# AXIOMATIC APPROACHES TO TRUTH I

## Philosophical and mathematical background

Volker Halbach

Axiomatic Theories of Truth

Neuchâtel

6 September 2014

# Plan of the lecture

In an axiomatic theory of truth, the truth or satisfaction predicate is taken to be a primitive expression.

Usually the term *axiomatic theories of truth* refers to formal axiomatic theories.

*History*

Tarski [17] formulated and studied formal axiomatic truth theories. His adequacy condition can be understood as an axiomatization.

Davidson advocated axiomatic truth theories in the 1960s without formulating the theories.

In the late 1970s Feferman started to work on axiomatizations of Kripke's theory.

Truth and satisfaction predicates are frequently and essentially used in various areas of philosophy.

Examples

- $S$ knows $p$ iff $S$ believes $p$, $S$ is justified in his/her belief $p$ and $p$ is true (and some Gettier condition is satisfied).
- Whatever I clearly and distinctly perceive is true.
- There are unknowable truths. $\exists x\,(\neg \Diamond Kx \wedge Tx)$
- There are contingent a priori truths.
- 'Stealing money is wrong' is neither true nor false.

For philosophical applications we need a truth predicate that is not relativized to a model or structure.

The truth predicate is global in the sense that it is sensibly applicable to arbitrary sentences of one's language (or, perhaps, to all propositions).

Don't trust the logicians!

How nonclassical logic spreads:

> *S* knows *p* iff *S* believes *p*, *S* is justified in his/her belief *p* and *p* is true (and some Gettier condition is satisfied).
>
> I know that for any sentence the sentence itself or its negation is true.
>
> Necessarily, the truth teller is true.
>
> All sentences of the form $\phi \rightarrow \phi$ are analytic.

A truth predicate can serve certain purposes in reductions. Adding a truth or satisfaction predicate to a language has effects that are similar to adding propositional quantifiers or second-order quantifiers.

Hence commitment to second-order objects can be eliminated by the use of a truth or satisfaction predicate.

Truth predicates can be used to express soundness claims of the sort ‘All theorems of theory $\mathcal{T}$ are true.’

Truth predicates relativized to a model or structure are heavily used in logic.

In logic, truth predicates relativized to certain language fragments are also used, e.g., ‘partial’ truth predicates.

A global, unrelativized truth predicate, however, is usually not definable, as Tarski’s theorem on the undefinability of truth shows. But one can still add such a predicate axiomatically.

*Proof theory*

As the truth predicate needs to apply to objects (sentences, propositions), the truth axioms are added to a theory of truth bearers (the base theory). The most popular choice is arithmetic.

Adding truth axioms is very similar to adding second-order quantifiers to arithmetic. This can help with formulating and analyzing subsystems of second-order arithmetic.

## Example

The theory $\widehat{\mathrm{ID}}_1$ is equivalent to PA plus uniform disquotation for T-positive formulae:

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

## Example

Systems $\mathrm{RA}_{<\alpha}$ of ramified analysis are important in the analysis of predicativity (Schütte, Feferman). They are difficult to formulate. Using ramified truth predicates, one can give more straightforward formulations of equivalent systems.

Sometimes truth theories are useful in proving theorems:

## Example

Consider the system PA plus elementary comprehension without parameters, that is, $\exists X \, \forall y \, (y \in X \leftrightarrow \phi(y))$ where $\phi(y)$ is pure first-order. By interpreting this system in UTB it can easily shown to be conservative over PA.

*Model theory*

### Definition

A type over a model $\mathfrak{M}$ is a finitely satisfied set of formulae $\phi(x, \overline{b})$ that have exactly the variable $x$ free and contain at most the parameter $\overline{b}$ for one fixed object $b \in |\mathfrak{M}|$. A type $p$ is recursive if and only if the set of codes of formulae $\phi(x, y)$ with $\phi(x, \overline{b}) \in p$ is recursive.

### Definition (recursive saturation)

A model $\mathfrak{M}$ of Peano arithmetic is recursively saturated if and only if every recursive type over $\mathfrak{M}$ is realized (i.e. satisfied).

$\mathrm{CT}{\restriction}$ is the system with the axioms of PA and the 'Tarski clauses' as axioms.

**Theorem (Kotlarski et al. [8], Lachlan [11])**

For all countable models $\mathfrak{M}$ of PA:

There is $S$ s.t. $(\mathfrak{M}, S) \models \mathrm{CT}{\restriction}$ iff $\mathfrak{M}$ is recursively saturated.

The theory is conservative over PA. There is further recent work by Enayat & Visser and Leigh.

1. Is truth definable? Is it reducible, conservative or eliminable?
2. Which axioms and rules about truth can be sensibly combined?
3. What's the expressive and deductive power of truth? What's the role of truth in reasoning?
4. To what extent can theories of truth replace second-order quantifiers and play a role in foundations?
5. How can truth be used to make explicit assumptions implicit in the acceptance of theories?
6. How compare different conceptions of truth? Are compositional truth theories always stronger than disquotational ones? How compare classical with the various nonclassical theories?
7. Is the theory of truth finitely axiomatizable?
8. Can a truth theory be categorical in some sense?
9. How are semantic concepts like compositionality related to mathematical concepts such as predicativity?

# AXIOMATIC APPROACHES TO TRUTH II

## A survey of systems

Volker Halbach

Axiomatic Theories of Truth

Neuchâtel

21 July 2014

# Classification of truth theories

1. non-classical theories
   There is much work on paraconsistent and other nonclassical logics but less on full theories of truth
   Field [6], Kremer [9], Feferman [3], Halbach & Horsten [7], Leigh & Rathjen [12]
2. classical theories
   They can be categorized along the following criteria:
   1. typed vs type-free
   2. disquotational vs compositional

All truth theories are relative to a base theory.

Before we can formulate a theory of truth, we should have a theory of truth bearers, e.g., a theory of syntax, propositions, or the like.

I'll use (first-order) Peano arithmetic here. Many of the results can be be applied to other base theories, e.g. a theory of concatenation, Zermelo-Fraenkel set theory.

So the truth theories are formulated in the language of first-order arithmetic plus a unary predicate $T$ for truth.

All theories considered below are formulated in classical logic.

## Tarskian disquotation

TB$\restriction$ contains all axioms of PA plus all sentences

$$T\ulcorner\phi\urcorner \leftrightarrow \phi$$

for all sentences without $T$. TB$\restriction$ has only induction without $T$.

This is the minimal theory that is adequate in the sense of Convention T (after adding $\forall x\,(Tx \rightarrow \mathrm{Sent}(x))$).

Theories similar to TB$\restriction$ play a role in deflationary accounts of truth.

## Theorem

TB$\restriction$ is conservative over PA and thus consistent (Tarski). The truth predicate $T$ isn't definable in PA (undefinability of truth). Any model of PA can be extended to a model of TB$\restriction$.

Here are some ways to strengthen TB$\restriction$.

## Tarskian disquotation with full induction

TB contains all axioms of PA including all induction axioms with $T$ plus all sentences

$$T\ulcorner\phi\urcorner \leftrightarrow \phi$$

for all sentences without $T$.

## Theorem

TB is still conservative over PA. But not any model of PA can be extended to a model of TB (Engström 2009).

# A disquotational theory of satisfaction

## uniform Tarskian disquotation with full induction

UTB contains all axioms of PA including all induction axioms with $T$ plus all sentences

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

where $\phi(x_1, \ldots, x_n)$ is a formula of the language of $\mathcal{L}$ with exactly $x_1, \ldots, x_n$ free.

## Theorem

UTB is conservative over PA.

*Typed* disquotational theories are weak (conservative over PA). Also, they don't prove generalizations such as:

$$\forall x \, \forall y \left( \mathrm{Sent}(x \dot{\wedge} y) \rightarrow (T(x \dot{\wedge} y) \leftrightarrow T(x) \wedge T(y)) \right)$$

Untyped disquotational theories can be very strong.

**Theorem**

Any theory extending PA can reaxiomatized by a set of disquotation sentences $T\ulcorner\phi\urcorner \leftrightarrow \phi$ over PA (McGee [16] using a variant of Curry's paradox).

But these theories are not well motivated.

## uniform Tarski disquotation for T-positive sentences

PUTB contains all axioms of PA including all induction axioms with $T$ plus all sentences

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

$\phi(x_1, \dots, x_n)$ is a formula of the language with $T$ with exactly $x_1, \dots, x_n$ free such that $T$ does not occur in the scope of $\neg$ ($\wedge$ and $\vee$ are the only connectives).

## Theorem

PUTB is equivalent to $\mathrm{RA}_{<\epsilon_0}$ and KF below.

There are reasonable, stronger disquotational theories. For instance, one can get second-order arithmetic (without second-order parameters) from a disquotational truth theory (Schindler).

Generally, disquotational theories can vary significantly in their properties, depending on how the paradoxes are blocked.

All known *natural* disquotational theories fail to prove generalizations such as

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

Thus various philosophers have tried to add them as axioms, although Tarski was dismissive of such attempts.

Compositional axioms may allow finite axiomatizability.

## typed compositional truth

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

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

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

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

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

Induction is restricted to sentences without $T$.

## Theorem ([8],[2],[13])

CT↾ is conservative over PA.

There are claims in the literature to the effect that the 'Tarskian' clauses fix the extension of the truth predicate.

If a model of PA can be expanded to a model of $\mathrm{CT}{\restriction}$ at all (in which case it will be rec. saturated), there will be uncountably many ways to do so.

Also $\mathrm{CT}{\restriction}$ does not prove hat all sentences of the form

$$0=0 \wedge 0=0 \wedge 0=0 \wedge \ldots \wedge 0=0$$

are true.

So let's add induction.

## typed compositional truth with full induction

CT is $\mathrm{CT}\restriction$ plus all induction axioms in the language with $T$.

Using induction with $T$ one proves in CT that all axioms of PA are true, and then using induction again, one proves that all *theorems* of PA are true. Since $0 \neq 1$ is not true, CT proves that PA is consistent.

## Theorem

CT fails to be conservative over PA. The effect of adding the CT axioms to PA is the same as adding elementary comprehension. CT is equivalent to ACA.

No truth theory containing TB is categorical (Beth's theorem).

But CT 'fixes the extension of the truth predicate' in the following way:

### Theorem

Let CT′ be CT with the predicate $T'$ instead of $T$. The theory $\mathrm{CT} \cup \mathrm{CT}'$ *plus induction in the mixed language* proves $\forall x \left( \mathrm{Sent}(x) \to (Tx \leftrightarrow T'x) \right)$.

To get even stronger theories one can iterate the theory CT along some ordinal notation system. The system turns out to be equivalent with iterated predicative comprehension.

Type-free compositional theories are usually extensions of CT.

Criteria for choosing theories:

1. consistency, $\omega$-consistency, and absence of non-trivial models
2. deductive and conceptual power
3. 'transparency' or symmetry
4. Leitgeb: *What theories of truth should be like (but cannot be)*

All theories below have full induction.

FS is a system of *classical* and *symmetric* truth.

## Friedman–Sheard

The system FS has all axioms of PA plus the following axioms and rules:

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

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

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

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

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

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

NEC $\dfrac{\phi}{T\ulcorner\phi\urcorner}$ $\qquad$ $\dfrac{T\ulcorner\phi\urcorner}{\phi}$ CONEC

## Theorem

The liar sentence is neither provable nor refutable in FS.

CONEC is not needed for proof-theoretic strength; NEC can be expressed via iterated reflection.

Natural models of FS can be obtained via finitely iterated revision (in Gupta–Herzberger style).

FS is $\omega$-inconsistent. (McGee [15])

FS is equivalent to finitely iterated Tarskian truth, i.e., iterated CT.

This is an axiomatization of Kripke's[10] theory of truth with Strong Kleene logic.

## Kripke–Feferman

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

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

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

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

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

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

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

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

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

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

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

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

KF12 $\forall t\, (T(\underset{.}{T} t) \leftrightarrow T t^\circ)$

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

KF14 $\forall x \big( \mathrm{Sent}_T(x) \rightarrow \neg (Tx \wedge T \underset{.}{\neg} x)\big)$

Several variants of KF can be found in the literature. This is my version.

The last axiom – called the consistency axiom – excludes truth-value gluts.

## Theorem

KF is an axiomatization of Kripke's theory of truth with the Strong Kleene schema.

It's an axiomatization of a partial notion of truth in classical logic.

KF proves the liar sentence, as it excludes truth-value gluts.

Even without the consistency axiom KF14, KF cannot be closed consistently under NEC and CONEC. It's essentially asymmetric.

KF is equivalent to $\epsilon_0 = \omega^{\omega^{\cdot^{\cdot^{\omega}}}}$ iterated Tarskian truth (Feferman [4]). Feferman's analysis proceeds in terms of infinite conjunctions.

Further systems:

- Variations of KF: Feferman's [4] strong reflective closure of PA, weak Kleene, Feferman's [5] DT
- Cantini's [1] VF and supervaluations
- axiomatizations of stable truth?

# Some preliminary conclusions

The metamathematical property of the axioms systems differs significantly, depending on which solution of the paradoxes is adopted.

*Typed* disquotational theories are proof-theoretically weak (conservative); type-free disquotational theories can be very strong. So, don't assume that disquotation is weak.

Typed truth theories give at most elementary comprehension. Compositional truth theories don't seem to exceed the strength of predicative systems.

Strong type-free truth predicates tend to be partial (or 'paraconsistent'); but presenting these truth predicates in a nonclassical logic will weaken them.

Conclusions

for the logician: Axiomatic theories of truth expand the playground.

for the philosopher: We have various truth theories on offer. But they differ wildly in their properties. Pick your favourite!

Andrea Cantini.
A theory of formal truth arithmetically equivalent to $ID_1$.
*Journal of Symbolic Logic*, 55:244–259, 1990.

Ali Enayat and Albert Visser.
Full satisfaction classes in a general setting (part I).
Technical report, 201?
to appear in *Logic Group Preprint Series* , available at http://http://www.phil.uu.nl/preprints/lgps/.

Solomon Feferman.
Towards useful type-free theories I.
*Journal of Symbolic Logic*, 49:75–111, 1984.

Solomon Feferman.
Reflecting on incompleteness.
*Journal of Symbolic Logic*, 56:1–49, 1991.

Solomon Feferman.
Axioms for determinateness and truth.
*Review of Symbolic Logic*, 1:204–217, 2008.

Hartry Field.
*Saving Truth From Paradox*.
Oxford University Press, Oxford, 2008.

Volker Halbach and Leon Horsten.
Axiomatizing Kripke's theory of truth.
*Journal of Symbolic Logic*, 71:677–712, 2006.

Henryk Kotlarski, Stanislav Krajewski, and Alistair Lachlan.
Construction of satisfaction classes for nonstandard models.
*Canadian Mathematical Bulletin*, 24:283–293, 1981.

Michael Kremer.
Kripke and the logic of truth.

*Journal of Philosophical Logic*, 17:225–278, 1988.

Saul Kripke.
Outline of a theory of truth.
*Journal of Philosophy*, 72:690–712, 1975.
reprinted in [14].

Alistair Lachlan.
Full satisfaction classes and recursive saturation.
*Canadian Mathematical Bulletin*, 24:295–297, 1981.

Graham Leigh and Michael Rathjen.
The Friedman–Sheard programme in intuitionistic logic.
*Journal of Symbolic Logic*, 77:777–806, 2012.

Graham E. Leigh.
Conservativity for theories of compositional truth via cut elimination.
1308.0168 [math.LO]; http://arxiv.org/abs/1308.0168, 2013.

Robert L. Martin, editor.
*Recent Essays on Truth and the Liar Paradox.*
Clarendon Press and Oxford University Press, Oxford and New York, 1984.

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

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

Alfred Tarski.
Der Wahrheitsbegriff in den formalisierten Sprachen.
*Studia Philosophica Commentarii Societatis Philosophicae Polonorum*, 1:261–405, 1935.

reprinted as 'The Concept of Truth in Formalized Languages' in [18, 152–278]; page references are given for the translation.

Alfred Tarski.
*Logic, Semantics, Metamathematics.*
Clarendon Press, Oxford, 1956.