AXIOMATIC AND SEMANTIC APPROACHES TO TRUTH

Volker Halbach

Numbers and Truth
Göteborg

21. Oktober 2012

I'm interested in theories of truth for arithmetic.

Two families of approaches:

— axiomatic theories of truth (Tarski, Davidson, Feferman)

— semantic theories of truth (Tarski, satisfaction classes, Kripke, revision theory)

$\mathcal{L}_{\text{PA}}$ is the language of first-order arithmetic; $\mathcal{L}_T$ is $\mathcal{L}_{\text{PA}}$ augmented with a new unary predicate symbol T.

A natural coding is assumed and expressions are identified with their codes.

Models of $\mathcal{L}_T$ take the form $\langle \mathfrak{M}, S \rangle$ where $\mathfrak{M}$ is a model of arithmetic and S is a subset of the domain of $\mathfrak{M}$. A standard model is a model $\langle \mathbb{N}, S \rangle$ where $\mathbb{N}$ is the standard model of arithmetic and $S \subseteq \omega$.

Classical logic is assumed unless I explicitly specify another logic.

An axiomatic theory of truth is a set of rules and axioms formulated in the language $\mathcal{L}_T$.

An axiomatic theory of truth is a set of rules and axioms formulated in the language $\mathcal{L}_T$.

A semantic theory is given by a set of models for the language $\mathcal{L}_T$. Many semantic theories determine standard models. So a semantic theory needs to determine a class Γ of sets (extensions) of sentences, e.g., the class of all sets that are fixed points of an operator.

Γ may contain only a single set.

The models $\langle \mathbb{N}, S \rangle$ with $S \in \Gamma$ are then the intended models.

On many semantic accounts classical logic is abandoned. In this case models may be defined in a different way; for instance, there may be pairs of extensions and antiextensions and models take the form $(\mathbb{N}, S_1, S_2)$.

I assume that the motivation for axiomatic theories is clear. But what purpose is served by the semantic theories?

1. Semantic theories provide insights into the model theory of axiomatic theories of truth.

I assume that the motivation for axiomatic theories is clear. But what purpose is served by the semantic theories?

1. Semantic theories provide insights into the model theory of axiomatic theories of truth.
2. Semantic theories elucidate the real meaning of the truth predicate.

I assume that the motivation for axiomatic theories is clear. But what purpose is served by the semantic theories?

1. Semantic theories provide insights into the model theory of axiomatic theories of truth.
2. Semantic theories elucidate the real meaning of the truth predicate.
3. Semantic theories are tools in the development of axiomatic theories (and perhaps vice versa).

What does it mean to say that a formal deductive system *captures* or ‘formalises’ a semantic construction or a semantic theory of truth?

First proposal: A theory TT captures a semantic theory (specifiying a set Γ of extensions) *iff* the axioms and rules mirror the clauses of the semantic definition of Γ.

STRUCTURAL SIMILARITY OF SEMANTIC CLAUSES WITH AXIOMS

$\phi \in S$ iff

(ϕ is a true atomic sentence) or

(ϕ is a true negated atomic sentence) or

(ϕ is of the form $\psi \land \chi$ and $\psi \in S \land \chi \in S$) or

$\dots$

$$
\begin{aligned}
&Tx \leftrightarrow \\
&\quad \exists s, t\,(x = s \underset{\cdot}{=} t \land s^\circ = t^\circ) \lor \\
&\quad \exists s, t\,(x = \underset{\cdot}{\neg}(s \underset{\cdot}{=} t) \land s^\circ \neq t^\circ) \lor \\
&\quad \exists y, z\,(x = y \underset{\cdot}{\land} z \land Ty \land Tz) \lor \\
&\quad \dots
\end{aligned}
$$

Because we cannot axiomatize $\mathbb{N}$ categorically in (first-order) logic, we can try to fix the standard model (ie. consider only ω-models) and see whether the axioms for truth can fix the extension of T categorically.

Second proposal: A theory TT captures a semantic theory (that specifies a class Γ of extensions) *iff* for all $S \subseteq \omega$

$$\langle \mathbb{N}, S \rangle \models \textsc{tt} \ \text{ iff } \ S \in \Gamma$$

Second proposal: A theory TT captures a semantic theory (that specifies a class Γ of extensions) *iff* for all $S \subseteq \omega$

$$\langle \mathbb{N}, S \rangle \vDash \text{TT} \ \text{ iff } \ S \in \Gamma$$

All the systems I'm going to consider comprise the axioms of PA and the axiom $\forall x\,(Tx \rightarrow \mathrm{Sent}(x))$ (if the system is typed) or $\forall x\,(Tx \rightarrow \mathrm{Sent}_T(x))$ (if it's type-free).

THEOREM

$\langle \mathbb{N}, S \rangle \vDash_{\text{TB}}$ iff S is the set of true $\mathcal{L}_{\text{PA}}$-sentences.

$\langle \mathbb{N}, S \rangle \vDash_{\text{CT}}$ iff S is the set of true $\mathcal{L}_{\text{PA}}$-sentences.

$\langle \mathbb{N}, S \rangle \vDash_{\text{KF}}$ iff S is a SK-fixed point of Kripke's operator.

$\langle \mathbb{N}, S \rangle \vDash_{\text{FS}_n}$ iff S can be obtained by revising a set n-times in the sense of revision semantics.

The base theory becomes irrelevant under this approach.

What amount of induction we have is irrelevant for these adequacy theorems.

Conjecture

The sequent $\Gamma \Rightarrow \Delta$ is derivable in PKF, *iff* the following holds for all SK-standard models $(\mathbb{N}, S_1, S_2)$ and all sequences $\vec{n}$ of numbers:

(i) If $(\mathbb{N}, S_1, S_2) \models_{\text{SK}} \psi(\vec{n})$ for all formulae $\psi \in \Gamma$, then there is a formula $\phi \in \Delta$ with $(\mathbb{N}, S_1, S_2) \models_{\text{SK}} \phi(\vec{n})$. Here $\phi(\vec{n})$ is the result of substituting the free variables in ϕ uniformly with the corresponding numerals from the sequence $\vec{n}$.

(ii) If $(\mathbb{N}, S_1, S_2) \models_{\text{SK}} \neg\psi(\vec{n})$ for all formulae $\psi \in \Delta$, then there is a formula $\phi \in \Gamma$ with $(\mathbb{N}, S_1, S_2) \models \neg\phi(\vec{n})$.

There is no theory TT such that:
$\langle \mathbb{N}, S \rangle \vDash$ TT iff S is the minimal fixed point of Kripke's operator.

Thus Burgess' (KF_μ) or Cantini's (KF+GID) attempts to capture the minimal fixed point fail to be categorical in the same way as KF.

Cf. ID_1.

There is no theory TT such that:
$\langle \mathbb{N}, S\rangle \models$ TT iff S is the minimal fixed point of Kripke's operator.

Thus Burgess' (KF_μ) or Cantini's (KF+GID) attempts to capture the minimal fixed point fail to be categorical in the same way as KF.

Cf. ID_1.

Would we say that the minimal fixed point model cannot be captured by an axiomatic theory, while the class of *all* fixed point models can?

Stable and nearly stable truth can also not be captured by an axiomatic system is this sense.

There is no theory TT such that:
$\langle \mathbb{N}, S \rangle \models \textsc{tt}$ iff S is the set of stable truths.

Conclusion: By considering the standard models only in the way above, one cannot decide whether an axiomatic theory captures a semantic theory.

Conclusion: By considering the standard models only in the way above, one cannot decide whether an axiomatic theory captures a semantic theory.

A proof-theoretic analysis may be useful. But it's not sufficient. Example:

The theory PUTB of positive disquotation doesn't capture the set of Kripke fixed-points: Even though it can define the truth predicate of KF, it admits standard models beyond the Kripke fixed-points.

In the general case it's not clear to me what it means to say that an axiomatic theory of truth captures a semantic theory.

The value of semantic constructions of high complexity for the development of axiomatic theories remains somewhat unclear (other than for pure consistency proofs).

Definition (TB)

The theory TB comprises all axioms of PAT, that is, of Peano arithmetic formulated in $\mathcal{L}_T$ including all instances of the induction schema with the truth predicate. Moreover all sentences of the form $T\ulcorner\phi\urcorner \leftrightarrow \phi$ are axioms of the theory where ϕ is a sentence of the language of $\mathcal{L}_{\text{PA}}$.

Definition (CT)

The system CT↾ is given by all the axioms of PA and the following axioms:

CT1 $\forall s\,\forall t\,\big(T(s \underset{\cdot}{=} t) \leftrightarrow s^\circ = t^\circ\big)$

CT2 $\forall x\,\big(\mathrm{Sent}(x) \to (T(\underset{\cdot}{\neg} x) \leftrightarrow \neg Tx)\big)$

CT3 $\forall x\,\forall y\,\big(\mathrm{Sent}(x \underset{\cdot}{\wedge} y) \to (T(x \underset{\cdot}{\wedge} y) \leftrightarrow T(x) \wedge T(y))\big)$

CT4 $\forall x\,\forall y\,\big(\mathrm{Sent}(x \underset{\cdot}{\vee} y) \to (T(x \underset{\cdot}{\vee} y) \leftrightarrow T(x) \vee T(y))\big)$

CT5 $\forall v\,\forall x\,\big(\mathrm{Sent}(\underset{\cdot}{\forall} v x) \to (T(\underset{\cdot}{\forall} v x) \leftrightarrow \forall t\, T(x(t/v)))\big)$

CT6 $\forall v\,\forall x\,\big(\mathrm{Sent}(\underset{\cdot}{\exists} v x) \to (T(\underset{\cdot}{\exists} v x) \leftrightarrow \exists t\, T(x(t/v)))\big)$

Definition (FS)

The system FS$_n$ is given by all the axioms of PAT (including induction axioms containing the truth predicate), at most n successive application of the necessitation rule $\frac{\phi}{T\ulcorner\phi\urcorner}$ and the following axioms:

FS1 $\forall s\, \forall t \left(T(s \underset{.}{=} t) \leftrightarrow s^{\circ} = t^{\circ} \right)$

FS2 $\forall x \left(\mathrm{Sent}_T(x) \rightarrow (T \underset{.}{\neg} x \leftrightarrow \neg T x) \right)$

FS3 $\forall x\, \forall y \left(\mathrm{Sent}_T(x \underset{.}{\wedge} y) \rightarrow (T(x \underset{.}{\wedge} y) \leftrightarrow (Tx \wedge Ty)) \right)$

FS4 $\forall x\, \forall y \left(\mathrm{Sent}_T(x \underset{.}{\vee} y) \rightarrow (T(x \underset{.}{\vee} y) \leftrightarrow (Tx \vee Ty)) \right)$

FS5 $\forall v\, \forall x \left(\mathrm{Sent}_T(\underset{.}{\forall} v x) \rightarrow (T(\underset{.}{\forall} v x) \leftrightarrow \forall t\, T(x(t/v))) \right)$

FS6 $\forall v\, \forall x \left(\mathrm{Sent}_T(\underset{.}{\exists} v x) \rightarrow (T(\underset{.}{\exists} v x) \leftrightarrow \exists t\, T(x(t/v))) \right)$

Definition (KF)

The system KF is given by all the axioms of PAT and the axioms:

KF1 $\forall s\, \forall t \big(T(s \underset{\cdot}{=} t) \leftrightarrow s^\circ = t^\circ \big)$

KF2 $\forall s\, \forall t \big(T(\underset{\cdot}{\neg} s \underset{\cdot}{=} t) \leftrightarrow s^\circ \neq t^\circ \big)$

KF3 $\forall x \big(\mathrm{Sent}_T(x) \to (T(\underset{\cdot}{\neg}\underset{\cdot}{\neg} x) \leftrightarrow Tx) \big)$

KF4 $\forall x\, \forall y \big(\mathrm{Sent}_T(x \underset{\cdot}{\wedge} y) \to (T(x \underset{\cdot}{\wedge} y) \leftrightarrow Tx \wedge Ty) \big)$

KF5 $\forall x\, \forall y \big(\mathrm{Sent}_T(x \underset{\cdot}{\wedge} y) \to (T\underset{\cdot}{\neg}(x \underset{\cdot}{\wedge} y) \leftrightarrow T\underset{\cdot}{\neg}x \vee T\underset{\cdot}{\neg}y) \big)$

KF6 $\forall x\, \forall y \big(\mathrm{Sent}_T(x \underset{\cdot}{\vee} y) \to (T(x \underset{\cdot}{\vee} y) \leftrightarrow Tx \vee Ty) \big)$

KF7 $\forall x\, \forall y \big(\mathrm{Sent}_T(x \underset{\cdot}{\vee} y) \to (T\underset{\cdot}{\neg}(x \underset{\cdot}{\vee} y) \leftrightarrow T\underset{\cdot}{\neg}x \wedge T\underset{\cdot}{\neg}y) \big)$

KF8 $\forall v\, \forall x \big(\mathrm{Sent}_T(\underset{\cdot}{\forall} v x) \to (T(\underset{\cdot}{\forall} v x) \leftrightarrow \forall t\, T(x(t/v))) \big)$

KF9 $\forall v\, \forall x \big(\mathrm{Sent}_T(\underset{\cdot}{\forall} v x) \to (T(\underset{\cdot}{\neg}\underset{\cdot}{\forall} v x) \leftrightarrow \exists t\, T(\underset{\cdot}{\neg} x(t/v))) \big)$

KF10 $\forall v\, \forall x \big(\mathrm{Sent}_T(\underset{\cdot}{\exists} v x) \to (T(\underset{\cdot}{\exists} v x) \leftrightarrow \exists t\, T(x(t/v))) \big)$

KF11 $\forall v\, \forall x \big(\mathrm{Sent}_T(\underset{\cdot}{\exists} v x) \to (T(\underset{\cdot}{\neg}\underset{\cdot}{\exists} v x) \leftrightarrow \forall t\, T(\underset{\cdot}{\neg} x(t/v))) \big)$

KF12 $\forall t \big(T(\underset{\cdot}{T} t) \leftrightarrow T t^\circ \big)$

KF13 $\forall t \big(T\underset{\cdot}{\neg}\underset{\cdot}{T} t \leftrightarrow (T\underset{\cdot}{\neg} t^\circ \vee \neg\, \mathrm{Sent}_T(t^\circ)) \big)$

Definition (PUTB)

The theory PUTB is given by the axioms of PAT including all the induction axioms in the language $\mathcal{L}_T$ with the truth predicate and the set of all sentences

$$\forall t_1 \ldots \forall t_n \left(T\ulcorner \phi(\underset{\cdot}{t}_1, \ldots, \underset{\cdot}{t}_n)\urcorner \leftrightarrow \phi(t_1{}^{\circ}, \ldots, t_n{}^{\circ})\right),$$

where $\phi(x_1, \ldots, x_n)$ contains at most the variables $x_1, \ldots, x_n$ free and all occurrences of the truth predicate in $\phi(x_1, \ldots, x_n)$ are T-positive.

The role of the T-sentences in Tarski's account of truth is stated in Convention T:

> CONVENTION T. *A formally correct definition of the symbol 'Tr', formulated in the metalanguage, will be called an adequate definition of truth if it has the following consequences:*
> *(α) all sentences which are obtained from the expression '$x \in Tr$ if and only if p' by substituting for the symbol 'x' a structural-descriptive name of any sentence of the language in question and for the symbol 'p' the expression which forms the translation of this sentence into the metalanguage;*
> *(β) the sentence 'for any x, if $x \in Tr$ then $x \in S$' (in other words '$Tr \subseteq S$').*

In the *Wahrheitsbegriff* Tarski considers disquotationalism, that is, an axiomatisation of truth based on the T-sentences.

In the *Wahrheitsbegriff* Tarski considers disquotationalism, that is, an axiomatisation of truth based on the T-sentences.

What he says on this axiomatic approach undermines the adequacy of Convention T.

In the *Wahrheitsbegriff* Tarski considers disquotationalism, that is, an axiomatisation of truth based on the T-sentences.

What he says on this axiomatic approach undermines the adequacy of Convention T.

At first Tarski proves that adding the typed T-sentences to a theory yields a consistent extension of the theory:

Tarski continues:

> *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 ‘$x\overline{\in}Tr$ or $\overline{x}\overline{\in}Tr$’. [“$\in$ 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.*

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.

The problem

On the one hand, according to Tarski, the theory generated by all T-sentences (and the axiom saying that only sentences are true) should be adequate because it satisfies Convention T; on the other hand, Tarski thinks that it isn't adequate because it doesn't prove the general principle of contradiction.

The problem

On the one hand, according to Tarski, the theory generated by all T-sentences (and the axiom saying that only sentences are true) should be adequate because it satisfies Convention T; on the other hand, Tarski thinks that it isn't adequate because it doesn't prove the general principle of contradiction.

It doesn't help to insist that Tarski's Convention T applies to *definitions* of truth only. Given a definition, one can define in that theory a truth predicate homophonically.

An example

There are real life examples of truth definitions that yield only T-sentences but that do not yield the general principles such as the law of contradiction.

An example

There are real life examples of truth definitions that yield only T-sentences but that do not yield the general principles such as the law of contradiction.

I suspect that many logicians thought that any serious definition of truth must prove that the truth predicate commutes with the connectives (and quantifiers) and thus proves the law of contradiction. If this were true then a truth definition would always yield a consistency proof (under fairly general conditions).

Thus the proof that Bernays-Gödel defines truth for Zermelo-Fraenkel (Mostowski 1950 and Wang 1952) caused some confusion because it was also known that BG is conservative over ZF (Novak 1950).