{Sent}(\mathrm{val}(x)))\big)$

   $\ldots$
3. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (T(x \underset{\cdot}{\rightarrow} y) \leftrightarrow (Tx \rightarrow Ty)\big)$
4. $\forall x \forall y \big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (T(x \underset{\cdot}{\rightarrow} y) \leftrightarrow (Tx \rightarrow Ty))\big)$
5. $\forall x \forall y \big(\mathrm{Sent}(\underset{\cdot}{\forall} xy) \rightarrow (T(\underset{\cdot}{\forall} xy) \leftrightarrow \forall z T \mathrm{sub}(y, x, \mathrm{q}z))\big)$

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

- T is not a symbol of $\mathcal{L}$: any axiom schemata of $\mathcal{A}$ contains only substitution instances from $\mathcal{L}$, that is, without T.
- 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 $\forall x(\mathrm{Sent}(x) \rightarrow (Tx \vee T\dot{\neg}x))$

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

- T is not a symbol of $\mathcal{L}$: any axiom schemata of $\mathcal{A}$ contains only substitution instances from $\mathcal{L}$, that is, without T.
- 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 $\forall x(\mathrm{Sent}(x) \rightarrow (Tx \vee T\underset{\cdot}{\neg}x))$

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

- T is not a symbol of $\mathcal{L}$: any axiom schemata of $\mathcal{A}$ contains only substitution instances from $\mathcal{L}$, that is, without T.
- 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 $\forall x(\mathrm{Sent}(x) \to (Tx \lor T\dot{\neg}x))$

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

- T is not a symbol of $\mathcal{L}$: any axiom schemata of $\mathcal{A}$ contains only substitution instances from $\mathcal{L}$, that is, without T.
- 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 $\forall x(\mathrm{Sent}(x) \rightarrow (Tx \vee T\dot{\neg}x))$

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

- T is not a symbol of $\mathcal{L}$: any axiom schemata of $\mathcal{A}$ contains only substitution instances from $\mathcal{L}$, that is, without T.
- 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 $\forall x(\mathrm{Sent}(x) \to (Tx \vee T\dot{\neg}x))$

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

- T is not a symbol of $\mathcal{L}$: any axiom schemata of $\mathcal{A}$ contains only substitution instances from $\mathcal{L}$, that is, without T.
- 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 $\forall x(\mathrm{Sent}(x) \rightarrow (Tx \vee T\underset{\cdot}{\neg} x))$

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 T) (cf Shapiro (1998), Field (1999), Halbach (1999), Ketland (1999)).

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

> *If is provable in $\mathcal{D}$, then φ is already provable in $\mathcal{A}$.*

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 T) (cf Shapiro (1998), Field (1999), Halbach (1999), Ketland (1999)).

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

If is provable in $\mathcal{D}$, then φ is already provable in $\mathcal{A}$.

So we would like to show that adding the following sentences as axioms doesn't allow us to prove more sentences in the language $\mathcal{L}$ without the truth predicate.

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 $\mathcal{L}$)
4. A universally quantified $\mathcal{L}$-sentence $\forall x\varphi(x)$ is true iff $\varphi(\overline{e})$ for all objects e.

So we would like to show that adding the following sentences as axioms doesn't allow us to prove more sentences in the language $\mathcal{L}$ without the truth predicate.

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 $\mathcal{L}$)
4. A universally quantified $\mathcal{L}$-sentence $\forall x\varphi(x)$ is true iff $\varphi(\overline{e})$ for all objects e.

So we would like to show that adding the following sentences as axioms doesn't allow us to prove more sentences in the language $\mathcal{L}$ without the truth predicate.

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 $\mathcal{L}$)
4. A universally quantified $\mathcal{L}$-sentence $\forall x \varphi(x)$ is true iff $\varphi(\overline{e})$ for all objects e.

So we would like to show that adding the following sentences as axioms doesn't allow us to prove more sentences in the language $\mathcal{L}$ without the truth predicate.

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 $\mathcal{L}$)
4. A universally quantified $\mathcal{L}$-sentence $\forall x\varphi(x)$ is true iff $\varphi(\overline{e})$ for all objects e.

Surprisingly this is hard. There are incorrect proofs by famous and less famous logicians for this claim (I'm one of the latter).

In fact, conservativeness follows from a result proved by Kotlarski et al. (1981). Graham Leigh, Ali Enyat and Albert Visser might have proofs as well.

Thus we still cannot prove any new truth-free sentences.

The situation changes if we apply additional syntactic axioms that allow us to prove the all theorems of $\mathcal{D}$ are true. In this case conservativeness is lost.

Surprisingly this is hard. There are incorrect proofs by famous and less famous logicians for this claim (I'm one of the latter).

In fact, conservativeness follows from a result proved by Kotlarski et al. (1981). Graham Leigh, Ali Enyat and Albert Visser might have proofs as well.

Thus we still cannot prove any new truth-free sentences.

The situation changes if we apply additional syntactic axioms that allow us to prove the all theorems of $\mathcal{D}$ are true. In this case conservativeness is lost.

Surprisingly this is hard. There are incorrect proofs by famous and less famous logicians for this claim (I'm one of the latter).

In fact, conservativeness follows from a result proved by Kotlarski et al. (1981). Graham Leigh, Ali Enyat and Albert Visser might have proofs as well.

Thus we still cannot prove any new truth-free sentences.

The situation changes if we apply additional syntactic axioms that allow us to prove the all theorems of $\mathcal{D}$ are true. In this case conservativeness is lost.

Surprisingly this is hard. There are incorrect proofs by famous and less famous logicians for this claim (I'm one of the latter).

In fact, conservativeness follows from a result proved by Kotlarski et al. (1981). Graham Leigh, Ali Enyat and Albert Visser might have proofs as well.

Thus we still cannot prove any new truth-free sentences.

The situation changes if we apply additional syntactic axioms that allow us to prove the all theorems of $\mathcal{D}$ are true. In this case conservativeness is lost.

In the theory $\mathcal{D}$ we have only finitely many axioms for truth. This is in contrast to TB, which has infinitely many axioms for truth.

This was the main reason for Davidson to prefer $\mathcal{D}$ over TB and to emphasize the importance of compositionality.

In the theory $\mathcal{D}$ we have only finitely many axioms for truth. This is in contrast to TB, which has infinitely many axioms for truth.

This was the main reason for Davidson to prefer $\mathcal{D}$ over TB and to emphasize the importance of compositionality.

There are still many natural generalizations that are not provable in $\mathcal{D}$.
For instance, if A is a trivially true sentence, one would like to prove that all sentences

> *A and A and A and A and ...*

with arbitrarily many As are true.
More formally, one would like to prove:

> *All sentences of the form*
> $\overline{\forall} = \overline{\forall} \wedge \overline{\forall} = \overline{\forall} \wedge \ldots$ *are true.*

But this is not provable in $\mathcal{D}$.

However, one can add induction principles. This leads to a stronger theory CT, which is no longer conservative over the basic theory of syntax.

I won't go into details here and turn to other theories of truth.

However, one can add induction principles. This leads to a stronger theory CT, which is no longer conservative over the basic theory of syntax.

I won't go into details here and turn to other theories of truth.

The truth predicate of $\mathcal{D}$ still applies only to sentences without the truth predicate. Can we strengthen the axioms so that they do apply also to such sentences?

One way is to add further truth predicate T_1, T_2,… that are added in the same way as T. This gives Tarski's hierarchy of languages.

So we have the language $\mathcal{L}$ without a truth predicate, then a language with the truth predicate T_1, then a language with T_1 and T_2, then a language with T_1, T_2 and T_3, and so on. Each new predicate is axiomatized as the truth predicate for the preceding language.

The truth predicate of $\mathcal{D}$ still applies only to sentences without the truth predicate. Can we strengthen the axioms so that they do apply also to such sentences?

One way is to add further truth predicate $T_1, T_2, \ldots$ that are added in the same way as T. This gives Tarski's hierarchy of languages.

So we have the language $\mathcal{L}$ without a truth predicate, then a language with the truth predicate T_1, then a language with T_1 and T_2, then a language with T_1, T_2 and T_3, and so on. Each new predicate is axiomatized as the truth predicate for the preceding language.

The truth predicate of $\mathcal{D}$ still applies only to sentences without the truth predicate. Can we strengthen the axioms so that they do apply also to such sentences?

One way is to add further truth predicate T_1, T_2,...that are added in the same way as T. This gives Tarski's hierarchy of languages.

So we have the language $\mathcal{L}$ without a truth predicate, then a language with the truth predicate T_1, then a language with T_1 and T_2, then a language with T_1, T_2 and T_3, and so on. Each new predicate is axiomatized as the truth predicate for the preceding language.

The theory FS↾ is given by all axioms of $\mathcal{A}$ plus the following truth-theoretic axioms and the rule below:

Definition

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 sentence $\neg\varphi$ is true iff φ is not true (φ is a sentence possibly containing T).
3. A conditional $\varphi \to \psi$ is true iff φ is false or and ψ is true (φ and ψ are sentences possibly containing T)
4. A universally quantified $\forall x\varphi(x)$ is true iff $\varphi(\overline{e})$ for all objects e ($\varphi(x)$ is a formula possibly containing T).

rule: If you have proved a sentence φ, you may infer that φ is true.

The theory FS↾ is given by all axioms of $\mathcal{A}$ plus the following truth-theoretic axioms and the rule below:

Definition

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 sentence $\neg\varphi$ is true iff φ is not true (φ is a sentence possibly containing T).
3. A conditional $\varphi \rightarrow \psi$ is true iff φ is false or and ψ is true (φ and ψ are sentences possibly containing T)
4. A universally quantified $\forall x\varphi(x)$ is true iff $\varphi(\overline{e})$ for all objects e ($\varphi(x)$ is a formula possibly containing T).

rule: If you have proved a sentence φ, you may infer that φ is true.

The theory FS$\upharpoonright$ is given by all axioms of $\mathcal{A}$ plus the following truth-theoretic axioms and the rule below:

Definition

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 sentence $\neg\varphi$ is true iff φ is not true (φ is a sentence possibly containing T).
3. A conditional $\varphi \to \psi$ is true iff φ is false or and ψ is true (φ and ψ are sentences possibly containing T)
4. A universally quantified $\forall x\varphi(x)$ is true iff $\varphi(\overline{e})$ for all objects e ($\varphi(x)$ is a formula possibly containing T).

rule: If you have proved a sentence φ, you may infer that φ is true.

The theory FS$\upharpoonright$ is given by all axioms of $\mathcal{A}$ plus the following truth-theoretic axioms and the rule below:

Definition

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 sentence $\neg\varphi$ is true iff φ is not true (φ is a sentence possibly containing T).
3. A conditional $\varphi \to \psi$ is true iff φ is false or and ψ is true (φ and ψ are sentences possibly containing T)
4. A universally quantified $\forall x\varphi(x)$ is true iff $\varphi(\overline{e})$ for all objects e ($\varphi(x)$ is a formula possibly containing T).

rule: If you have proved a sentence φ, you may infer that φ is true.

So FS$\restriction$ is like $\mathcal{D}$ except that we have removed the restriction to T-free sentences.

FS$\restriction$ proves many sentences with iterated applications of T.

FS$\restriction$ is ω-inconsistent.

So FS↾ is like $\mathcal{D}$ except that we have removed the restriction to T-free sentences.

FS↾ proves many sentences with iterated applications of T.

FS↾ is ω-inconsistent.

So FS$\restriction$ is like $\mathcal{D}$ except that we have removed the restriction to T-free sentences.

FS$\restriction$ proves many sentences with iterated applications of T.

FS$\restriction$ is ω-inconsistent.

*References

Hartry Field. Deflating the conservativeness argument. *Journal of Philosophy*, 96:533–540, 1999.

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

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.

Stewart Shapiro. Proof and truth: Through thick and thin. *Journal of Philosophy*, 95:493–521, 1998.